HAL
open science

Digital Scholarly Editing: Theories, Models and Methods

Elena Pierazzo

▶ To cite this version:

Elena Pierazzo. Digital Scholarly Editing: Theories, Models and Methods. 2014. hal-01182162

HAL Id: hal-01182162
https://hal.univ-grenoble-alpes.fr/hal-01182162

Preprint submitted on 30 Jul 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Digital Scholarly Editing: Theories, Models and Methods

Elena Pierazzo, Université de Grenoble 'Stendhal', France

Contents

List of Figures and Tables

Acknowledgments [TO COME]

Introduction

1 **Traditional and Emerging Editorial Models**

1.1 Traditional Editorial Models: an Overview

1.2 Digital Editing, Digital Editions

1.3 Digital Editorial Models

1.3.1 Phylogenetic

1.3.2 Social Editing and Social Editions

1.3.3 Crowdsourcing and Editing

1.4 The Source-and-the-Output Model: Paradigmatic Editions

1.5 Raising the Stakes: Interactive Facsimiles and Gamification

1.6 Functional Analysis of Digital Scholarly Editions: an Edition Is What an Edition Does

2 **Modelling (Digital) Texts**

2.1 Modelling and the Edition

2.2 Modelling Texts and Documents

2.3 The Material and the Immaterial Text: more on Dimensions

2.4 Expanding the Work: Versions and Derivative Works

2.5 Other models

2.6 A Born-digital Model: the OHCO Model

3 **Modelling Text Transmission: from Documents to Texts, and Return**

3.1 Textual Transmission: a Communication Model

3.2 Theories and models of transcriptions

3.3 Returning to the Document: the Success of Digital Documentary Editions

4 **What's on the page? Objectivity and Interpretation in Scholarly Editing**

4.1 Record and Interpretation

4.2 Objectivity in Textual Editing

4.3 The Digital Facsimile and the 'Real Thing'

4.4 A Faithful Representation

4.5 Markup and Interpretation

5 **Work and Workflow of Digital Scholarly Editions**

5.1 Modelling the Edition

5.2 Computing the Edition: the Two Paths and the Third Way

5.3 The Role of Standards in Textual Editing: the Text Encoding Initiative

5.4 Editorial Work, Collaborative Work

6 **The Publication of Digital Scholarly Editions**

6.1 Work in Progress

6.2 The Shape of the Digital

6.3 Dual and Hybrid Publications

6.4 How Much Information? Open Source Publications

7 **Using Digital Scholarly Editions**

7.1 Readers and Readings: the Audience of Scholarly Editions

7.2 Editorial Formats and the Nature of Texts

7.3 Readers and Users of Digital Scholarly Editions

7.4 Looking for New Models: the Critical Apparatus

7.5 Interface Design: Principles

7.6 A Standard User Interface?

7.7 Are User Interfaces Necessary?

8 **Trusting the Edition: Preservation and Reliability of Digital Editions**

8.1 Data and Metadata: Standards, Formats and Preservation

8.2 Preserving Data, Preserving Interfaces

8.3 Preserving Data: Digital Repositories

8.4. 'Lots Of Copies Keep Stuff Safe': Redundancy and Preservation

8.5 Open Access and Preservation

8.7 Trusting Editions: Heraclitean Editions

8.8 Trusting Digital Editions? Peer Review and Evaluation of Digital Scholarship

9 **The Present and the Future of Digital Scholarly Editions**

9.1 A glance toward the future

9.2. Ethics of Digital Editions

Index

List of Figures and Tables

Figures

1.1 Abbreviation of ‘domino’

2.1 Conceptual model of texts and documents

2.2 Conceptual model of texts and documents (UML)

2.3 Conceptual model of texts, documents and works

2.4 Guillaume Apollinaire, *Calligrame* – Catalogue of the exhibition *Léopold Survage - Irène Lagut* (1917)

2.5 The materiality and immateriality of textual theories

2.6 The materiality and immateriality of editorial models

2.7 The FRBR model (FRBR 1998, p. 13)

2.8 The Pluralistic Text (Sahle 2013)

3.1 Letter ‘A’

4.1 Cod. Sang. 189, p. 76

4.2 Cod. Sang. 189, p. 76 (2)

5.1 A simple wireframe

5.2 Workflow of editing (Rehbein and Fritze, 2012, p. 52).

7.1 XML source of the last two lines of *The Watsons*, p. [1]

7.2 Web display of the last two lines of *The Watson*, p. [1], with and without styling

Tables

1.1 Functional grid (Vanhoutte, 2010, p. 142)

2.1 Material and immaterial dimensions

2.2 Dimensions and the FRBR model

4.1 Epistemic virtues in time (Daston and Galison, 2010, p. 371)

Introduction

Scholarly editorial practices are undergoing deep structural and theoretical changes. Such changes are connected to the adoption of computers both for supporting editorial work and for disseminating it. In spite of what was believed by earlier adopters, computers are proving to be far more than just electronic research assistants able to offer relief from 'idiotic' work; in fact they are leading us to question the hermeneutics and heuristics of textual scholarship. Traditional editorial practices and theories have been developed and shaped by print culture, but, after the first stage of digital editing where the assumption was that it was possible to simply transfer such practices to the new medium, new editorial models have emerged, along with new workflows. New research goals have become pursuable, with the result that even the basic assumptions that were taken for granted are open for scrutiny.

Several books and articles published since the 1990s have discussed various aspects of digital scholarly editions,[1] presenting case studies or particular points of view on some of the activities involved in the preparation of digital editions, but a comprehensive approach to digital scholarly editing from a methodological and theoretical viewpoint is still missing.[2] In this gap this book finds its place, providing a coherent treatment of the subject, in the hope that many more contributions will follow. Digital textual scholarship is both over and under-theorized, in the sense that while there have been a lot of contributions to the fine points of this or that editorial practice, not many contributions have attempted to analyse the significance that changes introduced by digital scholarship have brought into editorial practices.

Before presenting what this book is about, however, perhaps it is better to discuss what this book is not about. It does not provide an historical overview of the field, nor does it discuss how editors have come to engage with computers; others have done this before and, most likely, better than I could ever do.[3] This book is not a guide to producing a digital edition either, as it does not provide explanations of how to use specific tools and techniques. Rather it investigates the changes and the methodological implications of the application of computational methods to all stages of the editorial workflow, and the impact of these on the main protagonists of such processes: editors, readers and editions. In a deep dialogue with previous publications on the topic, the book presents foreseeable lines of development in digital textual scholarship without attempting to propose a new

[1] There are too many books to mention here, but they will be the objects of discussion in the chapters to come.

[2] The exception to this is represented by Sahle 2013, for which see below.

[3] Cf. for instance two contributions by Edward Vanhoutte, one on the history of digital textual scholarship (Vanhoutte, 2010) and the most recent one on the history of Digital Humanities (Vanhoutte, 2013).

editorial approach, claiming instead that the alterations in traditional editorial practices necessitated by the use of computers impose radical changes in the way we think and manage texts, documents, editions and the public.

A book on digital scholarly editing: is it useful or is it an absurdity? And is editing an activity worth pursuing in the age of the Internet? The generalized access to a publication platform such as the web, and the low technical requirements that are needed to produce web content, have brought into question the necessity of selecting, controlling and ultimately editing altogether: 'since tools are now available that make it possible for users to exploit electronic data in a variety of ways, straightforward digitization that makes data available quickly is preferable to a critical edition which is never finished, even if less scholarly value is added' (Ore, 2009, p. 113). Yet, cultural heritage texts are complex objects which have reached us inscribed on stratified, diversified and conflicting documents that require the cultural and dialectical intermediation of expert readers: the editors. This requirement has not changed with the advent of the Web: on the contrary, the easier it becomes to publish on the Web, the larger the quantity of textual materials available, the more important becomes the guidance of the editor. Quality still matters. Any act of publication, of making available something that was not before requires judgement, evaluation and ultimately interpretation of what is worth publishing and what not, and of what is likely to better represent the text, the document or the work one is editing.

Is digital scholarly editing the same as scholarly editing but in a new medium, a new methodology or a new discipline altogether? This book suggests answers to these questions by guiding its reader along the ideal 'textbook' timeline and workflow of the preparation of an edition: from the elaboration of abstract models of the objects to edit, to the choice of the material, to the editorial work, to post-production and publication, to the use of the published edition, to long-term issues and the ultimate significance of the published work. But rather than aiming to be a handbook on digital scholarly editing, its goal is to examine from a theoretical and methodological point of view the issues and problems that emerge along these stages with the application of computational techniques and methods.

Two main contributions by Peter Shillingsburg can be considered among the main interlocutors of this book, because of their focus on digital editorial practices and their breadth (Shillingsburg, 1996 and Shillingsburg, 2006). In his latter contribution (Shillingsburg, 2006, p. 88) he asks if we are there yet with digital editions and concludes that we are not because we are not yet exploiting all the potentials offered by the digital medium. He then adds of digital editions that: 'Yes they are better. No, they are not good enough' (Shillingsburg, 2006, p. 111). A few years on and we can give the same answer: No, they are not good enough. Most digital editions are only timidly engaging with the medium, trying to reproduce on the screen the same experience offered by the printed page, desperately trying to demonstrate that digital editions are as good as printed ones.

According to Patrick Sahle one of the main reasons why this is still happening is due to the power of what he calls 'the page paradigm'. In his major contribution on the topic of digital scholarly editions he maintains:

> One will have to state of traditional editors, even in case they will become aware of the fundamental methodological issues of concern to them, they hardly ever had a chance to think of their editions without constantly having a printed page layout before their eyes. They will always have been conscious of the only possible end point (the printed book) - maybe even more than of the starting point (the documents of the transmission) (Sahle, 2013, vol. 1, p. 270).[4]

The persistence of the cultural model provided by the book cannot surprise. Such a model has deeply shaped the way we work and our expectations of the way things should work. Yet new models have started to emerge, both as a way to answer to new scholarly questions and also under the pressure of new generations of young scholars, who are nurtured within and by different cultural premises. Some digital editions have been built around more complex and rich interrelations with the paper culture, and are pushing, gently, for the moment, the boundaries of textual scholarship to include new discourses and hermeneutics.

The emergence of such models and the disorienting feeling of being in a completely new territory have sometimes raised defence mechanisms in editors, worried that the changed conditions offered by the new environment will eventually undermine the specificity of editorial work. Hence a great theoretical effort has been deployed in order to distinguish digital archives from digital editions and documentary editions from critical editions (Vanhoutte, 2010, 131-5). An analogous effort has

[4] 'Man wird wohl konstatieren müssen, dass traditionelle Editoren, selbst wenn sie sich der fundamentalen methodischen Probleme ihres Anliegens bewusst waren, kaum eine Chance hatten, ihre Editionen zu denken, ohne dabei ständig ein gedrucktes Seitenlayout vor Augen zu haben. Immer wird ihnen schon der einzige mögliche Endpunkt (das gedruckte Buch) – vielleicht sogar noch stärker als der Ausgangspunkt (die Überlieferung) – bewusst gewesen sein'. I am grateful to Patrick Sahle for having provided me with the above quotation and its translation. Due to my scarce competence in German, I have been unable to read his monumental contribution to the field, a fact that I deeply regret. My only knowledge of Sahle's reflections come via personal communications, and extracts of his work published by Franz Fischer (2010; 2013), which means a great work of scholarship remains out of reach to those who lacks essential linguistic skills. I can only hope that one day Sahle contributions will be available also to non-German speakers.

tried to analyse the benefits of printed editions with respect to digital ones.[5] These defences are the results of an unsettling caused by the new medium and the fear that its impact on scholarship will be on a great scale. Indeed the use of computers in editing is actually having a big impact on the very foundation of the editorial work, so in this sense such fear is entirely justified. However, such changes are not necessarily all bad. Textual scholarship has been shaped by the print environment in profound ways, in some cases dictating the research outcomes and the type of investigations undertaken. The advent of a new medium has allowed editors to detach themselves from the medium itself (they have 'transmedialized' their work, to use Sahle's terminology), reflecting on the heuristics of their work, understanding better than ever the print medium: it takes the detachment provided by another medium to understand the previous one.

Nevertheless there are serious drawbacks with digital editions, and more generally digital products and scholarship, which should not be overlooked in the wave of enthusiasm caused by the possibilities opened by the use of computers. There are misunderstandings as well: regarding the faithfulness of objectivity of digital representations, for instance. In spite of the 'exaggerated verisimilitude' of digital images (Deegan and Sutherland, 2009), or perhaps because of it, one should never forget that computers only provide those representations that humans have brought into being. Interpretation and critical analysis pervade the whole exercise, as maintained by Paul Eggert (2009a, p. 68).

> If in years to come historians see this as a rudimentary first stage of humanities computing, they will surely trace the origins of a second stage to the growing appreciation of the problematic involved in representing pre-existing handwritten or printed documents in electronic form. The turning point will, I believe, prove to have been the recognition that every electronic representation of text is an interpretation.

The other main issue with digital artefacts is their stability and longevity. Digital is fragile, ephemeral, and mutable, all characteristics that are ill-suited for a medium used to convey scholarship. This fact, beside the obvious problems of the risk of losing the hard work put into the creation of a digital resource, is also a strong contributing factor to the general mistrust toward digital products by academics.[6] Digital resources are useful, often quite heavily employed within one's research, yet they are quite often not cited as source of information, whenever possible their print counterpart is preferred. Scholars in the Humanities do not willingly admit to using the Internet in their research, and yet its use is pervasive and ubiquitous. The problem is that resources and editions on the Internet

[5] Cf. respectively chapters 9 and 6.

[6] Cf. chapter 8, in particular section 8.8 'Trusting Digital Editions? Peer Review and evaluation of digital editions'.

look less scholarly, less serious and academic than printed ones. They risk 'coming and going', or changing overnight, all of which gives them a bad reputation in academic circles. This fact has consequences on the career developments of people involved in the production of digital products as well; this problem is aggravated by the fact that digital resources are highly collaborative enterprises, where credit is usually distributed among dozens of people, contrasting therefore with the main framework for the production of scholarship in the Humanities, namely the single scholar working on a monograph. Other major issues concern sustainability, high production and maintenance costs, and the definition of an appropriate level of technical skills required to produce such products.

Yet, it seems impossible to imagine a future when the scholars will cease to use and produce digital resources; therefore either the scholarly community engages with the digital framework, or the key decisions and models will be established elsewhere (by publishing houses and software developers, for instance), and imposed top-down. This book intends therefore to promote the empowerment of the (digital) editors, a sort of call to arms to become protagonists of digital textual scholarship, setting agendas and goals dictated by research purposes and not by the possibilities offered by the medium.

One theme that runs throughout the whole book is the centrality that is given to models and modelling. Since McCarty strongly argued for modelling as the defining character of Digital Humanities (McCarty, 2005), the scholarly community that identifies itself using the label of Digital Humanities has tried to investigate the value and the heuristics powered and enabled by modelling. Modelling is interpreted here as the key methodological structure of digital editing; in doing this, this book also inscribes digital editing within Digital Humanities. In the same way that textual scholarship is an interdisciplinary component of the Humanities (one edits texts in literature, history, philosophy, classics and so on), digital textual scholarship represents its counterpart within the Digital Humanities. Yet the intended audience of this book is not digital humanists, but editors, and not necessarily digital ones. The pathway outlined by the book examines the editorial work and the editorial domain in the light of the changes introduced by computers, assessing elements of continuity with respect to more traditional approaches as well as analysing innovative methods and tendencies.

Chapter one takes into consideration traditional and emerging editorial frameworks. After a quick overview of the most established editorial theories, such as stemmatics, copy-text, social editing, New Philology and genetic editing, the chapter introduces born-digital editorial practices, starting with phylogenetics. Phylogenetics epitomizes some of the characteristics of digital scholarship in general and Digital Humanities in particular because it not only implies the use of computational methodology for the establishment of the critical text, but it is also deeply interdisciplinary, using as it does theories and algorithms developed in biogenetics. The results of such an approach are not entirely accepted by the community of textual scholars; but, as in other

cases, it is the journey that matters more than the destination: phylogenetics may not be the sole solution, but it has opened the way to further experimentations and into questioning the role of computers in textual scholarship. The chapter continues with the presentation of social editions, where users are invited to play an active role in the life of the edited text, producing new content either by transcribing primary sources or by annotating them. Crowdsourcing has taken a central role in many projects in the digital humanities, aimed at both generating great quantities of data, unachievable by project teams alone, and at engaging the general public with research processes. Yet it is questionable whether crowdsourcing can be adopted fruitfully as a method for the production of high quality editions, even when it takes the shape of academic-sourcing. This book proposes then a classification of different types of crowdsourcing and an evaluation of the opportunity of using such methods in the various stages of the editorial work. Next chapter one considers the so-called paradigmatic edition, namely an edition that contains more text than it shows and actually has many outputs, each of them presenting a different combination of the available fragments of texts. From this to the interactive and gamified editions the step is short. Here highly dynamic interfaces that engage the user in a pleasant and ludic environment are examined, and their potential as well as their drawbacks are scrutinized: is a pleasant experience scholarly enough? Classification is a concern that pervades digital scholarship, and in fact a tumultuous development needs to be counterbalanced by a discrete analysis; the ergodic edition method of classification proposed by Vanhoutte (2010) fulfils this need.

Chapters 2 and 3 are concerned with the elaboration of conceptual models. Modelling is the activity at the heart of Digital Humanities, as computers are totally dependant on models to function. The use of computational methodologies in editing requires then the modelling of the realities of editorial practices: documents, authors, texts, works, readers, editors, editions. All have to be reconsidered from an abstract point of view, in the search for their ontological status as well as their reciprocate relationship. This is the focus of Chapter 2. The text is claimed to be a model itself, generated by the modelling activity of the user-reader from one or many documents. From this perspective, the edition can equally be considered as a model of a document (in the case of documentary editions) or of a work (in the case of critical editions). The chapter then considers other existing models and compares them; in particular the FRBR model (and its evolution FRBRoo) as well as the pluralistic text theory developed by Patrick Sahle. Chapter 3 continues with modelling by taking on text transmission and the theory of variations. In this chapter classic theories of communication and linguistics are adapted to examine textual transmission: how is the text separated from its support? How are variants generated? And how are these events formally described? The separation of text and document is then tackled from another point of view, namely the theory of transcription, a topic that has generated quite a few contributions in the past few years, in particular based on the work of a group of scholars: Michael Sperberg-McQueen, Yves Marcoux, Claus Huitfledt and Paul Caton. Their analyses investigate the mechanisms according to which a transcriber

sees a sequence of signs on a document, mentally analyses them into discrete units and produces another set of signs which are meant to correspond to the original sequence, at some level of abstraction. Other recent contributions on the analysis and classifications of script acts are also reported in this context. In both sections editing and transcribing are seen as the agency operating the separation and abstraction of texts from documents. The attention to documents and to documentary editing has characterized the digital editorial scene for the past few years, reversing a century-long tendency that has seen documentary editing as a minor form of editorial endeavour, from both a quantitative and qualitative point of view. The rest of the chapter analyses the reasons behind the uptake of documentary editing in the digital environment as well as its consequences from a practical and theoretical point of view.

Chapter 4 remains focused on documentary editions, examining their alleged requirement of being faithful and objective representation of the documents from which they derive. The chapter firstly considers a classification of epistemic virtues analysed by Daston and Galison (2010), seen as the guiding factor behind the production of photographic atlases of scientific specimens. The researchers examine 'virtues' such as true-to-nature, objectivity (mechanical and structural), and trained judgement, which in turn are used here to examine the attitudes of different editorial frameworks with respect to the general epistemic category of objectivity. The claim is that there is no such thing as objectivity in editing and, furthermore, that it is not helpful in evaluating editorial practice. Methodological rigour does not mean an objective representation of the object of study, which is both unattainable and non productive. Instead accountability should be pursued, namely by providing exhaustive, punctual documentation of the editorial work, a task for which markup seems mostly suitable. Text encoding and markup in fact provide an excellent way to make explicit editorial activities such as regularization and normalization, among others. The claim that texts without markup are texts without interpretation is here denied: a text without markup is a text that hides interpretation behind writing conventions, as texts do not exist outside the dialectic between editors and documents. Then the analysis investigates the role and the consideration of digital facsimiles as a way of representing manuscripts and their role in editing. Digital facsimiles have had a mixed reception from scholars: from the unconditioned enthusiasm of some scholars considering them as an objective, unmediated, better-than-life experience, to the open hostility manifested by others, convinced that only the real artifact is able to provide an understanding of handcrafted documents. The chapter provides a different framework to evaluate them, namely to consider them as models of the documents they reproduce, therefore to be understood as simplifications able to support some research tasks more than others.

Chapter 5 is concerned with the work and the workflow of editing in the light of the introduction of digital technologies. Firstly, the theme of modelling is revisited, this time in order to

model the editions that have to be created and therefore retrospectively to model the editorial work required to achieve such a model. Then the diverse attitudes of editors toward the computer are discussed, namely those that consider computers as a research assistant able to relieve the editor of mechanical tasks, and those that see the editor engaging with computational methodology and embracing its transformative potentials. The latter approach allows for the study of the evolution of the editor into an encoder and for an assessment of the technological threshold that editors have to overcome in order to access editorial production tools. The (long) list of skills and requirements is inversely proportional to the possibility of scholars accessing the resources offered by specialized research centres, therefore making the production of digital editions an activity reserved for those who have the stamina and the interest in investing in their technical development, an activity that may take years, or who have considerable financial resources, namely holding a research grant. From this scene young researchers, PhD students not hired within a research project, are normally excluded, which questions not only the work and workflow of digital editing, but its ethics as well. In this context the role of the standard developed by the TEI (Text Encoding Initiative) is considered. The TEI in almost thirty years has not only supported research in digital editing, but has also contributed to shaping it. Yet its cultural hegemony has been contested by many, in particular those who consider the provision of embedded markup as an essential obstacle to interchange and progress in scholarship. Finally the transformation of the digital edition from the work of a single scholar into the effort of a 'village' is discussed with its many implications, including as the potential for the creation of innovative resources, and issues such as, for instance, the necessity of negotiating authorship and credits.

If chapter 5 studies the evolution of editors into encoders, chapter 6 studies their evolution into web publishers. The production of digital editions has seen the deterioration of the relationship that for centuries has bound editors and publishers together. The new medium has empowered editors, who felt that the constraints of publishing formats and market requirements were too limiting. However the relationship of the digital editions to the printed world is often more complex than may appear at first. In the past few years, in fact, hybrid editions have become more common, publishers are becoming more interested in the production of digital artefacts, and new forms of collaborations are being established. The chapter also examines new born-digital publication models. One of them is the 'work in progress' kind of publication, where texts partially edited are published on the Internet and changed and updated over a period of time, a fact that undermines the established rule that one cannot publish scholarly contributions (edition, monograph) until they become highly polished and stabilized. In contrast, the work-in-progress practice changes the relationship of the user with the resource, the latter being more of an evolving organic being than an established authoritarian site of knowledge. This also determines changes in other scholarly practices, such as, for instance the provision of post-publication reviews and citations. The chapter then takes into consideration the rapid establishment of the eBook market. However, eBooks could not be more different from the

model of digital publication offered by digital scholarly editions, as eBooks present themselves as a digital version of the printed book. This format has on its side the power of the big hardware producers as well as the major publishers and is imposing itself as the model for digital publications, narrowing as a consequence the space for digital scholarly editions.

Chapter 7 takes us into the realm of the readers and users of digital editions. In the digital environment readers seem to have been substituted by users, as digital resources seem objects to be used rather than texts to be read. Yet reading is a complex activity and is not always linear. An examination of different types of reading demonstrates how in reality scholarly editions are more like objects to be used than texts to be read even when they are in print. Their format does not invite the type of leisurely reading that we normally associate with works of literature: with all their commentary, apparatuses, introductions, appendices and indexes they are much more appropriate for non linear reading, targeted to an audience that is interested in the history of the text rather than in the text alone. Digital editions seem to fit this type of reading perfectly well. However, because of their free availability on the Web, their editors seem to think that they should also appeal to a much larger audience, an assumption which is not necessarily true. Here the claim is that pursuing a large readership is not necessarily a desirable outcome for research editions, as to make them accessible to the degree necessary to reach the larger public could compromise their fundamental scholarly function. On the other hand, the creation of simplified, public editions where textual facts and the history of the text are 'de-technicized' and presented in a playful way for the larger public could be pursued in parallel, perhaps offered for the tablet computer rather than on the web. eBooks and eReader devices have been profoundly shaped by the book, which represents a very powerful model, able to extend its influence over digital editions on the web as well. Yet the market seems eager to move away from such models, therefore the potential of tablet computers to deliver scholarship in ways that take advantage of the hapticity and playfulness of the device could be explored. The chapter then concludes with the examination of interface design principles.

User interfaces occupy an important role in the transmission of the editorial vision of the edited texts. Yet they present heavy drawbacks: because they rely on several layers of software and hardware to work, they are consequently extremely fragile and very hard to sustain in the long term. Sustainability and reliability of digital editions is therefore the topic of Chapter 8. These factors not only affect the longevity of the editions, but they also determine their trustworthiness: one can only rely on things that are perceived as stable and durable. The chapter examines several factors that may affect the trustworthiness of digital resources, offers guidance on how to build durable resources, and debates who should have responsibility for sustaining them, and at what price. And the price seems to be the interface: while the digital scholarly community has developed meaningful ways to support the longevity of the dataset, the same cannot be said about interfaces. This is however a very high price to

pay, given the fundamental role that interfaces play in delivering the message and the editorial interpretation. A few scholars have tried to offer solutions to what seems at the moment a crucial problem, but their proposals are not entirely convincing. The game is still open.

The last chapter finally attempts to give a terminological and substantial definition of what digital scholarly editions are or are supposed to be. By following the analysis Kenneth Price (2009) in his quest for the right definition for endeavours like the *Walt Whitman Archive*, the chapter examines proposals and rebuts ideas about the evaluation of digital resources. In fact such resources shake the core of the formats and the paradigm that have been inherited from print culture, making more and more difficult the possibility of describing and understanding the shape of the new. Yet we are only at the early stages of what digital editing will become. A few tendencies that have been observed allow us to draw some lines of the possible development of digital editions, which go in directions that seem to contradict the analysis proposed by this book. In particular the booming market of eBooks seems to suggest the imposition of an editorial model strongly shaped by the previous printed medium, where the text is presented in an even more rarefied version with respect to the printed book, disembodied as it is and deprived of any defining grounding in historical documents. Editors here are called to raise the stakes, instead of accepting such a simplistic publication model. Research and experimentation within the digital medium can only help to define the format of digital scholarship of the next generation, which should be as rich and as engaging as ever.

1. Traditional and Emerging Editorial Models

1.1 Traditional Editorial Models: an Overview

Under the umbrella of scholarly editing are a variety of practices and approaches, which are defined by the way they handle the evidence offered by primary sources that include the text to be edited, the way they reconcile contrasting readings from different sources, the role of the editor and the importance given to authorial intention. Distinctions are also made by the nature of the materials to be edited, their age, and the discipline within which they are edited. In particular, when a text is transmitted by more than one source, editors have to decide how to handle the contrasting readings that are witnessed by the document to hand. It is beyond the scope of this publication to give a detailed account of the main theories and approaches available to textual scholars, David Greetham has already done this in his excellent volume *Textual Scholarship* in 1992 and again in 2013, and, more synthetically, by Tanselle (1995). It will be sufficient here to sketch only the main theoretical positions which will become relevant to the discourse of the book.

The first aspect to notice is that various editorial frameworks are roughly determined by period, with stemmatics being more or less the framework for medieval and classical texts preserved in medieval manuscripts, the copy-text theory characterizes the edition of Anglo-American early modern print materials, and genetic criticism mostly for authorial drafts and contemporary authors. This distribution and periodization is nevertheless only theoretical, as it is also dependent on the country and the discipline of the scholar. For instance, the copy-text framework has found little favour outside the English speaking world, while the genetic criticism approach is more likely adopted by European scholars, with the exception of the UK. Biblical scholars traditionally use an eclectic approach, while historians will be more likely to use a documentary approach, but mostly in the US; and so on.

For the Anglo-American world from 1950, the most influential approach has been the so-called theory of 'copy-text' as defined by Walter Greg and refined (and radicalized) by Fredson Bowers and Thomas Tanselle.[7] This theory mainly applies to a situation in which we have an autograph manuscript of a given work as well as a later printed edition of the same; in such cases, the editor is invited to use the *substantive* readings coming from the later printed edition, combined with the *accidentals* (punctuation, spelling, capitalisation, etc.) of the autograph manuscript. The theoretical justification that lies behind this methodology is the pursuit of what is thought to have been the intention of the author: had authors been able to control the printing process in the same way they had controlled their own handwriting, surely they would have employed their own accidentals, as

[7] For an evaluation of the differences between American and British implementation of these theories, see Sutherland (2013).

is shown by autographic evidence. With the centrality given to authorial intention, which can only conjecturally be postualted, the 'copy-text' theory has attracted fierce criticism, mostly grounded in the famous essay by Roland Barthes *The Death of the Author* (1968). This argument has been used by many, in particular by Jerome McGann, who has offered in opposition to the copy-text his theory of the social text. According to this theory, 'the literary "text" is not solely the product of authorial intention, but the result of interventions by many agents (such as copyists, printers, publishers) and material processes (such as revision, adaptation, publication)' (Siemens *et al*., 2010). Contextually, McGann then claims the existence and the importance of the 'bibliographical codes' of a work beside its 'linguistic codes', that is, factors such as typesetting, layout, orthography, binding are to be considered together with the actual verbal content of any given text (McGann, 1991, p. 57), and therefore, as the author is unable to control every aspect of the 'bibliographic signifiers', 'the signifying process of the work become increasingly collaborative and socialized' (p. 58). According to this view, authorship is shaped by external conditions and therefore any attempt to reconstruct an uninfluenced authorial intention is misguided. Around the same time Donald F. McKenzie (1986) was elaborating his 'sociology of text', focusing on typography, format, binding, and layout which are influenced by the social context in which the authors wrote, but which also become part of the authorial intention. For both scholars, then, a work cannot be determined by words alone; but a critical approach that combines readings from several sources will, by definition, privilege linguistic codes over bibliographic codes, with an inevitable loss of meaning.

The scenario that is at the heart of McGann's and McKenzie's theoretical framework, as well as of copy-text theory, presuppose the print industry, excluding therefore all pre-Gutenberg texts and cases for which we have no manuscript or final printed edition.[8] For these cases, stemmatics remains the main theoretical framework, even if such approach has been harshly contested for more than one hundred years. Stemmatics is also known as the 'Lachamnian method' and it aims at reconstructing the 'original' work using significant errors made by scribes as a guide to reconstructing the genealogical relationships (organized in the *stemma codicum*) among the surviving manuscript sources (the witnesses). However, the legitimacy of reconstructing texts from multiple sources has been contested since the beginning of the twentieth century, starting with the critical works of Joseph Bédier in 1928 (his theory is known as the 'best-text' or the '*bon manuscrit*'), and has more recently taken new force, inspired by the work of Bernard Cerquiglini (1989, English translation 1999), according to which 'instability (variance) is a fundamental feature of chirographically transmitted literature: variation is what the medieval text is "about"' (Driscoll, 2010). Cerquiglini's work has

[8] It is worth mentioning, however, that the copy-text theory has been adopted even in the absence of authorial preserved manuscripts, when one printed edition could be, for some reason, considered closely controlled by the author and for which it is possible to suppose the accidentals are ascribable to the author.

deeply influenced the birth of a movement that goes under the name of 'new' or 'material philology', and has found many points in common with the theory of the social text proposed by McKenzie and McGann with their focus on the 'bibliographical code'.

One of the main distinctions between editorial practices, however, is determined by the object of editing; editors can edit texts preserved by only one source, hence editing 'documents' or, to use Tanselle's terminology (1989), 'texts of documents', or editors can try to provide an edited text combining readings coming from multiple sources, hence editing a 'texts of works'.[9] While the latter is normally called 'critical editing', the former is mostly known as 'non-critical' or 'documentary' editing (Greetham, 1994, pp. 347-51). For this editorial typology, however, the label 'documentary editing' seems to be preferred to 'non-critical' editing[10] as it better describes the object of the endeavour and it avoids being defined by a negative statement, which, whatever the intentions of the scholar using it, always sounds slightly detrimental. In both editorial practices, the purpose of the editor is to propose to the readers the best text they are capable of assembling given the documentary evidence they have to deal with and their theoretical approach to textual editing.

In spite of these strong commonalities, these two forms of editing are normally kept very distinct, in particular in the Anglo-American tradition, in the sense that they have been given different theoretical and practical frameworks, and typically have also been associated with different disciplines. In fact the edition given a single witness has been linked to historical evidences, while the multiple witnesses scenario relates to literary texts. Of course, this scenario is hugely simplistic. As we will see, however, it has been at the heart of a fierce debate between literary scholars and historians since the late 1970s. The distinction between the two approaches, implicit for a long time, was 'sanctioned' in 1978 with the establishment of the US Association for Documentary Editing (ADE) (Eggert, 2009a). But, at the same time that the ADE was being established, Thomas Tanselle, coming from a literary background, attacked the standards of documentary editing practised by historians, comparing them with the greater rigour practised by literary documentary editing (Tanselle, 1978). The debate that followed had the twofold consequence of producing a better definition of the respective editorial practices and an interdisciplinary cross-evaluation of such practices, as well as an understanding of disciplinary boundaries (Kline and Perdue, 2008, pp. 19-22). This debate also demonstrated that documentary editing is not the exclusive realm of historians and that scholars from different backgrounds have their own opinion and practices on how this should be done. For instance, documentary editing represents the editorial model of choice for the so-called 'new' or 'material philology', a theoretical approach that has been developed by literary scholars,

[9] On the concept of 'work', see Chapter 2.

[10] And as such it figures as the title of Kline and Perdue (2008). Cf. also Chapter 3.

with the conviction that '[l]iterary works do not exist independently of their material embodiments, and the physical form of the text is an integral part of its meaning' (Driscoll, 2010), and therefore any reconstruction of a text obtained by the combination of multiple sources is likely to miss some essential component of the text.[11] The philosophical differences between the two approaches are well summarized by Thomas Tanselle (1995b):

> These two kinds of editions imply two approaches to the past: the documentary method focuses on past moments as seen in the physical objects that survive from these moments; the critical approach recognizes that surviving documents may be misleading guides to the past and may therefore require altering or supplementing through the creative action of informed individuals. (Tanselle, 1995b, Online, § 9)

Digital editing has challenged these boundaries as well as many other assumptions of traditional editing. This fact has been more or less recognized by Tanselle himself, who, in the article mentioned above, concedes that digital editions ("hypertexts" in his 1995 terminology) are able to present both approaches as complementary as in fact they are they are in common editorial practice, and although critical editing is seen by Tanselle as the most valuable editorial activity, he still recognizes that editors have to consider texts of documents before they can reconstruct critical texts: 'critical editing is the natural complement to the presentation of documentary texts, and hypertext admirably supports both activities' (1995b, Online, § 22). But in spite this early intuition, the fact that digital editing supports different types of scholarly editing at once seems to have fallen out of the discourse of textual criticism.

For authorial draft manuscripts, the French school of *critique genetique* represents the most famous editorial approach. This methodology aims at investigating the writing and authoring processes as witnessed by the working manuscripts or *brouillons*, which can be organized and studied within a *dossier génétique* (genetic dossier), a term which should be preferred possibly to the alternative of *avant-text* (pre-text), because the research of genetic criticism is focused on the act of writing more than on the production of texts (Grésillon, 1994, pp. 108-109). This approach also allows one to define common stages within the authoring process, from planning to sketching, fleshing, drafting, revising, correcting, and so on. Although this method has produced remarkable theoretical reflections, its editorial products have not been immune from criticism: in fact genetic editions have been reproached for being very hard to read because of the heavy deployment of diacritics (the markup) used to signify the extreme complexities of the phenomenology of the written

[11] But on New Philology, see again Chapter 3.

page (Hay 1995; Grésillion 1994, pp. 195-202). On the other hand, genetic editions may be considered not the most representative outcomes of genetic criticism, but only a by-product, necessary to deploy the genetic criticism itself, namely the critical analysis of the authoring process. Nevertheless, the impasse produced by trying to represent the process of authoring in the bi-dimensional space of the printed page may be properly addressed by moving these editions to a more versatile and flexible space: cyberspace.[12]

1.2 Digital Editing, Digital Editions

Can all of the above-mentioned methodologies be pursued digitally or does the digital medium necessarily provide a new theoretical framework? Or, in other words, is digital simply a new medium for 'old' methods or is it an entirely new methodology? The question is left open for the moment, but we will see that the impact of computational technologies in editing requires us to evaluate the editorial work from different points of view with respect to the traditional ones, where not only the editorial approach needs to be considered, but also the functionalities, the typologies, the goals and the targets of the digital product.

But firstly: what is digital editing? Is it the use of digital tools in the production of an edition? Or is it the publication of an edition in digital format? One could indeed use digital methodologies and tools to produce either a print or a digital edition, or both (O'Donnell, 2008). However it is worth asking here if the type of digital tools employed in the production of such editions are also to be used as a discriminating factor. One could ask if the use of a word processor is enough to qualify one's edition as digital, or if something more advanced and sophisticated, perhaps developed specifically for a particular editorial work is implied by the label 'digital editing'. Here the distinction introduced by Rehbein (2010) between 'classical thinking' and 'digital thinking' may be useful. In his analysis, Rehbein qualifies the so-called classical thinking as output-driven: the purpose of the digital elaboration is to produce something that looks good on the page; while digital thinking is qualified by being input and user-driven, where the purpose is to produce something that captures the nature of the content elaborated. If we maintain this distinction, then editions produced with the digital support only of a word processor cannot qualify as 'digital editions' since they are the product of 'classical thinking' and their purpose it to look good on the page. Yet, the distinction may lie more deeply inside the type of workflow adopted, the type of output produced, and the ideas of the editors.

So far, computer-assisted scholarly editing has been aimed mainly at simplifying the traditional editorial work and at preparing traditional types of edition which are only enhanced by

[12] This in fact is the suggestion that comes from Grésillon (1994) and Lebrave (2008).

being offered as hypertexts or provided with some searching and indexing facilities, with the idea that editors can use the computer to speed up their editorial workflow without really changing the nature of their work (Shillingsburg, 1996).[13] A similar idea is also at the heart of what Andrea Bozzi defines as 'computational philology' (Bozzi, 2006), which is focussed on the design of user-friendly tools for assisting the editorial work, which, in turn, remains unvaried in its founding principles. However, this presupposition has been questioned by a number of scholars (McLoghlin, 2010; Vanhoutte 2010; and Sutherland and Pierazzo 2012, for instance), and will be discussed in detail in Chapter 5. In fact, whether fully appreciated by its practitioners or not, digital editing has ultimately challenged the concept that it is possible to use computers to alleviate mechanical work without questioning it, resulting in a series of innovative editorial practices disguised by traditional labels. As a matter of fact, the very distinction between editorial work and editorial product is now being challenged, as the same tools used by editors for their editorial work are themselves being made available to the users: this is what Greg Crane defines as 'ePhilology' (2008), where the editorial function is shared by editors and users, and where text is decomposed, analyzed and processed algorithmically. In the past few years there have been a few attempts to categorize digital scholarly editions (Vanhoutte, 2010, and Siemens *et al*., 2012b, for instance), which have tried to understand what is so special about them. Such attempts, which will be fully analyzed successively, have mostly concentrated on functionalities of the finished products and, more marginally, on the methods of production, leaving in the shade the question whether there is any new epistemological surplus implied in emerging digital editorial models. It is, however, the argument of this book that there is such a thing as an epistemological and theoretical significance in digital editing, which will be presented in this chapter and those to follow.

When not only the working methods are digital, but also the outcome (or one of the outcomes) of such a work, then we can normally talk of a digital edition, a term that, in spite of its apparent 'normality' has been discussed and questioned repeatedly. According to Patrick Sahle we should, for instance, distinguish between 'digital editions' and 'digitized edition', whereas the former is guided by principles that transcend the idea of the page as an informing paradigm, while the latter is a simple remediation of a printed edition (2008, Online):

> Digital scholarly editions are not just scholarly editions in digital media. I distinguish between digital and digitized. A digitized print edition is not a 'digital edition' in the strict sense used here. A digital edition can not be printed without a loss of information and/or functionality. The digital edition is guided by a different paradigm. If the paradigm of an edition is limited

[13] Edward Vanhoutte considers the computer-as-research-assistant as the first historical phase in the development of electronic editions (Vanhoutte, 2010, pp. 121-3).

to the two-dimensional space of the 'page' and to typographic means of information representation, than it's not a digital edition.

However, the label itself 'digital edition' is not an uncontroversial acquisition in its own right, and the past few years have seen many efforts to find a defining label. In his contribution of 2009 Kenneth Price asked 'Edition, project, database, archive, thematic research collection: What's in a name?' arguing that the way we call things determines the way we represent and conceptualize them, emphasizing the fact that the 'stuff' we produce has hardly any real counterpart in the pre-digital world. And, indeed, what digital scholarly editions are and how we should name them should not be taken for granted, as the discussion in Chapter 9 reveals.

1.3 Digital Editorial Models

1.3.1 Phylogenetic

The first demonstration that the use of computers can indeed produce new theoretical approaches is demonstrated by the development of what has been called a 'born-digital' editorial model, namely the phylogenetic/cladistics approach to textual editing. The principle behind this theory is that text variation as present in scribal manuscripts behaves similarly to the genetic mutation in molecules of DNA in living creatures; therefore it is possible to apply phylogenetic algorithms and methodology to group witnesses around their similarities and then to reconstruct texts (Macé and Baret, 2006). This approach represents perhaps the most radical attempt of taking advantage of the processing capabilities offered by modern computers, as not only does it use complex algorithms to group variants, but it also requires an automatic collation to produce such variants, therefore giving strong incentives to the development of collational algorithms and tools.[14] Furthermore, it exemplifies the interdisciplinary vocation of Digital Humanities, as it uses and adapts methods borrowed from computer science and biology, applying them to a typical humanities question. However, this methodology has not been immune from criticism and is still not completely accepted by the scholarly community (Hanna 2000, Cartlidge 2001, but see an answer to them in Windram *et al*., 2008). The main problem with this approach is represented by the difficulty of determining how, and if at all, to

[14] For collation tools see the latest development offered by Juxta Commons (http://juxtacommons.org/) and CollateX (http://collatex.net/, cf. Dekker and Middell, 2011). Pior than these, the most influential collational tool is offered by COLLATE, developed by Peter Robinson (for which cf. Robinsons 1989), but cf. Gilbert (1973) for a much earlier approach.

weigh variant readings (Andrews and Macé, 2013): what is indeed the effect of not distinguishing orthographic and linguistic variations from variants and errors? The answer is not obvious, as contrasting opinions among scholars demonstrate (Salemans 2000; Spencer *et al*., 2004).

From the early attempts of using phylogenetic and cladistics methods in textual editing, a new set of methods is now being developed, which goes under the name of digital stemmatology or 'computationally assisted stemmatology'. This method consists of the statistical and computational analysis of the results of automatic collation – by far more rich and precise than any manual collation – and the consequent building of *stemmata*, with the purpose of analysing textual transmission and, more occasionally, constructing critical texts, with promising results and various degrees of success and reliability (Andrews and Macé, 2013).

1.3.2 Social Editing and Social Editions

If the first born-digital editorial model has been developed by borrowing methods from sciences, the second is born from the evolution of the digital society. The Web, and in particular its user-centred functionalities that go under the label of Web 2.0, have put the emphasis on the user and user-generated content. This movement, fuelled by the ubiquity of social networks, could not but impact digital editions as well as any other endeavour that has made the web its home, and has created the premise for the birth of the so-called collaborative and social edition. Collaborative editions are not a novelty: large editorial endeavours have always been a familiar concept in editing, whether in the field of literary studies (such as the full edition of the works of Jonathan Swift or Ben Jonson published by Cambridge University Press, to give a couple of recent examples), or historical studies (such as the edition of the Thomas Jefferson papers), or philosophy and science (such as the editions of Isaac Newton and of Charles Darwin). In all the aforementioned editorial endeavours, a team of selected scholars, experts in the field, lead by one or more chief editors, share principles, tools and resources in the preparation of a (usually) multi-volume edition (Bree, 2009). Furthermore collaboration is also at the core of most digital editions (cf. chapter 5). But the social edition is something different. The idea here is to offer the text to the community not only for contributions such as annotation, comments, and translations but also for the editing of existing texts or the addition of new texts. The meaning of 'social' is then twofold: social in the way a text is edited (social editing) or in the way it is released to the public ('social edition'). The community that produces the text or enriches it can be controlled (that is, only selected users, where credentials and competence has been checked by some sort of editorial board) or uncontrolled and open to anybody. The term 'social edition' in the contest of digital editing has been mostly connected to the editon of the Devonshire Manuscript produced by Siemens *et al*. (2012c), where the '*et al*.' label has to be interpreted in the

broadest possible way, since it is offered to for contribution to the any interested user. One of the aims of this edition is to change the role of the editor:

> Our goal, through community engagement via Wikibooks, Twitter, blogs, and Drupal-based social media space, is to use existing social media tools to change the role of the scholarly editor from the sole authority on the text to a facilitator who brings traditional and citizen scholars into collaboration through ongoing editorial conversation. (Siemens *et al*. (2012c), Online, *A Note on this Edition*)

This editorial model is grounded in Jerome McGann and Donald F. McKenzie's theories of the social text which undermines the author as sole authority, focusing instead on the circumstances of the production of the work (Siemens *et al*. 2012b, p. 450); from here to undermining the *editor* as the sole authority for publishing the work is an easy step. McGann presents the social dimension of the text as a key component not only of the production, but also of the reception: the text and its meaning are made by the readers, each of them generating the 'plural text' (McGann, 2004, p. 206). This plurality of reading undermines the role of the privileged reading provided by the editor and opens the edition to the contribution of communities of practice (Siemens *et al*., 2012a). This change in the role of the editor has been advocated by Greg Crane (among others) in a contribution from 2010: 'We need to shift from lone editorials and monumental editions to editors … who coordinate contributions from many sources and oversee living editions'. Of the same opinion is Thaller (2004) who theorizes that the advent of digital editions will produce the blurring of the roles of editors and the reader. However, Bordalejo (2013), amongst others, heavily criticizes this approach, lamenting the lack of definition on what exactly the transformation of the role of the editor from authority to 'facilitator' means, and, more importantly, considers that the opening up of the editorial process to the public as a demonstration of 'a lack of understanding of textual critical theory and of the process of production of scholarly editions', hinting that the availability of social software (its 'affordance', to use Gibson's 1979 concept of 'opportunity for action given by the existence of a new tool') is likely to have 'cloude[d] our judgement'.

It is undoubtable that the 'social edition' as embodied by the edition of the Devonshire Manuscript raises a certain number of question, in particular on the nature and quality of the work provided by the online, potentially anonymous, contributors. What Siemens *et al*. (2012b), Crane (2010) and, in some measure, Robinson (2010) seem to have in mind is not a situation where anybody contributes to the edition, but a model where a community of practitioners who share interests and competences is collaboratively engaged in producing an edition. However, it is not clear how this community is to be defined or selected. Siemens *et al*. (2012b, p. 450) define 'community of practice' as 'a group that forms around a particular interest, where individual members participate in collaborative activities of various kinds', and they give as an example of such communities the Text

Encoding Initiative and the Digital Humanities at large. But when it comes to scholarly editing, the situation is more complicated. Editing is not an easy task, and a keen interest on a topic is not a sufficient condition for making an editor out of a reader. So, how is it possible to ensure that the contributors will actually have the necessary competences? How to attract qualified editors and how to reward them in order to convince them to contribute?

One possible approach is the one adopted by the *Suda On Line* (SOL) project. The project aims at crowdsourcing the translations into English of a Byzantine encyclopaedia written in the 10th century CE. The project has an editorial board, the members of which have scholarly qualifications in Ancient Greek. Such editors verify and edit the translations that are provided by the volunteers that produce the actual translations (Mahoney, 2009 §9). Anne Mahoney has compared this model to peer review, with the original translator and the editor both being credited for the entry. In summer 2014 it was announced that, after 16 years of works, SOL had reached a remarkable milestone: all of the 31,000 plus entries have been translated; the new goal of the project is now to improve the translation and the commentary (SOL, Online, *About SOL*).

Another model is provided by the *Early English Laws* project and it is based on a selection of the suitable editors. The project provides a technical and intellectual framework for the preparation and publication of a critical edition of English laws up to the Magna Carta. An individual or a team can apply to edit one or more of the available texts: their credentials, competences and the validity of their editorial projects are evaluated by the editorial board of the project, and if these are accepted, then the applicants are invited to edit the text.[15] The proposal guidelines distinguish between 'experienced editors', defined as 'those who have published editions of medieval texts', and 'new editors … who have not edited any text or texts of length or complexity'. The latter group is required to provide a fuller proposal, including a full *curriculum vitae* in order to be allowed to edit any of the available texts. It is arguable that this model represents a social edition in the sense proposed by Siemens *et al*. (2012b), as this can be better interpreted as an attempt to transpose a traditional editorial model (like the one adopted by the Jonathan Swift edition mentioned above, for instance) in a digital framework, the novelty of which is the provision of uniform editorial tools and a centralized web-based support.

The social edition as proposed by Siemens *et al*. (2012b and 2102c) has certainly caused a bit of a stir in the editorial community, and rightly so: scholarly editing is a specialized activity, performed by experts who are likely to have to spend years in acquiring the necessary skills and competences to perform it, and years to actually edit any given text, and therefore it is legitimate to ask how this activity will be performed by non-specialists and why specialists will take their very

[15] See the project's website: http://www.earlyenglishlaws.ac.uk/edit/

limited time to contribute to an edition outside the typical career-oriented pattern. Yet, the recognition of the inevitable sociality of the authoring and editorial processes cannot be denied, and this is the reason why it is worth devoting some attention to the issues raised by the edition of the Devonshire Manuscript. Even if editors and editorial projects leaders may prefer not to push the involvement of the readers as far as completely opening up the editorial process, many have embedded some sort of crowdsourcing and public engagement in their work, and many more are likely to do it in the future. The reasons behind this evolution are complex and varied, but the main two are probably on the one hand the pragmatic need to convince funders to support an editorial project (as candidly admitted by Causer *et al*. 2012, p. 131: 'no funding body would ever provide a grant for mere transcription alone'), and on the other hand a general movement toward a democratization of culture. It is therefore essential to devote some attention to this phenomenon that is likely to deeply impact on editorial practices of the future, which is new in its current digital embodiment, but has deep roots in scholarly practice, in particular in the English speaking world.

1.3.3 Crowdsourcing and Editing

The social component of the Devonshire Manuscript edition is accomplished through a form a crowdsourcing which could be defined as the use of social media tools to enable users to contribute to a specific project or platform with new content.[16] Wikipedia is one of the world's most famous crowdsourced projects, to which anyone can contribute, if they feel they have the knowledge and they wish to do it. When this project was launched there was a fair bit of scepticism about the possibility of the 'crowd' to produce valuable content in an unsupervised way; however 'Wikipedia proved the skeptics wrong' (Rockwell, 2012, p. 139). But is crowdsourcing really an option for digital scholarly editions? And if so, in which form? Both the *Suda Online* and the *Early English Law* projects try to combine some sort of crowdsourcing with the essential requirement of guaranteeing a quality sufficient to be acceptable by a scholarly community, one by checking and editing the materials produced by contributors, the other by pre-selecting the type of contributors as well as peer reviewing the results. An increasingly degree of rising attention devoted to public engagement by public and private funding bodies, combined with the possibilities offered by the Web 2.0, have given crowdsourcing a great deal of attention, prompting investigation into its role in scholarly editing (Rockwell, 2012). However, crowdsourcing, in particular the form applied in the Humanities, is not an invention of the age of the Internet: the name may well be (Howe, 2009), but the concept is definitely not. In fact, the origin of the *Oxford English Dictionary* can be described as one of the earliest crowdsourcing project in history: from 1879 the editors of the OED, and in particular the

[16] The first to analyze the emerging phenomenon of crowdsourcing seems to be Howe (2009).

general editor, James Murray, issued a public call appealing to readers in all the English speaking world to source new words and provide examples of usage (Murrey, 2001). For the following seventy years, hundreds of people contributed in various ways to the first issue of the OED which was published in 1928.

Indeed, crowdsourcing has been very successfully employed in a number of scientific, academic endeavours, but it still seems to struggle to affirm itself for scholarly editions, the reasons for which are mentioned above: editing is no simple task and requires a large number of competences. One of the most successful crowdsourcing initiatives is represented by the *Zooniverse* 'the internet's largest, most popular and most successful citizen science projects'[17], which features, among others, projects like *Old Weather* and *Ancient Lives* which have involved about 600,000 people each in transcribing manuscripts.[18] On the other side, the *Transcribe Bentham* project had more mixed results in the number of people involved and transcriptions achieved (Causer *et al*., 2012, p. 132).

> [C]rowdsourcing, particularly crowdsourcing manuscript transcription, is plainly not necessarily a cheap, quick, or easy solution … Editors of the Papers of Abraham Lincoln found that, when experimenting with using non-academic transcribers, they spent more time correcting errors than they would have had they carried out the transcription themselves.

These difficulties should not come as a surprise to anybody who has ever undertaken any manuscript transcription, which is a complex task even to an experienced editor. Furthermore, transcription is only the preliminary stage of any editorial work, whether based on a single document or on multiple witnesses. The work of analysing readings, making sense of them and, if necessary, correcting them seems still solidly in the hands of trained editors. From these considerations, it seems clear that crowdsourcing will not put editors out of work, nor make it necessarily cheaper or faster, but public engagement, the opening of the editorial endeavours to a community of practice and to new theoretical positions make crowdsourcing still very tempting, if carefully controlled. In fact, once the quality of transcriptions provided by volunteers has been ascertained by members of the editorial board (or by a self-moderating community like the one that congregates around the *Suda On Line* project), such transcriptions can be used for the editorial work, to speed up its pace, as the editors of the *Transcribe Bentham* project have recognized (Causer *et al*., 2012, p. 133). But how does crowdsourcing work and why do people get involved (or not)? Again it is the *Transcribe Bentham* project that provides some useful indication of how to engage people and how to keep them engaged, which is, perhaps, even more challenging. Manuscripts and primary sources are not everybody's cup

[17] Cf. <https://www.zooniverse.org/about>.

[18] Cf. <http://www.oldweather.org/> and <http://ancientlives.org/>, respectively; both of these projects have involved about 600,000 users in the transcription of manuscripts.

of tea: some documents may be fascinating and able to capture people's interest, others much less so, and the more complex these materials are, the fewer people they are likely to capture. For crowdsourcing to work well, the tasks required of the users need to be intuitive, easy to understand, manageable in a small portion of time, and to give immediate reward. In this way not only do people become engaged in a project, but they also will remain so. Rockwell (2012, p. 147) writes about 'small autonomous tasks' and quotes Howe's 'Keep it simple and Break it Down' rule (Howe, 2009, p. 285). However, this may be a challenge if the task is to transcribe a manuscript, which could be long and complex. In these cases, the stress could be put on the reward that is given to the contributors for their effort. For many, the public recognition of their work and the competition with the other contributors is the most valuable reward. The mechanism of crediting both the translator and the editor is one of the main concerns of the SOL project, in the hope to attract scholars, and in particular young researchers as contributors, who can use such creditation in their CV. Others, on the other hand, may find other, more playful types of reward more engaging; for them solutions borrowed from computer games may be more appropriate.

Crowdsourcing is an inclusive label that contains a wide range of phenomena and types of activity; an exhaustive classification of these phenomena in an academic context has been attempted only recently (Dunn and Hedges, 2012), but an analysis of its viability, relevance and significance within an editorial context has not yet been considered in full.[19] Without any pretension of being exhaustive, I propose here such an evaluation, where crowdsourcing activities concerning the edition and publications of texts are classified according to five parameters: Context, Participants, Tasks, Quality Control and Role in the Project.

1. *Context*: Crowdsourcing projects can be hosted and supported by:
    a. *Universities, research centres and cultural heritage institutions*, such as Libraries and Museums. This is the case with some of the projects mentioned above (*Transcribe Bentham* is hosted and supported by UCL, for instance), and of the National Library of Australia's *Historic Newspaper Digitisation* Project, where users have been asked to correct OCRed articles from historical newspapers.
    b. *Non-governmental Organisations and other private initiatives*: This applies, for instance, to *Project Gutenberg*, which began 1971 from the vision of its founder, Michael Hart and continued since thanks to donations.[20]

[19] A partial analysis is in Dunn and Hedges (2012, pp. 24-7).

[20] http://www.gutenberg.org/wiki/Gutenberg:The_History_and_Philosophy_of_Project_Gutenberg_by_Michael_Hart

   c. *Commercial*: This applies, for instance, to Google, which uses the ReCAPTCHA service, asking users to enter words seen in distorted text images onscreen, a part of which comes from unreadable passages of digitized books, thus helping the correction of the output of the OCR process while protecting websites from attacks from internet robots (the so-called 'bots').

2. *Participants*: the main question here is how they are recruited and which skills they should possess to be allowed to contribute. Some projects issue open calls, for which anybody can enrol and contribute, with no particular skill being required other than five minutes to spare; other projects require their contributors to possess specific skills which are checked before the user is allowed to become involved in the work. The former is the case for the *Historic Newspaper Digitisation* or *Project Gutenberg*, the latter for the *Early English Laws* project. Many projects locate themselves between these two categories, closer to one end or the other. In the SOL project, for instance, users are assumed to read and understand Greek, but their competence is verified by the quality of their translations, although to register as an editor, users are expected to declare their competences which are then checked by the editorial board.
3. *Tasks*: While there are many tasks that users may be requested to perform (See Dunn and Hedges, 2012, p. 21), the tasks that are relevant for textual scholarship could be one or more of: [21]
   a. *Transcribing* manuscripts or other primary sources, like in the case of *Transcribe Bentham*.
   b. *Translating*, as in the case of SOL.
   c. *Editing*, which is requested by the *Early English Laws* project.
   d. *Commenting* and *Annotating*: as in the case of the *Pynchon Wiki* (Schroeder and den Besten 2008)
   e. *Correcting*: this is the case, for instance, for the National Library of Australia's project seen above and of the *Project Gutenberg*, where users not only contribute by uploading new material, but also by taking on proofreading texts in the archive.
   f. *Answering* specific questions: this is the case for the *Friedberg Genizah Project*, for instance, which uses the project's Facebook page to ask specific questions of its followers about, for example, a particular reading of a passage, or if the hand of two different fragments is the same, and so on.[22]
4. *Quality control*: the quality of the work produced by the contributors can be assessed by professional staff hired for that purpose (e.g. *Transcribe Bentham*), or could be assured by the community itself, with super-contributors whose controlling roles are gained "on the field" by

[21] Please note that neither task *c.* or *f.* are listed by Dunn and Hedges 2012, p. 21

[22] See <http://www.facebook.com/pages/Genizah-Research-Unit/125835514119021>

becoming major contributors (e.g. Wikipedia), or because of their qualifications (e.g. SOL), or both.

5. *Role in the project*: for some projects the provision of crowdsourced material can be the final aim, as with the *Project Gutenberg* or the *Historical Newspapers Digitization* project, or it could be a product that will be used in other stages of the project. The transcriptions produced within *Transcribe Bentham*, for instance, serve a double purpose: they represent the main outcome of the project as, once their quality has be ascertained, they feed into UCL's digital repository, but they are also meant to be used for the edition of *The Collected Works of Jeremy Bentham* in preparation since 1958 (Causer *et al*., 2012, p. 133).

This loose classification may prove helpful for evaluating the content, the scholarly level and the reliability of the editions available on the Internet. Indeed the difficulties in assessing the methodology and the quality of digital editions seem to be one of the main factors that prevent a more widespread success of digital editions by the scholarly community. Back in 1996 Peter Shillingsburg lamented that 'only one tenth of 1 percent of the available texts on the Internet were reliable for scholarly work – 99.9 percent of the texts were who knows what' (p. 138). The situation has changed little since Shillingsburg's discomforting assessment, and while the digital awareness of the scholarly community has certainly risen since, it has not risen as much as would have been required for the academic community to be confident in its evaluation of the quality of digital resources. This fact is demonstrated, for instance, by the proliferation of guidelines on how to evaluate digital resources issued by various universities, by the high number of attendees to the MLA's workshops on evaluating digital resources[23] and by the publication of their *Guidelines for Evaluating Work in Digital Humanities and Digital Media* in April 2012.[24] Chapter 7 will return to this topic at length.

1.4 The Source-and-the-Output Model: Paradigmatic Editions

But it is not only Web 2.0 and crowdsourcing that challenge traditional editorial models: even when digital editions simply aim at transferring established editorial models into a digital framework, changes take place which are perhaps less evident but equally disrupting and which prelude the establishment of new editorial models.

---

[23] 2011: *The Good Web: Workshop in Teaching Your Students How to Evaluate Web Resources* <http://www.mla.org/web_wkshp>; 2012: *Evaluating Digital Work for Tenure and Promotion* <http://www.mla.org/resources/documents/rep_it/dig_eval>.

[24] See the website <http://www.mla.org/guidelines_evaluation_digital/>.

Let us start our examination by looking at editions based on only one witness, that is to say documentary editions. Among them we find several types of edition: type facsimile or ultra-diplomatic, diplomatic, semi-diplomatic, and reading editions – which are distinguished mostly by different levels of editorial intervention on the text as transmitted by a primary source.[25] They are also differentiated by the layout and the amount of editorial comment and annotation. However, in the digital world such differentiation may prove to be meaningless or need more effective ways to be ascertained. One of the reasons for this lies in the fact that most digital scholarly editions are based on markup, expressed in XML for the greatest part; when adopting this technology editors (encoders) add tags to their text in order to annotate some feature or another (a scribal error, a tear on the page, the presence of a date in a calendar other than the Gregorian, and so on). These tags are then interpreted by a script (a computer programme) and transformed into a conventional format for display (by adding '*sic!*', or a square bracket containing dots or letters, and so on). This fact then implies that we need to distinguish the data model, where the information is added (the source), from the publication where the information is displayed (the output). In this framework, type-facsimile or diplomatic or indeed editions that present the text in a clear uncluttered way (the so-called reading editions), each represents only one of the possible outputs of the source. These possibilities are offered by digital editions based on markup languages, and in particular those based on TEI, which are scholarly oriented. The key factor is that many features (or facts, as we will see) of a primary source may be, and often are, encoded in many ways. For instance, it is possible to transcribe the text with both abbreviations and expansions; with typos and without; with unconventional spellings and with regularized ones at the same time; furthermore it is possible to record features that one may not want to display all the time or at all, but use them to generate statics or indexes. These overabundant, paradigmatic types of encoding, which have become more or less the norm in digital editions, imply that we should redefine what we mean by labels such as diplomatic or reading edition. For instance, according to its classic definition, a diplomatic edition is a published version of a transcription which reproduces as many of the characteristics of the original document as the medium permits or as the project requires. If we consider a digital edition based on text encoding such as the one described just above, this will normally be composed of many parts, including at a minimum the following:

- A source file (or files) which will contain the transcribed text. At the current time XML-TEI markup represents the best practice of text encoding in the Humanities, but other formats are (and will continue to be) possible as well. One of the reasons why the TEI model is particularly effective is because it enables the encoding and transcription of several alternatives for the same segment allowing, for example, the encoding of abbreviated forms along with the expanded ones, or of erroneous passages and their corrections.

[25] For an historical overview on different types of edition see Greetham, 1994, pp. 347-372.

- A set of scripts able to *read* the source file and transform it into some particular 'view' of the text. The scripts therefore express which combination of encoded features composes the particular views. These scripts will typically be XSLT or XQuery if the source file is XML.
- One or more outputs, each of which will represent one of the aforementioned views or, to use a more established word, one 'edition'. This may include a diplomatic output if the editors have decided that this is a purpose of the edition.
- One or more styling files to refine the display of the edition. For output in HTML, this role will typically be filled by some combination of CSS and JavaScript.

So, which of these components is, say, the diplomatic edition? If the transcription is recorded as an XML source file, such file will record as many characteristics as considered relevant by the project team, however it will not *look* like the original document in any meaningful way since the words on the page[26] are interspersed with tags, and so it is not a diplomatic edition. The output displays the transcribed text as closely as possible to the original document, which is then perhaps a more suitable candidate for the label of 'diplomatic edition', but the knowledge (the scholarship) of how to produce such an edition from the source is stored in the scripts. Let us consider the following example:

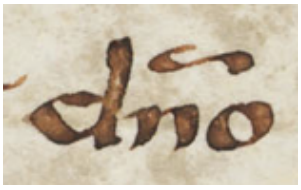


FIGURE 1.1: THE ABBREVIATION OF 'DOMINO'[27]

The sequence on the page can be transcribed using XML-TEI as follows:

```
d<choice><am>ñ</am><ex>omin</ex></choice>o
```

The encoding shows that the sign of abbreviation composed by the letter 'n' topped by a curly line can be expanded as the sequence 'omin' to form the word 'domino'. In turn such encoding can be then output as follows:

[26] Or the words that according to the editor are on the page, as argued in chapter 4.

[27] St. Gallen, Stiftsbibliothek, Cod. Sang. 189, p. 76 – *Eucherii Instructiones; Isidori liber differentiarum; S. Hieronymus super Daniel* (<http://www.e-codices.unifr.ch/en/list/one/csg/0189>). Reproduced with permission of the Stiftbibliothek St.Gallen.

- ‘dño (diplomatic output)
- ‘d[omin]o *or* d*omin*o or d(omin)o (semi-diplomatic output)
- domino (reading output)

One could also wish to display the word ‘domino’ capitalized in its reading output; to do so, more code will have to be added:

```
<choice><orig>d</orig><reg>D</reg></choice><choice><am>ñ</am><ex>omin</ex></choice>o
```

Such encoding will allow us to output ‘Domino’ or ‘D*omin*o’ or any combination of capitalized/un-capitalized and abbreviated/un-abbreviated form, according to the editor’s purposes.

One might use a culinary metaphor here: the source contains the ingredients, the scripts contain the recipe, and the output represents the cooked dish. This metaphor works, but only to a certain level of abstraction because while the source contains the ingredients to produce the edition, it can in many circumstances be considered an edition by itself: for a TEI expert, for instance, an XML source file will be even more eloquent than the output and in fact the importance of *published* source files as scholarly products has been widely recognized (Bodard and Garcés, 2009; Cummings, 2009). The question can be pushed even further: in the case of many digital editions, the diplomatic output is interactive and can be modified by the users,[28] meaning that *diplomatic* is only one of the possible, unstable, states of the output; we could therefore even conclude that these are not diplomatic editions at all, but that they are something else. In such cases one could also say that while the source file contains a diplomatic edition, it may also contains other editions – semi-diplomatic, reading, critical, interpretative, or others – all of which are contained simultaneously *in posse* within the same source file, with each of these editions requiring the application of different sets of scripts and styling to be realized. This proteiform, cumulative nature of digital editions challenges on the one hand the concept of a single edited text. Which is ‘the text’? The redundant, paradigmatically encoded text or one of the many possible outputs? Similarly challenged on the other hand is the commonly accepted way of distinguishing different editorial products, namely the different levels of editorial intervention. The handling of features such as corrections of scribal errors, preservation of original spellings, layout and paginations are normally used to distinguish diplomatic from semi-diplomatic from reading editions,

[28] See for instance The Van Gogh letters, where the user can switch on and off features of the original layout, or Langscape, where it is possible to convert the wynn letter into ‘w’, or *The Book of Margery Kempe*, where the user can (although not very intuitively) switch on and off abbreviations and marginalia (http://english.selu.edu/humanitiesonline/kempe/index.php).

with each of them characterized by an increasing weight of editorial intervention. For digital editions based on text encoding the editorial interventions are all present at once in the source, they are just not displayed at once.

Similar considerations could be made for digital editions based on many witnesses. In the case of the edition of Willem Elsschot's *Achter de Schermen* (edited by Peter de Bruijn, Vincent Neyt and Dirk Van Hulle, 2007), we are offered the XML source for each witness, as well as the scripts to generate multiple outputs for each of them, together with the capability of generating a combined critical and reading edition – actually, *many* critical editions, each of them based on a different copy-text, chosen interactively by the reader, some of which may even never be generated in practice. This model is ground-breaking for many reasons: not only does it challenge the definition of 'critical edition' as the text is generated on demand by the user, but it also blurs the distinctions seen at the beginning of the chapter between documentary and critical editions, as both approaches are equally represented. This fact is clearly present in the reflection of Edward Vanhoutte (2007), who, discussing the outcomes of an edition he prepared in 2000 in collaboration with Marcel De Smedt, declares (pp. 162-3):

> According to the Anglo-American theorists … this could have been called an electronic archive or a documentary edition. However, we included not only one but two critical reading texts. For their constitution we … introduced the German theory that only allows justified correction of manifest mistakes in the edited text. But we divert from this tradition by not including an electronic version of a traditional apparatus variorum … Instead, each paragraph of all the versions of all the texts included in the edition … could be consulted on their own or in any combination with what we called 'the orientation text'.

This and other digital editions based on the separation of source and output, while still aiming at presenting traditional editorial formats, are clearly already something else, in search of a theoretical definition. I have called them 'paradigmatic editions' (Pierazzo, 2014b), as the choices offered to the reader are collocated in the paradigmatic axis, the axis of variation, but another name could be 'generous editions', in the sense that they offer more to their readers than a simple one-dimensional text.

Print-based editorial models are defined by various parameters: the way the edited text relates to the primary sources that transmit the text, the way variant readings are handled and combined in the final text, whether variant readings are or are not kept separated from errors, if a stemma *codicum* is used to represent the relationships among the witnesses or is also used to constitute the edited text, and so on. Each editorial theory requires the editor to follow one route only. But digital editions allow

you the luxury of not choosing, allow to have your cake and eat it, as there is potentially and theoretically no limits to the amount of information it is possible to include: it will be then left to the user/reader to choose which editorial principle to follow, instantiating the preferred edition. There are of course limits, as an article of 2011 clearly states (Pierazzo, 2011), which are not only necessary, but also desirable. However, these limits can (should) be determined by scholarly purposes, more than by the limitations of the publishing technology.

However rich and generous, paradigmatic editions are not 'definitive editions' in the sense that they may contain many editions, but they will never contain all editions. Both Sperberg-McQueen (2009) and Shillingsburg (2006) warn that there are infinite sets of facts about the work being edited, and that no matter how inclusive a digital edition might be, one could always have done something else, as in Robert Frost's poem 'The Road Not Taken' (cited by Shillingsburg, 2006, p. 155).

1.5 Raising the Stakes: Interactive Facsimiles and Gamification

In 2011 the TEI has introduced a new approach to text transcription that proceeds page-by-page and zone-by-zone instead of textual unit-by-textual unit (Pierazzo, 2014a). This innovation was introduced with the twofold purpose of providing an encoding framework for the transcription of complex documents such as, for instance, heavily corrected drafts,[29] and for the production of genetic editions. In order to test these new possibilities, such encoding has been used in the creation of a prototype featuring a few openings of a Notebook of Marcel Proust (André and Pierazzo, 2013; Pierazzo 2014a).[30] Besides testing the encoding, the biggest innovation in this prototype was the research into new ways in which the text could be delivered to the reader. The prototype then explored innovative approaches to both editing and delivering scholarly editions, in ways that tried to avoid the bi-dimensional 'page paradigm' (Sahle, 2008). Traditional diplomatic and even ultra-diplomatic editions aim at presenting texts in a format that tries to mimic the layout of the manuscript page as much as the publishing medium allows (Pierazzo, 2011). And while this type of edition presents many advantages, it is unable to represent the dynamic of the writing process. Furthermore, even when published on the web, such editions are also normally presented side by side with the digital facsimile of the page, but this representation, inherited from print, has proven unsatisfactory for more than one reason: first, it creates an alternative new space which tries to mimic the original without ever being able to reproduce it in full, giving rise to all manner of frustration in attempting the unachievable goal of reproducing the exact layout, spacing and 'feeling' of the draft page (Sutherland and Pierazzo, 2011,

[29] But not only: see Chapter 3.

[30] The prototype itself is at < http://research.cch.kcl.ac.uk/proust_prototype/>

pp. 207-208). Second, it leaves to the user/reader the task of establishing the relationship between the transcribed and the inscribed text, a task which is tiring and uncomfortable, relying as it does on the rapid movement of the eyes from one area of the screen to another.[31] Finally, the side-by-side view is limited to present the source one page at a time, and not, for instance, by openings, given the constraint in width of the screen. However, such an approach, if applied to Proust's notebook, would have indeed misrepresented the documentary evidence which shows how Proust considered his writing space to be the opening as a whole, as he used to write only on the right side of the opening of his own notebooks and used the left side for additions, corrections and rewriting. Therefore, the page-by-page visualization that has become the standard for some types of digital editions was not an option; nor was it an option to present the facsimile of the opening alongside the transcription if either was to be at a readable size: the type of document leads us into moving the transcription *within* the facsimile. The integration of the transcription with the facsimile has already been attempted by, for instance, the edition of *Stirrings Still* included in the *Samuel Becket Digital Manuscript Project*'s website. This provides in fact a 'Zoom Topographic' view, where a draggable box reveals an ultra-diplomatic transcription as it is laid over the facsimile; a similar solution was explored also by Gabler (2007). Such visualisations represents a good attempt to integrate ('interpenetrate', to use Gabler's 2007 terminology) the edition and the facsimile, but they also present limitations: the integration is only partial and unstable, as it relies on the movement (or steadiness) of the mouse; the two layers are not aligned making the deciphering of the handwriting cumbersome; and finally they are not suitable for extended reading, any more than watching a movie through a keyhole would be. In order to address such issues and present a wider focus, the *Samuel Becket Digital Manuscript Project* offers several alternative visualizations, among which the "image/text" view, allowing for more extended reading and for the display of transcription over the facsimile in a more stable way, section by section.[32]

The Proust prototype-edition has been based on and embedded within the facsimile in a similar way, but avoiding the 'keyhole effect', as the text is displayed on-click in a stable way. Additionally, the prototype attempts to go a step further, as it is built around the idea of process, meaning that the zones that have been outlined and encoded have been assigned a relative order with

[31] Usability tests have demonstrated that users prefer concentrating on the left hand side of the screen only, a fact that, together with personal experience, suggests that the side-by-side layout may not be very effective (Nielsen 2010).

[32] See it at <http://www.beckettarchive.org/demo/MS-UoR-2934.htm?page=06&trans=basic&text=documentfacsimilestopo2&notes=on&metamarks=&type=linear&facs=2934-4r#2934-4r>[accessed 15 January 2014].

respect to their presumed sequence of writing and reading.[33] This information has then been used to create an interactive, accessible interface which tries to present the user with a representation of the writing process, not just the end product.

Draft manuscripts are complex, data-rich objects which require the long patient work of scholars to be made 'consumable' by people other than the specialist. Because of the complexity offered by these materials, they have rarely taken a central role in scholarship beyond that of their editors. The documents' complexity, combined with the inaccessibility of editions, has often discouraged even the bravest of readers. The printed book model has proven to be unsuitable for the task of presenting such material in an accessible way to scholars other than editors. The Proust prototype, even at this very limited stage, represents a step forward in that direction, borrowing as it does some ideas from computer games. The easy, intuitive interactivity makes the user experience enjoyable and fun, and suggests that even the most complex of cultural objects can be made enjoyable without compromising the level of scholarship. The idea of using game mechanics in a non-game context for solving problems is called 'gamification' and is a well-known approach in interface design, eLearning and advertising (Zichermann and Cunningham 2011). The principle is to take the user experience or problem to be solved and break it down into small tasks; once these are achieved, then users are rewarded somehow and are invited to proceed to the next, more complex level. Most gamified environments appeal to competitiveness, building communities of users and making them 'play' against each other. A light version of these principles has already been implemented successfully in academic projects, particularly in crowdsourcing such as the *Old Weather* project which uses the crowd to transcribe ships logs in order to study the weather,[34] or *What's on the Menu*, which aims to enable the study of food prices and eating habits by inviting people to transcribe old menus of restaurants.[35] Can (or should) a scholarly digital edition do the same? The idea behind the digital representation embodied by the Proust prototype offers two aspects of this: on the one hand it presents scholars with the possibility of exploring draft manuscripts in a much deeper and more accessible way, by representing the draft manuscript as the custodian of the authoring process, by

[33] Although possible in principle, we have not attempted to record the absolute timing of authoring but only the relative sequence of writing campaigns. This is due to both practical and theoretical reasons: the former because it is not yet clear exactly when Proust wrote in this specific notebook, and the latter because, in the impossibility of assigning all variants to a specific absolute time and therefore to clearly distinguish all layers of writing, this could lead to the reconstructions of texts that never existed (Pierazzo 2009, 185-186).

[34] The project is supported by a large consortium of partners including the UK Met Office, the National Archives and the University of Oxford.

[35] *What's on the Menu* is a project of the New York Public Library.

exploring the codicological aspects of the material object. On the other hand, the end result could also be enjoyable for people with non-specialist knowledge, opening new perspectives on the access of advanced cultural content to the wider public, a consideration that must increasingly find a place a place in the agenda of textual scholars.

1.6 Functional Analysis of Digital Scholarly Editions: an Edition Is What an Edition Does

What is a digital scholarly edition? What does it include? For whom is it produced and why? Foundational as these questions seem to be, they are neither easy nor have a straightforward answer, as it will become clear. In fact, these questions will be only addressed in chapter 9; it seems more fruitful to start this book by asking instead what a scholarly digital editions does or is supposed to do. While books all behave more or less in the same way, with some notable exceptions,[36] digital editions are all different, and the things they can do diverge considerably; this is because while the print technology has developed standard editorial templates and formats, more or less constrained by the physical boundaries of pages and bindings, the digital medium is still experimenting with the available possibilities and is not limited by space. This perhaps represents one of the most compelling reasons why we should produce digital editions to begin with, namely because it is possible to do new things, things that were not possible before. The fact that digital editions are also more accessible is also one of the reasons why scholars decide to produce a digital edition rather than a printed one, but only a secondary one.

Many scholars have undertaken such a functional analysis, with very divergent results. According to Patrick Sahle (2008)'s definition mentioned above, we should distinguish between 'digital' and 'digitized' editions: an edition cannot be called truly digital if by printing it we do not have a loss of functionality. What is probably meant here is that one can possibly print some parts of a digital edition (a reading version of the established text, for instance), but the printed version will not be able to retain the interactivity and the functionality offered by the online version. So, for instance, a PDF version as well as any other remediated version of a printed edition will not qualify as a digital edition, but only as digitized. Of course most digitized editions will provide some extra functionality which is added to the mere digital representation of the text, some only by the fact that they are in digital format: provided that they are machine-readable (and so not scanned images), they can be searched full-text and, provided they are online, they are available from many computers at once.

[36] Cf. for instance Goytisolo (2012) or Sardou *et al.* (2009), where there are and half-sheet fold-outs inserted within the main pagination throughout the books and liftable and foldable patches.

However, these low-level digital add-ons alone are not enough to provide the type of functionalities that Sahle considers essential for the definition of a digital edition (Sahle, 2012).

Sahle does not specify which functionalities are then necessary, nor how to group or examine then; instead this analysis has been undertaken by Edward Vanhoutte. In a contribution from 2010 he examines the characteristics of digital editions and elaborates a functional definition of what he calls the *ergodic edition*,[37] the name and the model of which are based on a now classic essay written by Espen Aarseth in 1997. In his work Aarseth examines the characteristics of cyber texts and cyber literature, and concludes that their most important, defining characteristic is the engagement and the interactivity required by the user/reader and her/his effort in achieving the ultimate goal of reading the text. Vanhoutte claims the same for digital scholarly editions, and applies to edited texts the same analytic grid developed by Aarseth (1997), which encompass the following variables: Dynamics, Determinability, Transiency, Perspective, Access, Linking and User function. Vanhoutte's grid is reproduced in Table 1.

Variable	**Possible value**
Dynamics	Static, intratextonic dynamics, textonic dynamics
Determinability	Determinable, indeterminable
Transiency	Transient, intransient
Perspective	Permanent, impermanent
Access	Random, controlled
Linking	Explicit, conditional, none
User function	Explorative, configurative, interpretative, textonic

TABLE 1.1: FUNCTIONAL GRID (VANHOUTTE, 2010, P. 142)

In this model, the intrinsic dynamics of the text are given by the interaction and tension of *textons* 'strings as they exist in the text' and *scriptons* 'strings as they appear to the readers' (Aarseth, 1997, p. 62). This distinction between *textons* and *scriptons* is of fundamental importance to understanding digital editions and reflects a first defining characteristic of many of them, namely the distinction between what is visible to the final user and what is data stored in the computer's memory;

[37] *Ergodic* is a term first used in this context by Aarseth, 1997, deriving from the Greek ἔργον – work and ὁδός – path, namely something ergodic implies a difficult laborious pathway.

in other words, the distinction between the raw source data and the rendered output.[38] Dynamicity and interactivity is only possible when 'what you see' is the result of some process applied to some data stored and codified in some way that is retrieved and displayed 'on demand' by the user. As for cyber literature, for some digital editions it may be possible that the user will never be presented with all the material potentially available: what the user will be able to see depends on her/his choices, the paths taken through the digital edition. According to this model, while for printed editions all the data available is presented to the reader in the order and format established by the editor and the publisher (bearing in mind that this does not necessarily imply that the reading has to follow that order), for many digital editions it is the user who chooses the order and the format of the text. This possibility is contemplated by the variable 'user function' in Vanhoutte's model; such function is 'configurative', when the user chooses or partially chooses the *scriptons*, and 'explorative' when the users choose which path to follow. According to the same model, considering the variable of dynamics, an edition can be 'static' if all the *scriptons* are constant; 'intertextonic dynamic', when the number of *textons* is invariable while the *scriptons* can change; and 'textonic dynamic', when also the *textons* can change. The last case implies an editorial model where the editors or the users can add more content to the edition after its publication. If the editors can add more material, we could call it an open-ended edition or a work in progress; if this possibility of adding more material is offered to the users, then we are in the realm of the so-called 'social edition' discussed above. The variable of 'determinability' establishes whether 'for every *scripton* its adjacent *scriptons* are always the same', and here Vanhoutte declares how for scholarly editions determinability appears to be a *conditio sine qua non*. However, it is not only possible to envisage an edition based on game models which, as Vanhoutte suggests, would challenge such a condition, but also editions where the users are given the opportunity of choosing their own variant readings among the ones offered by the tradition, henceforth questioning the editorial authority. This possibility (as well as having been discussed above) is contemplated by Tanselle as early as 1995, who also underlines how this is really not a novelty, but it is at the base of providing readers with a critical apparatus:

> Readers can of course make their own choices among variants, using whatever bases of judgment they wish, just as they have always been able to do with other forms of apparatus – though with hypertext they can more easily produce a smooth reading text of their own construction (1995b, par. 22).

[38] Such distinction is at the base of what we have defined the 'paradigmatic' edition seen before, namely the distinction between source and output.

Editions are then 'transient' if the passage of time causes the *scriptons* to change, and Vanhoutte imagines editions with a *play mode* where users can watch editions like they watch movies, for instance. This is the defining function of the prototype built around a few pages of a Marcel Proust's notebook (André and Pierazzo, 2012); this prototype (for which see above) uses interactive animation to showcase the author's writing process as well as the development of narrative sequences. It is worth stressing how crucial is transiency (or time-based animation) in the case of genetic editions; this function could well represent the key factor in making digital editions worth pursuing for genetic editing (Pierazzo, 2009).

Still following the analysis of Vanhoutte's model, as for perspective, editions are 'personal' if the users play a role in what happens to the text; otherwise, they are 'impersonal.' Moving on, editions have a 'random access' if their *scriptons* are always available to the users at any given time, otherwise they have 'controlled access'. Finally, editions can offer 'explicit links' to follow, 'conditional links' (that is, links available only in certain conditions) or none.

This model is indeed fascinating and disorienting at once: fascinating because, for the first time, we witness the attempt to explain what is special with digital editions using a 'born digital' model, instead of trying to map the new, digital product to the old, analogic, editorial model. It is also useful as it gives the reviewer tools to describe features of digital editions that are new and therefore difficult to account for using traditional analytical grids. But it is also disorienting because the model is different from anything seen before in textual scholarship, and it uses an unfamiliar terminology. Perhaps more familiar, if more simplified (a 'rough typology', according to its authors), is the functional categorisation proposed by Siemens *et al*. (2012a), based in turn on that by Lancashire (1989) and Faulhaber (1991). Again the model is a functional one, based on what the edition does and what the users can do with it. In the model of Siemens *et al*. (2012a), the first category is called 'dynamic text' and consists of 'a combination of a properly encoded electronic text with text-retrieval and analysis software' (Lancashire 1989). In this case the dynamicity is offered by the software that accompanies the edition and which allow for a non-linear exploration of the text.

The second category is the 'hypertextual edition' (Faulhaber, 1991), where the edited text is provided with ancillary material, textual apparatus, indexes etc. all available in a hypertextual way to the user. The two models are not mutually exclusive and, in fact, Siemens *et al*. (2012a) propose the third category of the 'dynamic edition' where the text is offered along side tools for dynamic exploration and a constellation of ancillary material accessible via hyperlinks. 'Dynamic edition' is also the label of choice of Rehbein (2009).

Although useful, both functional analyses proposed by Vanhoutte (2010) and Siemens *et al*. (2012a) seem to be missing an evaluation of the appropriateness and effectiveness of the

functionalities in the term of the scholarly purpose of the edition, and the capability of these functionalities to respond to the research questions. In other words: what do the explorative tools and interactivities available to the users mean for the understanding of the edited text? Do they represent valuable, scholarly tools or only accessories? And are the functionalities provided sufficient to fulfil the purpose of the edition? Finally, does the provision of tools, functionalities, texts produced digitally require a different editorial model? The following chapters will address these questions, starting with what has been considered the basis of any digital process: modelling.

2. Modelling (Digital) Texts

In the beginning was the model

2.1 Modelling and the Edition

What is an edition? And what relation does an edition establishes with the text it edits and with the work the text represents? This chapter will examine familiar entities such as texts, documents, works and editions from a point of view that is different and unfamiliar to many, the one of conceptual modelling. This attempt is not a total novelty by any means: Peter Shillingsburg has adopted a similar approach since 1991, and many others have come close to it (Caton, 2013; Gabler, 2012; Eggert, 2009; Tanselle, 1989; Sperberg-McQueen, 2009; Robinson 2009 and 2013; but also Barthes, 1968, and Ong, 1975, to name a few), but they have called it, not incorrectly, a 'theory'. In fact a conceptual model is a way of (re)presenting a theory, therefore a conceptual model of bibliographic and editorial entities reflects a theory of the same. The reason here to insist on the act of modelling more than on the theoretical architecture is given by 'the fundamental dependence of any computing system on an explicit, delimited conception of the world or "model" of it' (McCarty, 2005, p. 21): whatever theory is at the base, in order to compute editions a conceptual model will be required. Nevertheless, the theory 'behind the scenes' will also become evident as the argument progresses. The modelling of texts has been tackled before from different points of view: sociological, cultural, psycholinguistic, philosophical, historical and computational. Why, then, is it necessary to return to the same topic once again? Because our knowledge and understanding of the object to be modelled evolves; in our case, not only the knowledge, but also the object itself (digital scholarly editing) is also rapidly evolving, a fact that encourages the deployment of more theoretical efforts. But, even more importantly, because the model will adopt a different point of view: the editorial one.

Analysis and modelling are rightly considered to be at the heart of Digital Humanities; furthermore they are fundamental components of scholarly editing, whether digital or not. As a matter of fact, modelling is at the core of any critical and epistemological activity, as these require the selection of particular facts (among all observable ones) relative to a particular object or set of objects in order to be analyzed and understood. In the humanities we are not used to the words 'model' and 'modelling' used in this particular sense, but nevertheless the elaboration of a particular model (or concept) of an object of study is the preliminary, implicit activity that lies at the basis of most analytical endeavours.

Willard McCarty was the first scholar to have drawn the attention of the Humanities scholarly community to the importance of modelling and its connections with similar activities in other disciplines. In his fundamental volume of 2005 he forged the definitions of 'modelling' and of 'model' that have become foundational columns within the digital humanities: 'By "modelling" I

mean *the heuristic process of constructing and manipulating models*; a 'model' I take to be either *a representation of something for purposes of study*, or *a design for realizing something new*' (McCarty 2005, p. 24). He then adds that models 'are by nature simplified and therefore fictional or idealized representations' of real objects. The process of modelling is also an epistemological process: by modelling an object, an activity we come to know this object, and this activity.

In a contribution of 2009 (but reporting on discussions and debates at least a decade older) Michael Sperberg-McQueen declared that there are three *things* to consider when we edit a text: '1. There is an infinite set of *facts* related to the work being edited. 2. Any edition records a *selection* from the observable and the recoverable portions of this infinite set of facts. 3. Each edition provides some specific *presentation* of its selection' (2009, p. 31). By saying that an edition records a selection of facts related to the work, Sperberg-McQueen implicitly defines the edition as a model of the work. In the same way as a map can be considered an idealized model of the earth for a specific purpose, an edition of a work can be considered as a model of such a work, as the edition represents a selection of the infinite features of the work, organized according to particular point of view, the one of the editor. In fact, the act of selecting features from the uninterrupted continuity of the reality is the defining act of modelling. If there is an infinite set of facts to be observed within the work to be edited, a selection, and hence a simplification, will allow us to avoid creating a model which aspires to equal the object to be studied and therefore becomes useless: a model that is identical to the object it studies is not a model, it is the object; assuming that such a thing is possible, an edition that, for instance, is perfectly identical to the primary source it aims to represent in all respects, logical and material, is no edition, it is the primary source. The necessity of selecting in order to build useful models is a fact that has been demonstrated by the fathers of the theory of models:

> Let the model approach asymptotically the complexity of the original situation. It will tend to become identical with that original system. As a limit it will become that system itself. [...] Lewis Carroll fully expressed this notion in an episode in *Sylvie and Bruno*, when he showed that the only completely satisfactory map to scale of a given country was that country itself. (Rosenblueth and Wiener, 1945, p. 320).

So, we must model, we must select, and we must establish limits to the selection process; and these limits represent the boundaries within which the hermeneutic process can develop. The challenge is therefore to shape a model which is adequate to the scholarly purpose for which it has been created.

As Willard McCarty states (2005, p. 26), modelling is an iterative, learning process, the result of which is a better knowledge of the object to be studied; it requires a sound knowledge of the object (in our case, texts, documents and works) to be modelled and of the processing to be applied. In

addition, modelling is an on-going process: as knowledge of the object deepens within the context of the project in general and the editorial work in particular, so it is to be expected that the initial model will need to be refined many times over. McCarty explains that this refinement of the model, and therefore its effectiveness, reveals itself through failures: 'modelling succeeds intellectually when it results in failure, either directly within the model itself or indirectly through ideas it shows to be inadequate. This failure, in the sense of expectations violated, is fundamental to modelling (2005, p. 26)'. In editing, failures of the model may concern two main issues:

1. Wrong section of the features to include.
2. Misunderstanding of the relationships between these features.

In the first case, the editor may realize after having undertaken a good amount of work that she/he is including so many details in the edition that her/his progress is too slow, or alternatively that not enough features have been included, preventing the intended analysis. These features may be the transcription of minor orthographic variations or the inclusion in the apparatus of minor variants. The choice of what to include is a crucial one, and depends mainly on the purpose of the edition. In the second case, the editor may misunderstand correlations between readings and a particular physical feature of the supporting document, for instance, or the relations among witnesses, or again the succession of versions.

If an edition is a model, we may ask what is the process that is deployed in the creation of such model, what does it model, how does it happen that texts and editions are conjured as models, by whom and why. McCarty (2005) distinguishes two types of modelling, *modelling-of* and *modelling-for*, where the former is descriptive simplification of a complex object, and the latter is a design to guiding into something new. The editorial work and the edition which represents its outcome are complex objects, the results of many, often implicit, operations; complex objects are ill-suited to be handled by computational systems, which in turn require that each step of a complex process is explicitly and algorithmically traceable. In order to be able to develop tools and software to support the editorial work, it is fundamentally important to produce a conceptual and functional *model-of* it. In this way the *model-of* the edition will enable the creation of a *model-for* the edition.

In the context of digital editing, by modelling we mean at least two types of conceptualization: the one that tries to organize entities such as texts, documents, works, along with their relationships and how they have happened to come into being, and the analytical process of establishing the kind and purpose for the production of a new edition, its implied community of users and what features best represent their various needs. This chapter and the next focus on the former, while the latter will be examined in Chapter 5.

2.2 Modelling Texts and Documents

The first step towards producing a conceptual model of editions, texts, documents and works and their relationships requires a definition of such entities. However, a definition – any definition – of key terminology is bound to provoke disagreements and discussions, a possibility even more probable in the case of a highly theorized and cultural-rich concept such as the one of 'text'. Huitfeldt goes to the heart of the problems when he observes that defining text is a 'fruitless' operation:

> If possible, I would have liked to start out by suggesting possible answers to questions such as: What is a text? What is the ontological status of the text? What is the epistemological status of texts? However, I have come to think that these questions do not represent a fruitful first approach to our theme.
>
> The answer to the question what a text is depends on the context, methods and purpose of our investigations. (Huitfeldt, 1994, p. 235)

However in spite of the cultural overload that surrounds text – and possibly because of it – a definition is nevertheless necessary, in order to establish the theoretical framework of the following discussion.[39]

Let us start with the document, which constitute the primary sources of any editorial endeavour. **Documents** we take to be physical objects that contain some sort of inscribed information; therefore a book is a document, a leaf with some writing on it is a document, a stone is a document. More generally, a document is a physical object that has some text on it, or more formally, a Verbal Text Bearing Object, or VTBO. The definition willingly and knowingly leaves out non-verbal documents, as the object of the present research is to analyze and model written and verbal texts. All documents contain verbal texts as well as other things: images, graphs, musical notation, arrows, blotches, for instance, as well as including the 'bibliographical codes' discussed by McGann. The model proposed here does not exclude the non-verbal content of documents, quite the contrary, but it only concerns documents for which a verbal-content can be determined, since it is built to

[39] I acknowledge here the fundamental contribution of my colleague at the Department of Digital Humanities, King's College London, Geoffroy Noel. Geoffroy is a fine thinker and an expert in knowledge representation. The conceptual model presented here has been elaborated via lengthy discussions and a fair number of emails between the two of us. It is with his permission that I publish it here as part of this book. My gratitude goes to Geoffroy Noel not only for his invaluable contribution, but also for his great generosity.

explain the editorial work. A similar but more generalized definition is given by Huitfeldt and Sperberg-McQueen, who prefer to speak of 'marks' on a document, rather than of 'verbal text': 'By a document we understand an individual object containing marks. A mark is a perceptible feature of a document (normally something visible, e.g. a line in ink)' (2008, p. 297).

A document is defined by three sets of characteristics:

1. It has, or had in the past, some physical dimensions which can be measured (length, weight, number of leaves, numbers of bytes, so on).
2. It can, or could in the past, be found somewhere.
3. It has, or had in the past, some signs on it that can be recognized as words written in a language by a competent reader.

Documents do not need to exist as physical objects in the present, but they had to exist in time and space at some point.

Documents are handled by various agents, who perform with them various activities: storage, dusting, cataloguing, reading, and so on. For the purpose of this model, the type of agency that we are considering implies reading, therefore we could call the agent a **Reader**.

Documents have observable features or '**Facts'** (as maintained by Sperberg-McQueen, 2009) which are relevant for their readers; these facts have a meaning which is determined by the context in which they are found by the readers. Facts may be of quantitative nature (for instance the number and dimension of pages), but their meaning and their relevance is interpretational. Facts can be grouped into **Dimensions**, which are interpretative entities arbitrarily conjured by readers, according to an organizing principle, a point of view or a research purpose which makes such dimensions meaningful. For instance, what we could call a 'palaeographical dimension' reflects the interest in, among other things, the way a letter (part of the document) is written (fact). The way lines are marked on the writing surface could also be an element of the palaeographical dimension, or a so-called 'codicological dimension', the attribution to one or the other being the exclusive choice of the reader/researcher based on the purpose of the research undertaken. A fact could also not belong to any dimension, if researchers do not consider that fact relevant for their purposes. As dimensions potentially observable in a document are defined by the purpose of one's interest in the document, it is therefore impossible to draw a stable and complete list of such dimensions; however, for example they could include:

- **Linguistic dimension**: the language in which the text is written, its grammatical rules, and the way it has been realized.
- **Semantic dimension**: what the words mean and what they were intended to mean.

- **Graphematic/palaeographical dimension**: the letters that compose the words, their shape and succession, the type of script or typeface and the cultural implications of using one or another.
- **Literary dimension**: style, rhetorical features, genre, intertextuality, citations and allusions.
- **Genetic dimension**: by whom and when the document was filled with words, the revisions and the additions.
- **Artistic/iconic dimension**: how words and decoration look, their artistic quality, by whom they were created and why, what decoration represent and mean.
- **Codicological dimension**: how the words have progressively been added to the document and how the particular shape and structure of the document and its production have influenced the words, how the document has been made, and its significance.
- **Cultural dimension**: the value we attribute to it because someone special has written it, or it has been owned by someone special, or has played a role in a historical event; because of the scarcity or the value of its components, e.g. its role for religious devotion.

Another way of seeing these dimensions is to interpret them as channels of communication (or 'contacts'), each one with its own codes (a 'meaning' encoded as 'fact'), as they are made of signs which, in the intentions of their producers (the 'senders') were meant to communicate something (the messages) to someone else (the 'receivers'). This terminology will be resumed in Chapter 3 which models textual transmission as an act of communication; here it is enough to establish the connection between the multidimensionality of documents and that type of modelling.

These dimensions are only potentially available within a document, and until someone reads and inspects that particular document for some reason, the document itself has no particular meaning: it is an inert object with no particular significance. The place where meanings are conjured is the **Text** which comes into being when the document is read/analyzed by a reader: 'meanings are constructed in terms the reader controls or is influenced by' (Shillingsburg, 2006, p. 41).[40] Or, to go back to the communication model hinted at above, the document transmits its message (the text) when an interested receiver can be found (the reader).

The **Text**, then, is the meaning(s) that readers give to the subset of dimensions they derive from a document and that they consider interesting for their purpose. As a consequence, texts are immaterial and interpretative. Many people can read from the same document and understand slightly or radical different things, depending on their culture, their understanding, their disposition, their

[40] Cf. also Eggert (1994): ' "Text,"… requires the socialized reader's engagement in the raising the meaning from the document' (p. 2).

circumstances, and so on. There are facts in the object (the document), but their meaning is not factual, it is interpretative. For one reader the only interesting dimension could be the semantic one (what the text means, the plot, who is the murderer), for another could be the artistic value: maybe she/he cannot read the words written in an unfamiliar language, but she/he can still admire and make (some) sense of the iconography and its artistic value.

Graphically, this model could be then represented as follows, where Documents have infinite Facts (F1-F∞) which can be arbitrarily grouped (dotted lines) into Dimensions by a Reader; the result of which is the Text, which is then a function of the document conjured by a reader:

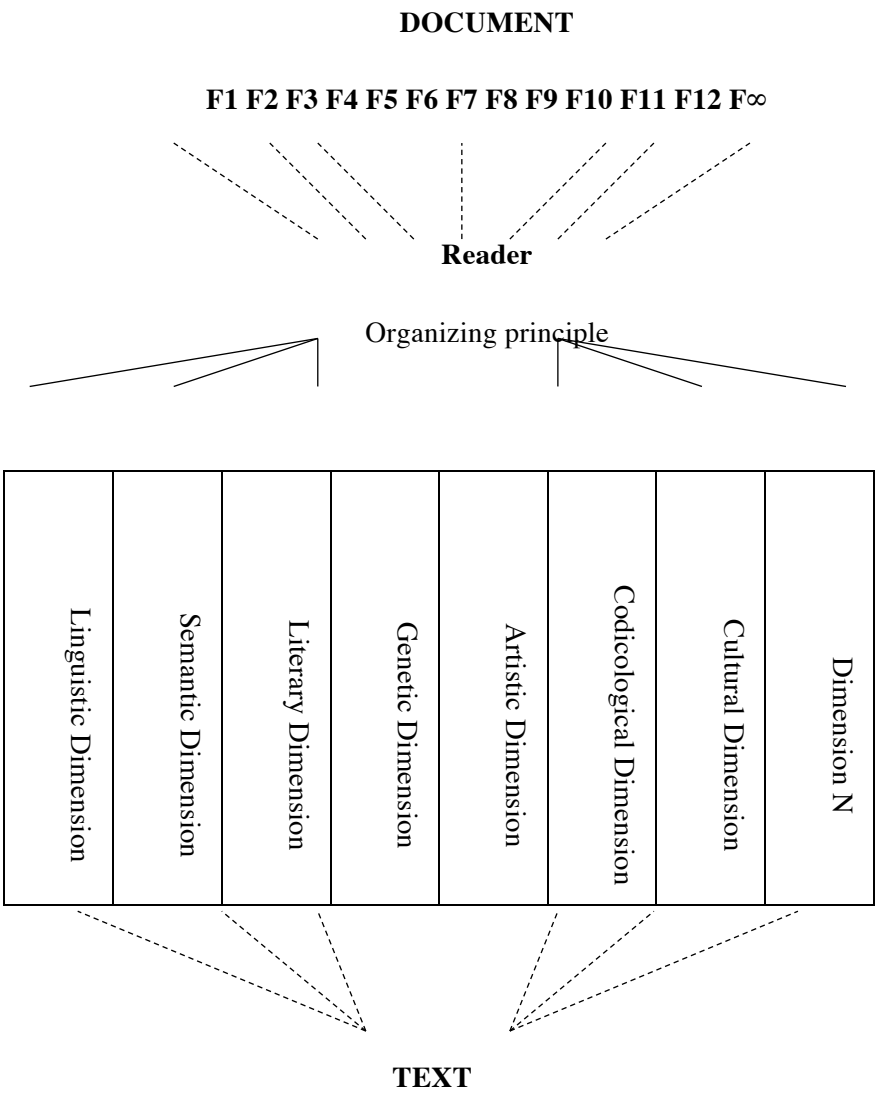


Figure 2.1: Conceptual model of texts and documents

Dimensions contribute to the generation of the meaning, and while all these dimensions are potentially present in any text transmitted by a document, we might not be interested in all of them (we may believe that some of these dimensions do not contribute to the message in any substantial way), or we may lack the ability to fully see some of them. Editors are very special types of readers and are likely to be engaged with many dimensions at once, but then, according to their scholarly approach and the purpose of their editions, they will weigh differently the facts included in each of these dimensions.

Again, while all the aforementioned dimensions may be drawn from documents, some of them seem to be more essential than others. For instance, we could have a user who

may not be interested in the verbal content at all, lacking the ability to understand the language (the code) in which the words are written, and be only interested in, say, the iconographic and codicological dimensions. Is the end result of such selection of dimensions still a text, then? The answer to this question of course depends on the point of view that has been adopted by this analysis, namely the point of view of textual scholarship; according to this point of view, then, a selection of dimensions that does not include consideration of the verbal content of a document is not a text, but must be *something else*. In order to define what this 'something else' actually is, it might be useful to reconsider the conceptual model and the way the entity we called 'text' has come into being. Text has been defined as a particular selection of dimensions operated by a reader according to specific organizing principle; this definition is then very close to the definition of a model provided at the beginning, the defining principle of which is the selection of an infinite set of facts with a purpose of study. We could then say that *a text is a model which, among the facts selected by the reader, includes the verbal content of the document*. We define then dimensions that include the verbal content of a document as **Verbal Dimensions**. Other selections which do not include the verbal content of the document are non-textual models. **Editions**, then presents texts in some formalized and specialized way, as maintained by Sperberg-McQueen ('Each edition provides some specific *presentation* of its selection': 2009, p. 31).

This generalization of the text as a type of model requires a similar generalization of the entity that has been called 'Reader': if reading is only one of the possible purposes of the models to be generated and if some of these models will not allow for reading, then we can possibly consider such entity as an User-function, one particular type of which will be the Reader. However, any act of reading is also an act of interpretation, as it is any other type of selection and grouping of the facts into dimensions, therefore readers could also be called **Intepretants**, to use Kralemann and Lattmann's terminology (2013, p. 3401). Editors have been previously defined as special types of readers, and even more precisely could be defined as a type of interpretants. Consequently, the edited text is a model produced by the interpretant-editor which combines some of the features observable in the document according to an organizing principle driven by a research purpose. There are of course many types of interpretants besides readers and editors: cataloguers, bibliophiles, art historians, collectors are just some examples of them; however this model, although generalizable, only focus on the editorial act and actors. Intepretants are the agents that selects facts, give them a meaning and organize them into dimensions; by doing this, interpretants make **Assertions** about facts and use such assertions to produce dimensions.

When modelling becomes a highly formalized activity, as in this case, it is methodologically more sound to adopt a modelling formalism to express the model in a standardized way. There are several such languages and notations: for instance Huitfeldt and Sperberg-McQueen (2008) use Alloy code using a literate programming system (p. 298). For the purpose of this exercise UML notation has been chosen instead, particularly for its simplicity and human readability.[41] The model, generalized and formalized looks as follows:

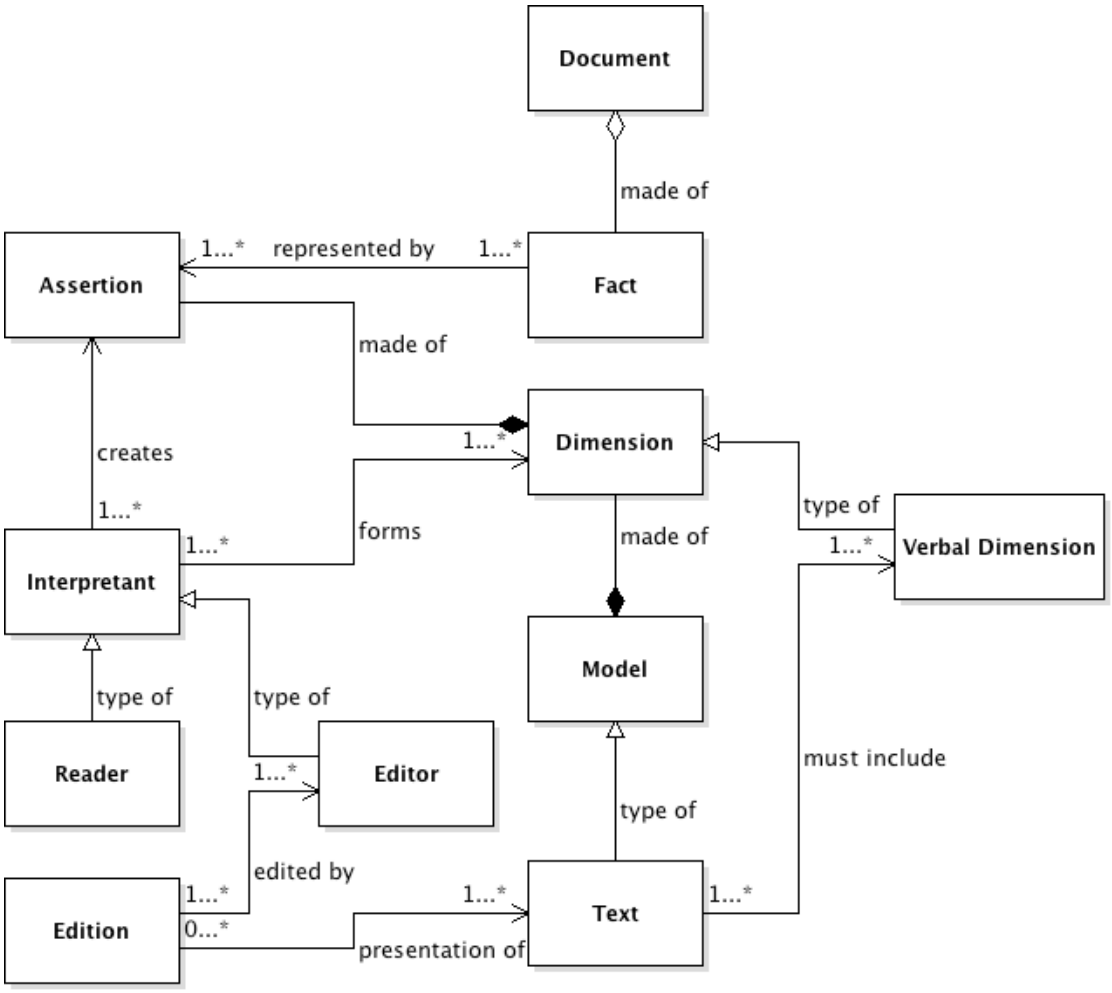


FIGURE 2.2: CONCEPTUAL MODEL OF TEXTS AND DOCUMENTS (UML)

2.3 Modelling Works

[41] Unified Modeling Language: see the website <http://www.uml.org>.

The dimensions that relate to the verbal content of a document have two characteristics in common that separate them from the other dimensions: they are immaterial and map to a linear sequence of finite, discrete set of conventional symbols, and therefore they can be transmitted across several documents or versions. Two documents from which more or less the same verbal content could be extracted or that could be understood to be elaborations (versions) of the same 'thing'[42] may share many of the features from the immaterial dimensions but not necessarily from the material dimensions. This observation requires the introduction of another entity into the conceptual model, that of **Work**. Eggert defines the work as 'a regulative idea' that helps in 'organising our remembered experiences of reading documents that are closely related bibliographically' (2009, p. 235). In this sense the work is not a physical object, but a combination of the immaterial dimensions of all the documents and all the texts that derive from them that have enough in common ('closely related bibliographically') to be defined as the same 'thing'. How much variation among the different texts and documents can be tolerated before it will be possible to define two different works? When can we speak of two versions of the same work or of two distinct works? Again, there is not a straight answer to this question, as this will depend on the interpretation of the special user that is the editor. In this sense a work represents an editorial assertion: a work is a work because an Interpretant-editor has declared that this is the case. But a work is not just summative, in the sense that it encompasses all the possible texts derived from all the documents by which it is manifested. Works may also be defined by Authorial Intention, in the sense that the work is the text as conceived by the unconditioned intention of the author (McLaverty, 1984; Bowers, 1991). The editorial work as represented by the model does not necessarily need an Author function, but it may, if the Interpretant-editor postulates it. With this, it is not meant here that the Author-function is not relevant or unnecessary for a theory of editing in the way proposed, for instance by Gabler (2012),[43] but that the model only requires it if the Interpretant-editor requires it; authorship needs to be assumed for any kind of document, however its relevance changes for many of the functions and activities of the Interpretant. Furthermore the label 'Author function' seems preferable to the one of Author, since the concept of Author is a rather controversial one; for instance,

[42] The expression is willingly vague as at this point the model concentrates on a limited range of variations of the verbal content, such as, for instance, the ones present in different printed editions. In the following sections more radical variations such as versions, translations and other derivative works will be considered and incorporated into the model.

[43] The terminology of the present model and Gabler 2012 is mostly coincidental. In fact, the model does not postulate real people but functions at all levels, from the author to the Intepretant. In particular, as the model is meant to describe many forms of editing, according to many scholarly practices, the use of optional functions rather than entities is necessary.

in case of highly elaborative traditions such as sermons or medieval epics, it is difficult to distinguish between an original author and a scribe who expands and elaborates, sometimes substantially, the original text. For Homeric works, for instance, the concept of 'author' is even more irksome.

If postulated by the Interpretant, the Author-function performs two main sub-functions, one *in posse* and one *in esse*, where the latter represents the activities of producing some of the facts present in the documents, especially, but not exclusively, the ones concerning the verbal content of the documents. The function *in posse* concerns instead the authorial intention, namely what the author function wanted to produce, but did not do it or if it did, the evidence of it is lost. It is *in posse* because it is unachieved, or if achieved is unknown. The Work is then at once the sum of all the meanings conjured by the users from all the extant documents, but could also be what the author intended to write. This latter concept could be intended as the exact sequence of words or other signs that the authors wrote as well as the meaning that they intended to give to them. Once again, the Authorial Intention is an assertion made by the Interpretant, the *deus ex machina* of the hermeneutics of verbal texts. Our model can then be defined as Interpretant-centric.[44]

Many theories of textual scholarship and of literary criticism may argue that the work level of the model is irrelevant or bares no particular epistemological value. However, it is still necessary to postulate the existence of such an entity in order to account for the fact that we are able to use the label, for example, *Pride and Prejudice* for many items that present more or less the same sequence of words even when inscribed onto different documents, using different fonts, over different materials, laid out differently with respect to the first edition, which in turn may be represented by many different objects (or items) that instantiate it.[45] The work is the concept that allows scholarly editors to combine readings from different documents, in order to create a critical edition; any text that combines readings from more than one source shall be called a **Critical Text**, or model-critical text. The ontological status of the model-critical text is similar to any other model-text considered so far, namely it represents a model of the combined facts and dimensions, with the difference that these facts and dimensions are derived from more than one document, rather than from one alone. Editions are the embodiment of a model-text, if derived from a single document, or of a model-critical text, if derived from many documents. Such models are based on a selection of dimensions and facts: not all facts, only some, namely the ones

[44] And in this sense it has some consonances with Eggert (2013) who emphasized the role of the reader.

[45] This concept is borrowed directly from the FRBR model, for which cf. below.

considered important and relevant by the Interpretant-editor. In fact, no edition will ever be able to account for all possible facts of a work; as Shillingsburg maintained '[n]o edition is a full representation of that which it attempts to edit. No edition was ever or will ever represent a work adequately. Full stop. The positive. The hopeful. The perfection. The adequacy. The triumph of scholarship. They will not occur' (Shillingsburg, 2006, p. 154).

The revised conceptual model now looks as follows, where the Author Function is split in two sub-entities, the **Producer of Document(s)**, which represents the historical agency that wrote the document, and considers therefore the writing of the document as an authorial function, and **Author as Hypothesis**, which is an assertion of the Intepretant; the Author as Hypothesis manifests a **Conjectured Authorial Intention**, which again is an assertion of the Interpretant.

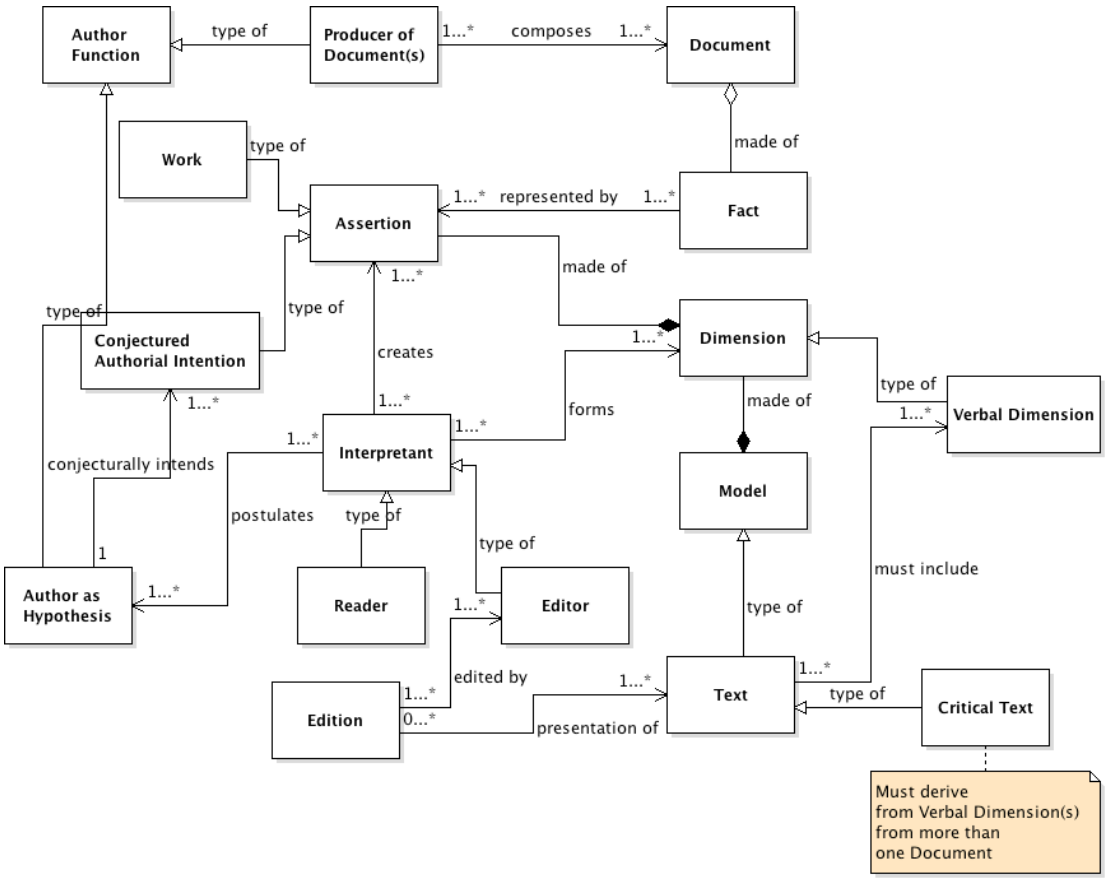


FIGURE 2.3: CONCEPTUAL MODEL OF TEXTS, DOCUMENTS AND WORKS

2.4 The material and the immaterial text: more on dimensions

In analyzing the conceptual model drawn above, it is worth considering in more detail the entity that has been defined as the Text, and some of its characterizing dimensions, namely the Verbal Dimensions which are connected with its verbal content.

The capability of texts to be abstracted from the support on which they were first inscribed and still retain their core essence is determined by their essentially linguistic nature: natural language can be transmitted orally, and in written form,[46] a fact that makes them transmittable across different supports and media. Like the Roman god Janus, texts have a dual nature: they are material and immaterial at the same time. This feature is not shared by other artistic artefacts such as paintings or sculptures, but finds an almost perfect equivalent in music. As for natural language, music can be written or produced as sound; texts and music can be performed in an auditory form or read silently from a written support. In both cases (texts and music) the transmission happens mostly in written form, and in fact editorial theory and practice of music has a lot in common with textual scholarship.[47]

Immateriality is then a characterizing feature of Verbal Dimensions, as shown by the table below which lists the documents' dimensions mentioned above.

Immateriality		**Materiality**
Linguistic dimension	**Cultural Dimension** **Genetic Dimension**	Graphematic/Palaeographic dimension
Semantic dimension		Artistic/Iconic dimension
Literary dimension		Codicological dimension

TABLE 2.1: MATERIAL AND IMMATERIAL DIMENSIONS

However, as displayed in Table 2.1, while most dimensions are either of a material or immaterial nature, the cultural significance of a text lies in between, as it cuts across dimensions and natures. For instance, we could say that an artefact such as the Book of Kells, preserved at the library of Trinity College Dublin, has a very high cultural value, with millions of visitors queuing every year to have a glimpse of it. Its value is given by its age, the state of preservation, the quality of the materials, the exquisite and abundant decoration, the role it has played and plays in the building of Irish national identity, and so on. Its

[46] This of course is a simplification that leaves out, for instance, sign language and other non verbal forms of communication.

[47] Cf. Grier (1996), p. 15: 'Since the beginning of scholarly editing in music, musicologists have borrowed the methods and approaches of philology'. Cf. also Broude (2012).

content, however, the words, are almost irrelevant: most visitors do not even know that the Book is a copy (a very corrupted one) of the Bible. At the other extreme is the so-called *Indovinello Veronese*, a small riddle made of nine words of no consequence written on the margin of a page of a mozarabic orational of the Biblioteca Capitolare of Verona (cod. LXXXIX, f. 3r), the value of which resides in the belief that it represents the first attestation of the Italian language; the *Indovinello Veronese* is perhaps the text with more criticism written per single words in Italy, if not worldwide.[48] Again, it is the cultural value that results in some manuscripts selling at auction for a few thousand pounds and other for millions.[49] In addition to this, there is another type of cultural value which is completely immaterial, and it is the value we attribute to literary works, among others. For instance, my personal copy of the *Commedia* by Dante has negligible cultural value (unless I become very famous), but the *Commedia* of Dante as a work has a great cultural value which is not bound to any specific object.

The genetic dimension is also located between materiality and immateriality, as the authoring (the genesis) of a text is normally recorded by traces on surfaces, the meaning and the understanding of which are inseparable from their physical inscriptions on the documents. Furthermore, most of these dimensions may be understood to have a dual nature: for instance the meaning of a word or a sentence may be understood differently whether it is found on the main body of the page or in the margins; as Ferrer maintains 'sign are toposensitive' meaning that their understanding and evaluation are strictly connected to the space where they have been inscribed by their author (1998).

Of the two natures of texts – materiality and immateriality – it is the latter that allows for their transmission: texts can be transmitted (that is transcribed and therefore made material) because they are also immaterial. Material aspects of texts can be studied, perhaps even understood, but their transmittability is questionable: documents can be transported, not transmitted. If we think, for instance, of some modernist forms of poetry, such as the *calligrammes* of Guillaume Apollinarie, the issue of the difficulty of transmission of the materiality of texts becomes clear: his poem in form of a horse can only be transmitted as a photograph, that is a surrogate materiality. Any other form of transmission, including transcription, would miss essential levels of meaning resulting in a total misunderstanding of the message.

---

[48] Discovered in 1924, the bibliography regarding the *Indovinello* counts hundreds of titles. Cf at least Petrucci and Romeo (1998), and De Angelis (2003).

[49] Not to mention that an object may also have a personal value which is irrespective of any broader cultural or market consideration.

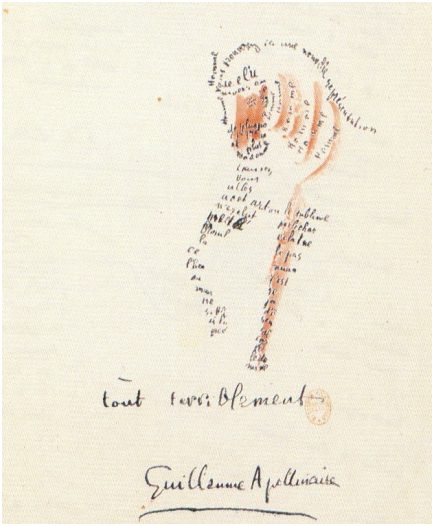


FIGURE 2.4 - APOLLANAIRE, *CALLIGRAME* – CATALOGUE OF THE EXHIBITION *LÉOPOLD SURVAGE - IRÈNE LAGUT* (1917) PARIS, 1917

Editorial theories and types of editions also define and differentiate themselves by the role they give to the materiality of texts, that is to the documents that transmit them. The graph below shows in a simplified and symbolic way the collocation of some of the most important editorial theories with respect to the centrality they give to the documentary evidence (the materiality of the text). Editorial approaches are located along a continuous spectrum between materiality and immateriality; in this context the dichotomy of materiality/immateriality could also be read as the predilection of a particular theory in producing text of documents or texts of works, in the sense established by Tanselle (1989) and discussed further below.

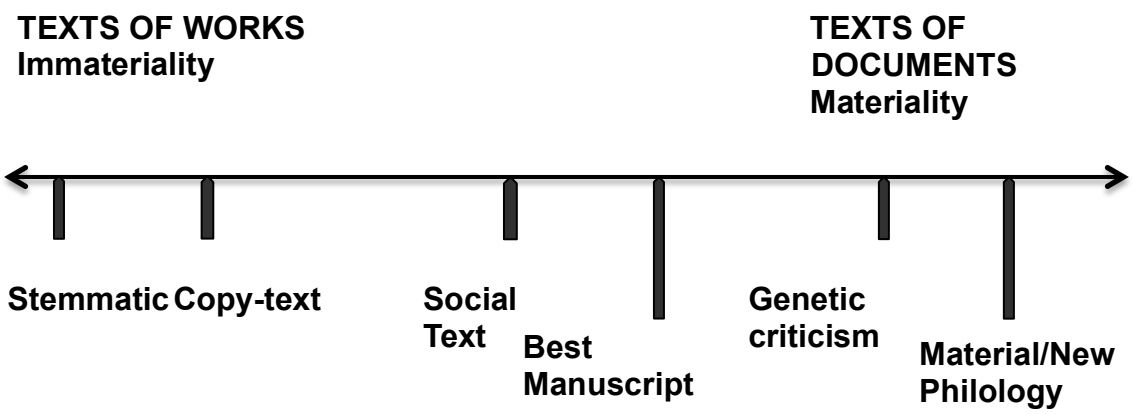


FIGURE 2.5: THE MATERIALITY AND IMMATERIALITY OF TEXTUAL THEORIES

Similarly, editorial format can be equally distributed along a continuous line:

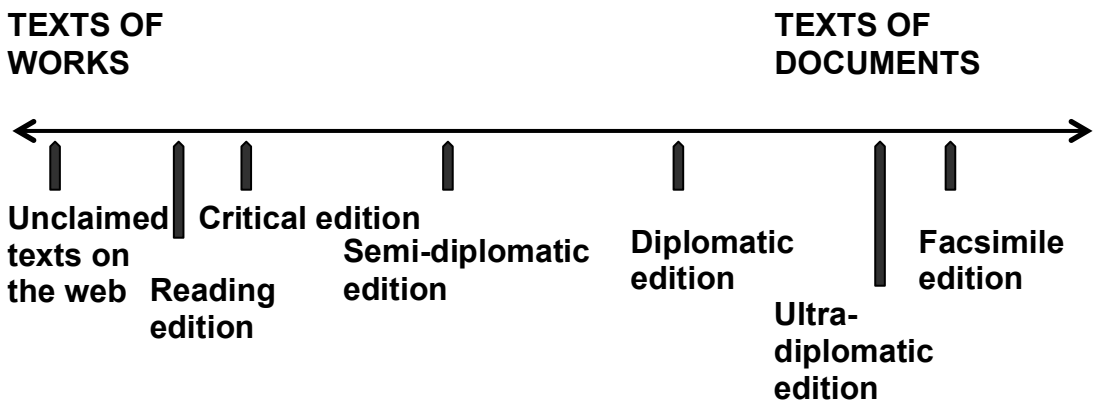


FIGURE 2.6: THE MATERIALITY AND IMMATERIALITY OF EDITORIAL MODELS

The maximum immateriality is offered by the category 'Unclaimed texts on the web', which refers to those texts which are presented without any reference to the material editions from which they are presumably drawn: no reference to editor, source edition or original document. While presenting a very complex reality of theoretical positions in a perhaps over-simplistic way, these graphs have the advantage of showing the path between (immaterial) text and (material) document as a continuum, with no clearly defined boundaries, as well as the relative positioning and therefore distance and vicinity of different theoretical approaches.

2.5 Expanding the Work: Versions and Derivative Works

Texts are not isolated entries, they live in relationship with other texts; some texts are very similar to the point that they are called versions of each other; some texts derive from others; some other texts are combinations of many texts. The concept of work has been introduced to account also for these types of phenomena, but it now requires expansion to include versions and their typology: versions produced by the same or ancillary authorial functions, such as re-elaboration of the same work, and derivative versions that imply the change of medium or a considerable change in the verbal content. For instance, let us consider the work that goes under the label of *Romeo and Juliet*: besides the play written by William Shakespeare, which exists in different (authorial?) versions, there are also a number of movies, TV adaptations, a ballet, and almost countless derivative works, such as translations or endeavours such as, for instance the musical *West Side Story* in both theatrical and cinematographic versions, and many, many others. All of these entities are related to one another in some way, since they all feature the tragic love story of two young people belonging to rival families, but they also present at the same time a level of individuality such that we are able to distinguish them when needed.

Textual versions are based on substantial differences in the verbal content of the documents and are determined by the interpretant-editor, on the basis of the readings offered by the documents. In other words, the existence of versions is established by Interpretant-editor who decides that textual variation among documents is irreconcilable with a single act of creation; it is again the Interpretant-editor who establishes if a version belongs to an existing work or it substantiates an independent work. Critical editions of multi-versioned works may draw from the many versions or from only one: the first, the last, the most completed, the uncensored one, and so on, depending on the circumstances and on the theoretical choices of the editor. As versions can be witnessed by more than one document, it is also possible to produce a critical edition of the versions themselves. Textual versions therefore behave as sub-works as they share all the same features of works, being hierarchically subordinate to them. In fact, derivative works, such a translations or filmic versions of novels, are not too dissimilar, however they have a different ontological status because in derivative works it is possible to recognize other types of authorship with respect to the work from which they are derived, being therefore much more independent compared to the versions. It would be scholarly unacceptable, for instance, to produce a critical edition of a work by combining readings of documents and filmic versions of the same work.[50]

[50] For translations, nevertheless, the situation is much more complicated. In particular for medieval texts, where the 'original' version is lost, translation may be the only relics of an otherwise lost text.

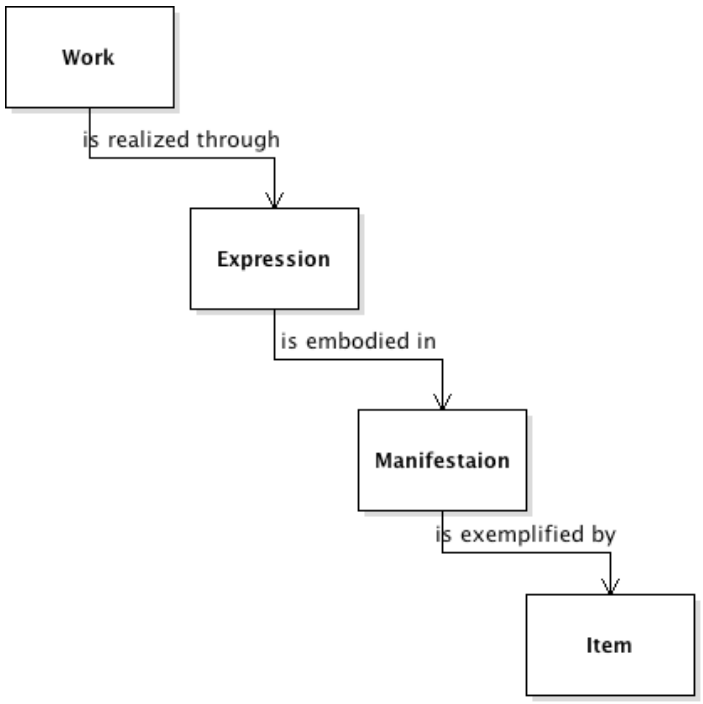
Derivative works, then, are related to the work from which they derive, but the level of dependency varies considerably. The model therefore will have to be recursive.

2.5 Other models

The library community has dealt with this level of complexity in order to catalogue coherently and group together their holdings, and to respond to users' requests. The most famous and influential model of the bibliographic domain produced by libraries is the *Functional Requirement for Bibliographic Records* (1998) known also as *FRBR* model. According to it, in the 'bibliographic universe' (Tillet, 2005, p. 197):

> The entities defined as *work* (a distinct intellectual or artistic creation) and *expression* (the intellectual or artistic realization of a *work*) reflect intellectual or artistic content. The entities defined as *manifestation* (the physical embodiment of an *expression* of a *work*) and *item* (a single exemplar of a *manifestation*), on the other hand reflect the physical form. (FRBR, 1998, p.12)

Schematically, it could be represented in the following way:



This is the case, for instance, of the *Book of the Marvels of the World*, which narrates the travels of Marco Polo. The original version was compiled by Rustichello da Pisa, following the narration of Marco Polo himself in, probably, a Pisan dialect. However this version is completely lost. The oldest versions that are preserved an Old French and a Latin translation.

FIGURE 2.2: THE FRBR MODEL (FRBR 1998, P. 13)

To go back to the *Romeo and Juliet* example, as a Work it represents the intellectual endeavour created by Shakespeare. The work is available in different Expressions: the version of the first Quarto (Q1), the version of the second Quarto (Q2), the critical edition established by Mack and Boynton, the critical edition established by Cheetham in 1987, and so on. It can be noted here how the FRBR model conflates in the concept of expressions both the versions of the work, in the sense of sub-works seen above, and model-critical text, but not its embodiment, which is covered by the next step. Expressions are Manifested by a printed edition published in 1597 by John Danter, a printed edition published by Cuthbert Burby in 1599, by a printed edition published by Boynton Cook Publishers in 1981 and in 1990, and so on. To each of these manifestations we can relate the individual Items that are preserved in public and private libraries. The FRBR conceptual model considers Manifestations and Items as physical and therefore material, but, in reality, only Items are truly physical objects, while the Manifestations are the idealized sum of all the items, a concept which is strictly related to the 'ideal copy' elaborated by analytical bibliography (Bowers, 1949, pp. 113-23). The distinction between Manifestation and Item, even if still valuable from a conceptual point of view, is less meaningful when it comes to a manuscript-based textual transmission as, by definition, each preserved manuscript is a Manifestation with one and only one Item. No two manuscripts are identical, both at word level and at physical level: each manuscript is then both a manifestation and an item. Surrogates such as digital or analogue facsimiles are, arguably, the only exception to this general rule, as they could be considered different Items of the same Manifestation.[51] However, the problem with the FRBR applied to manuscripts seems even deeper: in fact FRBR, modelled around printed books, fails to effectively account for manuscript transmission and circulation. Manuscripts of the same work are all different from one another, but whether these differences account for different Works, Expressions or Manifestations may not be very straightforward. Expressions are defined at word level: "Strictly speaking, any change in intellectual or artistic content constitutes a change in *expression*. Thus, if a text is revised or modified, the resulting *expression* is considered to be a new *expression*, no matter how minor the modification may be" (p.19). If we apply this strict approach, we have to assume that each manuscript of a given work represents a different Expression of the same Work. However, this concept does not allow, for instance, for a distinction between minor differences such as orthographic variants, errors, and readings which are connected to the

[51] For a discussion on the relationship between document and facsimile representation, cf. Chapter 4.

fact that a manuscript is written by hand by a human being prone to mistakes and to leave an individual imprint in the text,[52] and more substantial re-elaborations, which we have called versions, like in the case of Q1 and Q2 for *Romeo and Juliet*, where Q2, beside correcting many mistakes of Q1, presents around 800 more verses than Q1. If we identify an Expression for each manuscript, which in turn may have only one Manifestation and only one Item, we see how the FRBR model can be ineffective in describing conceptually the relationship of bibliographical items when the latter are manuscripts. The distinction between Manifestation and Item seems also meaningless as all the characteristics of the Manifestations are characteristics of the Item as well. It is useful here to remind ourselves that the FRBR is a conceptual model that aims to organize the holdings of a library, and therefore it has different aims to those of a textual scholar attempting to find some regulative principle for a complex reality of material and immaterial objects and ideas. This problem however has been addressed and solved by the so called FRBRoo model (FRBR object oriented), which represents a revision of the FRBR model created by the collaboration of the museums and libraries communities with the purpose of establishing an ontology 'intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation, and interchange of bibliographic and museum information' (CIDOC-CRM, 2009, Online). The FRBRoo model is much more complex than the FRBR model, and it is beyond the scope of this chapter to give a complete account of the model, it will suffice here to mention the fact that the model introduced a fifth entity, namely the 'Manifestation Singleton' (FRBRoo, 2009, p. 34), specifically thought for 'unique objects, with no siblings intended in the course of their production', such as 'manuscripts, preparatory sketches and the final clean draft'.

The FRBR model also proves inadequate to describe derivative or related works such as the ballet or the movies derived by Shakespeare's play. The FRBR model recognizes that people mean different things when they refer to *Romeo and Juliet*, but while the model explicitly declares that Shakespeare's play and Zeffirelli's movie with the same title are two different works as the movie 'involves a significant degree of independent intellectual or artistic effort' (IFLA, 1998, p. 17) with respect to the play, at the same time the model seems unable to recognize in a formal way the fact that they are related to each other. Pragmatically, some librarians have introduced the level of 'family of work' to represent such relationship (Tillet, 2003, p. 17), but this again fails to recognize, for instance that the

[52] Of the type meant by Shillingsburg when he speaks of "infelicity in transmission" (1991, p. 12), for instance.

movie is derived from the play, and not vice versa. Once again FRBRoo addressed the issue with the introduction of the concept of 'Complex Work' and 'Container Work'.

In spite of these limitations the FRBR model is useful in both its declinations as it helps to specify different levels of abstraction and of materiality/immateriality with respect to the ones assumed by our conceptual model. If we consider, for instance, the multiple dimensions listed above, we can trace them to the different levels of the FRBR model as follows:

WORK -Semantic dimension -Literary dimension EXPRESSION -Linguistic dimension -Semantic dimension -Literary dimension -Genetic dimension -Iconic/artistic dimension MANIFESTATION/ITEM -Iconic/artistic dimension -Graphemaic/palaeographical dimension -Codicological Dimension	Cultural dimensions

TABLE 2.2: DIMENSIONS AND THE FRBR MODEL

Both the semantic and the literary dimensions can be considered at Work and at Expression levels: as they are connected to the specific wording of a text, their instances need to be considered as representative of a particular Work and embodying a particular Expression. The iconic dimension characterizes the Expression as contributing to produce the set of meanings that constitutes the Expression, but they also have physical characteristics that are typical of the Manifestation/Item level. For instance, the elaborate iconographic programme of the Lindsfarne Gospel is not only a defining character of the particular copy of the Gospel which is preserved at the British Library and has shelf number Nero Cotton D.iv, but it also changes the way we read and understand the words in there, and the meaning we conjure from there. The graphematic and palaeographical dimension, as well as the codicological dimensions naturally belong to the Manifestation/Item level;

however, one could argue that a particular choice of script or types, of codex structure, of binding or of material may respond to a specific cultural programme and have a specific meaning, with the result that would not figure wrongly if we had listed it under Expression.

The FRBR model is also useful as it tries to organize in a coherent and hierarchical system different bibliographical concepts, following quite different principles with respect to the one adopted by the conceptual model elaborated in this chapter, demonstrating how a different functional requirement and a different point of view can produce a very different conceptual model. However, even when the point of view is similar, the result can be quite different. The theory of the pluralistic text elaborated by Patrick Sahle presents some cross-similarities with the FRBR model and with the conceptual model elaborated above. According to Sahle a text can be view in many ways:

1. Text as Idea, Intention: 'texts are what they mean'.
2. Texts as Work: 'a clearly defined narrative structure; thus, for example, the opening paragraph of Saint Patrick's *Confessio* is always *Confessio* 1, be it in the original Latin or in the German translation; it is always the same text'.
3. Text as Linguistic Code: 'as a specific expression, a *linguistic code*, a certain series of words'.
4. Text as Version: texts may present themselves with different wording.
5. Text as Document: texts may have different embodiments which can help, for instance, in understanding the chronology of the writing by studing the colour of the ink.
6. Text as visual Sign: how the text looks like, 'the physical appearance of medieval charters or, most notably, of the Bible as a book (e.g. in a liturgical context, or in court when you swear on it) is highly meaningful. The "text itself" does not matter here at all — the book might even contain blank pages — yet the text is identified by means of its physical appearance and, as such, points to a certain idea or truth'.[53]

Sahle organizes his pluralistic view as a wheel (fig. 6): in a circular, non hierarchical argument, Ideas are structured into Works, which are expressed though Language, any variation of which generates a Version, which are embodied in Documents via the means of Signs, the interpretation of which produces the semantic of the text or its Idea.

[53] All quotations from Sahle are taken from Fisher 2013, pp. 78-79.

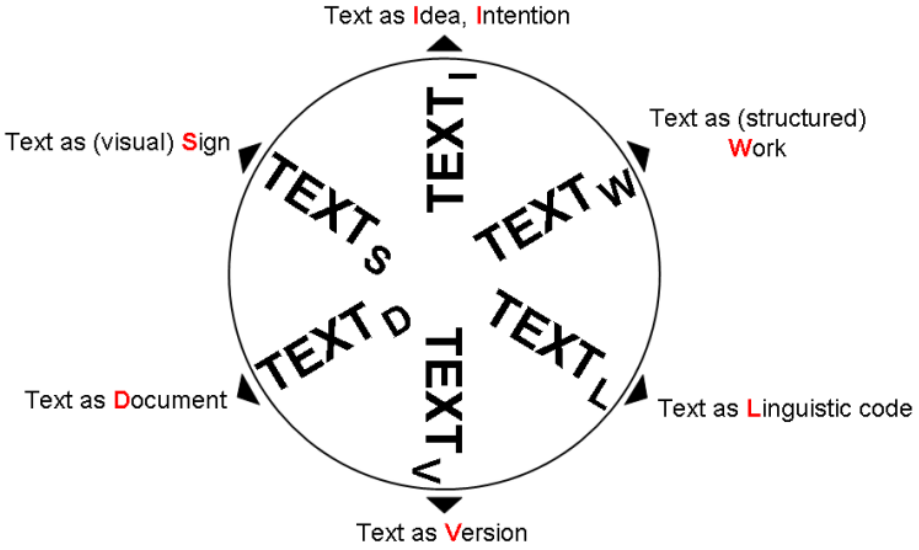


FIGURE 2.8: THE PLURALISTIC TEXT (SAHLE 2013)

It is clear from this outline that the pluralistic view of the text promoted by Sahle and the conceptual model described above have many points in common, but different granularities. They both claim that texts are complex objects, that mean different things according to the point of view, and that none of these points of view can claim to be The Text. There is also some overlap between Sahle's and the FRBR model: for example the Text as Linguistic Code and FRBR Expression try to capture more or less the same concept; however, the Text as Work and the FRBR Work entity do not coincide, as for Sahle translations represent the same Work.

The fact that there is more than one opinion on what is a work comes as no surprise; Paul Eggert devotes an entire chapter of his excellent *Securing the Past* (2009) to the analysis of the concept of work, overviewing a number of different theoretical positions. But perhaps, from a textual scholarship point of view, the most influential approach has been the one offered by Thomas Tanselle in his seminal *A Rationale of Textual Criticism*, published for the first time in 1989. In his vision it is essential to distinguish the texts of documents (that is the texts one can derive from physical documents) from texts of work, namely the ideal texts that the author had intended to write but which have never been realized in practice; therefore the work can be defined as the unachieved ideal authorial intention, to which editors ought to aspire in their attempt to reconstruct it. As one can see there are several points of contact between Tanselle's theory and the conceptual model presented here, but also many points in which they diverge. In Tanselle's vision, the existence of many texts derivable from documents presenting different versions of the same text demonstrates that each of these texts

of documents represent but do not present the work. This assumption allows for, and actually requires, the work of the editor to reconstruct the texts of works and to pursue the authorial intention, which, in his vision, ought to be the editor's goal. Here the differences between a theory of textual scholarship and a conceptual model become more evident: while Tanselle's theory has a prescriptive outcome, based on an intellectual agenda, a conceptual model describes a given domain in order to understand it. Therefore, while for Tanselle the purpose of the editor is to combine readings from different documents in the attempt of getting closer to the unattainable work, the conceptual model describes and equates different editorial choices.

For Peter Shillingsburg (1991) the work is a Conceptual Text which is not materially witnessed in any single copy. However, the work can only be experienced via the Material Text (or document), and when the Material Text is read, then it becomes a Performed Text or Reception Performance, in his taxonomy. Shillingsburg too attempts a hierarchical organisation of his classification, when he states that Linguistic Texts – which are made of "words and punctuation in a particular order" (p. 52) – have three forms: "Conceptual, Semiotic and Material Text", where the Conceptual Text represents the work, the Material Text the document and the Semiotic Texts the "signs found recorded in a physical form". Such a taxonomy has many intricate implications, and is provided as a complex architecture which cannot be easily summarized; this is perhaps one of the reasons for the uneasiness expressed by Paul Eggert. Eggert maintains that by trying to distinguish different levels of the text, as Shillingsburg does (and as I do in this chapter, most probably), 'we are necessarily abstracting from the messy processes that all agents [i.e. 'author, copyist, editor, typesetter and reader', p. 234] of textuality actually undergo throughout the production-consumption spectrum. When we nominalize, we are erecting our own methodologies for productive ends rather than pointing to an actual ontology of texts' (p. 236). In his opinion, then, these classifications only have a limited value, as the 'messy processes' that are at the core of the creation of texts, works, and documents, cannot but be studied as such, in all their messiness. In other words, Eggert seems to disavow one of the main advantages of modelling (in particular of the so-called 'model-of'), which is the simplification of complex and messy processes in order to study them. One could argue that only by analysing, classifying, and simplifying messiness it is possible to study it properly, not to mention the necessity of modelling imposed by computational analysis; however, Eggert himself recognizes that these simplifications may embody our editorial methodologies employed 'for productive ends'.

The concept of work, or a renewed concept of it, is at the heart of Eggert's own theory. He calls the work a 'regulative idea' that 'retains its function as a pragmatic agreement for organising our remembered experience of reading documents that are closely related bibliographically… the work unravels, in every moment of its being, into a relationship between its documentary and textual dimensions" (2009, p. 235).

The work, what it is and what its role is with respect to textual scholarship, how it relates to the text and the document, has been and is at the heart of the debate of textual scholars of many theoretical orientations. The last in the series to date is Peter Robinson who in his contribution of 2013, following Eggert, defines the work as 'the object we seek to know' (p. 120) and the text as 'the site of meaning which links the document and the work'. In his view, the task of each scholarly edition is to 'illuminate both aspects of the text, both text-as-work and text-as-document' (p. 123); in fact the 'document without the text of the work we construct from it is mute, simply marks on a surface' (p. 120).

Returning to the work, it seems that what all these definitions have in common is that the work is what the editors have in mind when they prepare a critical edition. In the division of material/immaterial seen before, the work is definitely leaning toward the immaterial; one could even say that the concept of work represents the quintessence of immateriality, because it not only includes a linguistic, semantic and literary dimension, as the immaterial text, but it can only exist *in posse*. Another characteristic of the work that is possible to draw from all of these models is that its concept is bound to change with time: because the work is more of an aspiration than an achievable editorial result, this aspiration changes every time a text is edited and every time new evidence is uncovered. It also changes with respect to the editorial beliefs of its editor, in the perceptions of its readers and with respect to cultural frameworks and trends. This ever-changing and never-achieved work is defined by the relationship between text and document, as stated by Eggert. The conceptual model described above, however, provides a different meaning to the Work, which is not the goal of a model-critical text, but the sum of all documents and model-texts derived from them, in the light (or not) of the authorial intention; the work is a statement of the interpretant, and changes therefore with the interpretant. This personalisation of the concept of work is somewhat implicit in Eggert's vision, as he relates the work to 'our remembered experience of reading documents'.

2.6 A Born-digital Model: the OHCO Model

What all of this has to do with digital editing? The last century of textual scholarship has been characterized by the debate between supporters of critical editions and authorial intention against the advocates of the documentary editions and historical evidence;[54] the introduction of digital methodology into the preparation and the delivery of scholarly editions has redefined such debate according to different parameters, giving new fuel to the argument. The discussion has developed along two main lines: on the wave of the enthusiasm generated by the discovery of markup languages, a group of scholars has concentrated their efforts on defining what text really is (with respect to documents) and which features should be considered ontologically relevant to its definition. A second line of discussion is focused instead on the role of documentary editing and the role and responsibility of editors in providing (or not) their readers with reliable, critical texts which combine readings from different sources. While the latter will be analyzed in Chapter 3, the former shall be discussed here, as it focuses on one of the most influential born-digital model of textuality, namely the OHCO model.

Markup languages, and in particular generalized markup languages such as SGML (Standard Generalized Markup Language) and its descendant XML (eXtensible Markup Language) are more or less the standard formalism for the encoding of digital editions and form the basis of the TEI. They possess two characteristics that have determined the elaboration of a highly debated and influential theory of (digital) texts, the so-called OHCO model (Ordered Hierarchy of Content Objects: DeRose *et al*., 1990). First, markup languages such as SGML and XML force their users to include texts within a single structural hierarchy, without overlapping. As for Russian dolls, each structure neatly contains a hierarchically subordinate one; in case of texts, it means that each chapter, for instance, has to end before a new chapter can begin, and that paragraphs cannot spread across chapters but must be contained entirely within them. This characteristic implies that if one wants to use markup to encode a poem, it will not be possible to mark both syntactic features and verses at the same time, as sentences normally spread across verses. The second aspect that has had a crucial importance in the development of the OHCO model is the fact that SMGL was invented with the purpose of simplifying the typesetting of text for printing, thus separating the content of a text and its "meaning" from the way it was to be presented on a screen or on paper (Renear, 2004, p. 220). So, SGML (and XML later on) made it possible to declare that a specific sequence was the title of a section, while storing in some other ancillary file the information that it should have been displayed, say, centred and with a font 24pt high. SGML

[54] The origin of this debate can be traced back at least from the essential criticism that Joseph Bédier made of the Lachamannian method (Bédier, 1928).

and XML, being completely agnostic on the type of characteristics to encode, do not force their users to separate format and content in this way, but as this separation represents one of the most important use cases that has determined the development of SGML in the first place (Renear, 2004, p. 221), it has generally been assumed (end explicitly enforced at least in the form of best practice recommendation) that a person wanting to encode their text using a markup language would want necessarily to separate format from content. The result of these two characteristics combined has been the elaboration of the OHCO model, or the definition of 'what text really is', that is to say one hierarchy of structures that neatly contains all other structures, these structures being features such as sections, subsection, chapters, paragraphs, and verses. These structures do not include pages or lines, however, as these have implicitly been considered 'not scholarly enough':

> [A] book can be divided into pages; a page into the header, the main text area (perhaps with several columns, embedded pictures, etc.), an optional footnote area, and a footer. However, even this model fails to provide the kind of text handling needed by authors and scholars. How can one find equations, poetry quotations, lines of verse, and the like?" (DeRose *et al*., 1990, § 10).[55]

However, the OHCO model only accounts for what we have called the immaterial Verbal Dimensions, and for one at a time, in the sense that this model cannot represent features belonging to, say, linguistic and literary dimensions at the same time. This model has been harshly contested and discussed since its first publication (Huitfeldt, 1994; Pierazzo and Stokes 2010), and rejected repeatedly as an overly simplified vision even of the immaterial text alone: hierarchies in texts, as a matter of fact, overlap, even not considering any material aspects, as the case of verses and syntax mentioned above clearly shows. Models, by definition, are a simplification of real objects with the purpose of some type of analysis; however they might be too simplified to be useful, and while the OHCO model may be useful in the context of digital publication of text for which reading is the main purpose, it is ill equipped to serve editors in their scholarly work of analysis of documents and preparation of critical editions. Such a drawback does not undermine the value of the model as hermeneutical object: as maintained by McCarty, in fact, 'modelling succeeds intellectually

[55] It is perhaps worth noting here, however, that in order to exemplify what a *text* really is, the authors of the OHCO model have chosen to describe the structure of a *book*, thereby slipping from the immaterial abstraction of the text to the material level of the document: 'A book for instance is a sequence of chapters, each of which is a sequence of major sections, each of which in turn is a sequence of subsections.' (Renear et al. 1996, § *OHCO-1*).

when it results in failure, either directly within the model itself or indirectly through ideas it shows to be inadequate' (McCarty, 2004, p. 26), and this is because the failure of the model pushes the research forward, toward a new conceptualization and a new model.

Despite this rejection, the OHCO model endures in various ways. The reason for this endurance of such an essentially unsuitable model is due to the fact that this model is actually an accurate description of the two characteristics of markup languages mentioned above, more than being a theory of what text really is. As in fact Allen Renear himself declared, it was by observing the many advantages for digital publishing that descriptive markup entailed, that lead 'some people to suggest that it was not simply a handy way of working with text, but that it was rather in some sense deeply, profoundly, *correct*, that "descriptive markup is not just the best approach… it is the best imaginable approach"' (Renear 2004, p. 224, quoting Coombs *et al*., 1987). From an empirical observation to the elaboration of a theory the step was a short one. One could call OHCO the abstract description of the way markup languages such as XML and SGML (conceived as stated above) force their users – or at least encourages – to consider texts. It is, so to speak, a 'matter-of-fact theory': if one wants to use XML to digitize a text, then the OHCO model is the only one to use, because this is the way XML works, more or less.[56] The encoding models provided by the Text Encoding Initiative (TEI) mostly follow this model, as the TEI is based on XML and therefore this approach is almost inevitable. Or is it? While the single hierarchy component of the OHCO model is somewhat unavoidable, as it depends on the nature of XML, the fact that the immaterial structures are to be privileged over material ones has no technical grounds, and therefore this part of the model can be and has been contested. The OHCO model was elaborated to describe the publishing process: one should describe what a text is and not how it should look once published, digitally or on print. But what about how the text looked in its source document? The OHCO model was elaborated by observing the steps and the requirements of the publication process, and not the editorial, scholarly work, so it comes as no surprise to discover that it falls short on the task of providing a viable model for the editorial work. In an article of 2010 the OHCO assumption that 'a book … is a sequence of chapters, each of which is a sequence of major sections, each of which in turn is a sequence of subsections' (Renear et al. 1996, § *OHCO-1*) has been demonstrated to be an oversimplified model of text (Stokes and Pierazzo 2010). This article lists a large number of texts for which the divorce of materiality from immateriality is not only painful, but comes at the price of a fundamental loss of information, deeply affecting our understanding of the text itself. The OHCO model, then, has right of citizenship in scholarly editing as it provides a

[56] See Chapter 5 for a discussion of stand-off markup.

shortcut label for the way the most established technology in the field works; all the criticism that it has raised only represents the criticism to the use of XML as underlying technology for textual scholarship in particular, and text representation in general. Other limitations and possible ways out from the XML impasse will be discussed in Chapter 5. Chapter 3, in the meantime, will continue with conceptual modelling of the bibliographical domain, with a particular focus on textual transmission.

3. Modelling Text Transmission: from Documents to Texts, and Return

Textual scholarship (and digital textual scholarship) engages with texts transmitted across time by means of mainly physical supports.[57] These supports can be made of different material (paper, papyrus, parchment, stones, fabric, wood, etc.) and can be handcrafted (manuscripts, epigraphy), or mechanically produced (printed books). Each combination of the above presents its own issues, problems, and requires specialized expertise and experience to be understood in full. Can a text be separated from the physical object on which it is inscribed without loss of information? The whole idea of transmission, edition and distribution of texts from the invention of writing requires us to answer 'yes' to this question. For centuries, if not millennia, human civilisation has relied on the fact that text can be transmitted independently from the document in which it was originally inscribed by its 'author'. On the other hand, it has been clear for millennia, again, that 'without loss' is a utopian statement and since the establishment of the philological school of Alexandria, scholars have tried to find the rules and principles to fix mistakes, record different readings, and restore the 'original' version of a given text – whatever meaning is given to the word 'original'. Traditionally, we date the birth of textual scholarship with the work of the scholars at the library of Alexandria and from the edition of the *Iliad* by Aristarchus (200 BCE). From the very beginning, what struck those scholars was the discovery that, comparing any two manuscripts of allegedly the same work, the texts contained within these documents were different, in some cases considerably so. Since then it has been the purpose of textual scholarship to make sense of this variation, to discover the reasons for which variation occurs and to reconcile this with the need of providing reliable and correct texts to the readers. What 'correct' actually means defines the different theories of textual scholarship.

Texts (intended as the verbal content of documents, i.e. the Verbal Dimensions) can be separated from the medium that originally hosted them because they 'are not *only* physical objects' (Eggert 2009), they are also immaterial, abstractions, and can be realized in many different ways, using different media, through different channels. This capability comes at a price: variation and progressive deterioration of the original verbal content. Textual scholars and philologists have studied the phenomenology of textual variation since antiquity, and why variants (and errors) happen is well understood. However, when dealing with computational methods, more than why something happens, it is important to establish how

[57] Increasingly, however, textual scholarship deals with born-digital texts, cf. for instance, Kirschenbaum and Reside (2013) , and Vauthier (2013).

does it work, and if this can be made computationally processable; in other words, how it can be modelled.

Text transmission can be generalized and modelled as a form of information transmission, and therefore can be considered a type of communication. If we take this route, it seems then appropriate to try to apply models developed by classic theories of communication. As we will see the experiment has proved quite successful, even if a certain number of adaptations have to be applied in order to fit the model to the use case.

3.1 Textual Transmission: a Communication Model

The capability of texts to be transmitted across time and media is at the core of a great part of human civilization and it is also the object of study of textual scholarship: editing itself is nothing if not a form a textual transmission, and critical editors are in a sense not so different to ancient scribes who were contaminating readings from more than one source, correcting what they thought was wrong in their antigraph, apart from the methodological rigour and the scholarship itself, that is. In the previous chapter we have modelled the way a text comes into being, namely by the act of an Interpretant who selected a series of Facts from one or more Documents, combining them into a Text, but how does such a Text get transmitted and what are the circumstances and the entities that can influence the way they are transmitted? We have defined Edition as the 'public' face of a Text, but how does it work as vehicle for distribution? Can the cognitive process that is enacted by the Interpretant when considering the Facts of the Document be formalized and described? If we consider textual transmission as an act of communication and adopt the terminology of communication theory, then we could refer to each of the Dimensions introduced in the previous chapter as 'codes'. We could then define Text (or a model-text) further as follows: 'a text is (also) a multidimensional message that conveys a set of meanings transmitted by various codes which are potentially understandable to at least one group of Interpretants with the capability and interest to decipher at least one of such codes'. According to this definition, we could then define textual transmission as a selective act of communication where the Interpretants-receivers willingly or unwillingly makes the selection of which code to decipher from the number potentially available from the source that contains them.

According to the classic theory of communication elaborated by Shannon and Weaver in 1948, a communication act is represented by the transmission of a message produced by a source from a sender to a receiver via a specific channel until it reaches its destination, using a shared code. If we apply this model to textual transmission, then we

could say that the source is where the production of the message happens, the responsibility of what we have called the Producer of Document(s) and which in turn could be the author, the scribe, the publisher, the scholarly editor who crafts a new critical edition, or any other agent (or group of agents) who creates or contributes to the creation of the text. The same entities can be considered the transmitters, in the moment they communicate the texts they have created (or modelled). The receiver is the Intepretant-reader, who, in a circular motion, could be the scribe, the publisher, the editor who has to decode the message to be able to retransmit again, while the destination is where the meaning is created, namely the model-text. Finally, the channel is the medium, the support that transmits the text itself, which in our model is represented by a set of the facts derivable from the Document.

According to the same classic communication theory, there are many factors that can affect a perfect transmission of the message from the sender to the receiver; for example we can think of the imperfect quality of the transmission channel or medium, which could be affected by the so-called noise, and the non-perfect coincidence of the code of sender and receiver. If we translate this into our textual transmission case, we can consider as noise any external interference with the support (the channel) that might affect the capability of deciphering the messages. Damage caused by water, pests and fires, or rebinding could be considered a type of noise. The change of medium should also be considered here. McLuhan declared that 'the medium is the message' and by this he points out how our perceptions are influenced by the way the message is conveyed to us; he actually pushes the argument to the point of saying that the way in which each medium influences the content becomes part of the message if not the whole message itself, as it transforms the message in such a way that it becomes unrecognizable (McLuhan, 1964).[58] This argument, of course, applies both to historical re-mediations, such as the transfer of classic literature from scrolls to codex or the revolution caused by Gutenberg's Galaxy (as outlined by McLuhan in a 1962 book which studies the sociological impact of the introduction of new technology in culture and society),

[58] Actually, the title of the book by Marshall McLuhan should more correctly be quoted as *The Medium is the Massage*. In the Commonly Asked Question of the website maintained by McLuhan's estate at the question 'Why is the title of the book "The medium is the massage" and not "The medium is the message"?', the son of Marshall, Eric McLuhan answers: 'Actually, the title was a mistake. When the book came back from the typesetter's, it had on the cover "Massage" as it still does. The title was supposed to have read "The Medium is the Message" but the typesetter had made an error. When Marshall saw the typo he exclaimed, "Leave it alone! It's great, and right on target!" Now there are four possible readings for the last word of the title, all of them accurate: "Message" and "Mess Age," "Massage" and "Mass Age."' < http://marshallmcluhan.com/common-questions/>

but also to the work of editors accessing primary sources from the screen of their laptops, and, by extension, to the consumption of digital editions instead of printed ones.

In 1960 Roman Jakobson returned to the Shannon-Weaver model, using the various components of the communication act in his functional analysis of language, and to the set of agents devised by Shannon and Weaver he introduced a fundamental correction by adding a two new components: the code, which was only implied in the Shannon-Weaver model, and the context where the communication takes place. The passage from the signal and information theory to communication and linguistics made the revised model extremely influential in the Humanities and Social Sciences. The classical communication theory assumes that senders and receivers share the same code and they are equally capable of coding and decoding the message. However, when we apply the model to natural languages, this is demonstrated to be impossible both in theory and in practice. In his *Course de Lingusitique Générale* published posthumously by his students in 1916, Ferdinand de Saussure introduced the fundamental distinction between *langue* and *parole*, which undermines our confidence that two people can speak and understand exactly the same language, as the personal performance, the *parole*, necessarily implies an element of individuality and therefore a certain distance from the established, shared rule signified by the *langue*. Differences in languages are both personal and social, and can be influenced by external factors. Since the Sixties scholars of sociolinguistics elaborated their classification of multilingualism, stating that in any given national language elements of multilingualism could be found, and that the following stratification and variation can be singled out: diachronic (language varies in time), diatopic (language varies in space), diastratic (language varies according to the social class of the speakers), diamesic (language varies according to the medium used for the communication), and diaphasic (language varies according to the style and rhetoric of communication).[59] All of these types of variation affect communication as well as textual transmission, from the work of the scribes and typesetters to the editorial work of contemporary scholarly editors. In fact, no matter how accurate the editing is, factors like level of education, distance in time and space in relation to the text being edited, the use of computers and the expectations of the scholarly community and the editorial textual theory all deeply and insidiously influence the editions themselves. No two editors will produce the same edition, in the same way that no two scribes will produce identical copies. When considering textual transmission, however, we also need to add other level of codification that goes beyond the ones of natural language. The same non-coincidence of codes affects the decoding of the script or the writing conventions, the iconographic

[59] Cf. Stewart (1968) and Coseriu (1981).

programme of the document and other aspects of the codicological and bibliographical codes (McGann, 1991) resulting in alteration of the message.

As for the context, we have to consider factors like where the document was produced and for which audience, the context of the creation (the authoring) of the text to begin with, as well as the context in which the reading takes place and how different they are with respect to what was envisaged by the author, and so on. For instance, serialized novels published in magazines during the Nineteenth Century assumed a different type of reading with respect to the one offered by volumes in which some of them were later collected, a fact that has a not negligible effect on the understanding of their structure and plot.

This rapid excursus into communication models, media and sociolinguistic theories demonstrates how texts are complex, data-rich "things" and their transmission can be affected by many factors. It also demonstrates that variation is not an accident of text transmission but a necessity, and that each variation can be traced to differences in the channel and its noise and on the inevitable variations of the code. The study of such variation can be a very rewarding scholarly approach indeed, as the work of Cerquiglini (1999), for instance, testifies. The goal of the editor is then to reconstruct the communications steps, working 'backwards' in an attempt to understand the relations and dependencies of the extant documents (the witnesses) that constitute the so-called tradition of a text, making some sense of the variation among them. In some cases, this 'making sense' means combining readings from different witnesses, discarding the ones considered erroneous according to the reconstructed genealogical dependence of the extant witnesses; in other cases, it means accepting that texts survive in different versions and that whatever the intention was of the author who originally produced the text, this authorial text does not exist anymore, has not existed for some time, and therefore has not been read by anybody since.

3.2 Theories and models of transcriptions[60]

Text transmission happens mostly in writing: whether it is matter of historical transmission (scribal copies, re-prints), or modern transmission in the shape of scholarly editions it all begins with a transcription.

[60] An earlier version of this section has been published in *Scholarly Editing* (2014), 34, availaible at the address <http://www.scholarlyediting.org/2014/essays/essay.pierazzo.html>.

In the past few years a few contributions have attempted to model transcription in highly theoretical way, trying to understand and to model the cognitive process that makes people recognize a given sign inscribed on a surface, map it to a mental meaning connected to that or similar signs, and then inscribe another sign with the same symbolic value on another surface, whether analogue or digital. These speculations may seem marginal for editorial discourse, but transcription is the first act of every editorial endeavour and, in many cases, it represents a very common digitization method, in the sense that it makes the chosen selection of the verbal content of documents digital and processable. Because of this a brief account of these debates and a discussion of their scholarly implications may be of interest.

The first contribution to this debate was produced by Claus Huitfeldt and Michael Sperberg-McQueen in 2008. The authors describe their work as an attempt to give 'a formal account of transcription as it is performed in scholarly editing and in the creation of digital resources' (p. 295). In particular, they are interested in the relationship between a document and the transcription of the document, and what it means '[w]hen we say […] that a particular resource is a transcription of a particular work'. They state that the purpose of making a transcription T of an exemplar E is to make a 'representation that is easier to use than E. For example, T may be easier to read, or easier to duplicate' (p. 296); when they declares that T is 'easier to read' than E, they then imply that T is a model of E as simplification is one of the defining characteristics of models. In the simplest case, in order for T to be a transcription of E, T must contain "the same sequence of letter, spaces, punctuation marks, and other symbols"; however, as exemplars (Es) are physical objects and letters, spaces and punctuation marks are abstractions, the latter cannot be 'contained' in the former except in a metaphoric sense. Adopting a terminology used to define instances of natural languages, they define marks as 'tokens' 'insofar they are instances of types'. Consequently, the relationship between T and E is summarized as follows by Caton (2013):

> under a set of reading conditions R, marks of E can be seen as a sequence of tokens each instantiating a type. So, abstractly, a document E is a sequence of types, and if under the same set of reading conditions R a document T can be seen as a token sequence whose corresponding type sequence matches E's type sequence, then E and T are t_similar'.

In the article of 2008, tokens are intended at letter level: for instance, the word 'sees' is composed of four tokens and two types ('s' and 'e'); however, in a successive contribution of 2010, Huitfeldt *et al*. postulate the existence of composite tokens at word and sentence

level. But what happens if the word 'sees' has been written in a particular document with a long s, 'ſees'? In this case, there are still four tokens, but how many types? Two ('s' and 'e') or three ('ſ', 'e' and 's')? In other words, how shall we treat the long 's', as an allograph of 's' or as a separate grapheme? The question of how to map the marks on the document page to discrete elements of the modern writing system is an open one (Driscoll 2006). Huitfeldt and Sperberg-McQueen debate over ligatures and abbreviations: are they different tokens or allographs? This is where the role of reading with its interpretational roles (and therefore the Reader as an d Interpretant, we may add) is conjured: 'a reading of a document interprets the marks in that document as instantiating a sequence of types' and '[t]he reading identifies some marks as constituting tokens' (p. 304).

Subsequent contributions (Sperberg-McQueen *et al*. 2009 and Huitfeldt *et al*. 2010) have attempted to extend the model at word, sentence and other structural levels, making more explicit the connection between the theory of transcription and the theory of encoding. Paul Caton has responded to these contributions on various occasions. In his presentation at the Digital Humanities conference of 2010 he argued for a more precise formulation and quantification of what is left behind in transcription, what is lost in the act of transcribing. He argues for the introduction of the concept of 'modality' which is the recognition of the fact that sometimes tokens 'display something that is in excess of – or deviant from – the norm'; for instance a *presentational modality* is to be recognized in tokens that are in italics, bold or underlined; an *accidental modality* in tokens that are recognized as erroneous, and a *temporal modality* which involves the effect of time on the token sequences. Caton then argues for text encoding, and in particular for TEI-based encoding, as a way to record such modality which is often lost in transcription. In his intervention of 2013 at the Digital Humanities Conference, Caton presents his take on what he calls 'Pure Transcriptional Markup', making a distinction between elements that wrap objects and ones that encode a property. His pure transcriptional markup substitutes or incorporates letters-tokens into elements-token in order to prove the theoretical similarity between an Exemplar E and an encode Transcription T.

What is the value of these speculations for the (digital) editor? Modelling is certainly a defining activity of the Digital Humanities and of any analytical endeavour, particularly if the latter takes place in a digital environment; and while models may or may not have an immediate application to projects and actual work, their epistemology is still valid. By modelling we learn how we know what we know. We learn to step back and look at things we take for granted from a different point of view, the point of view that allows us to know. For instance, by modelling and theorizing transcription we learn that transcription is a complex set of interpretative activities and not the mechanical application of criteria. We

learn that 'transcription necessarily involves the application of competent intelligence, and the selective suppression and alteration of information' (Huitfeldt and Sperberg McQueen 2008, p. 301); we also learn how to quantify and qualify the loss of information in textual transmission. It is useful here to re-introduce the distinction between *modelling-of* and *modelling-for* made by McCarty (2005) and discussed in the previous chapter. The model of transcription defined by Huitfeldt *et al*. and Caton is a model-of, but has the potential to be a model-for. According to McCarty 'a good model can be fruitful in two ways: either by fulfilling our expectations, and so strengthening its theoretical basis, or by violating them, and so bringing that basis into question' (McCarty, 2008, p. 395). The model of transcription elaborated by Huitfeldt *et al*. and Caton attempts to do both: on the one side it tries to account for the phenomenology of transcription, and on the other side it questions our beliefs that transcription is 'obvious' and 'trivial' (2008, p. 295). However, the Huitfeldt *et al*./Caton model has some limitations, and seems particularly weak when it tries to incorporate markup into transcription. While it is understandable that these scholars have tried to reconcile a theory of transcription with a theory of encoding, since in most cases in the Digital Humanities people transcribe and encode simultaneously, they are still two different domains and it would be probably more useful to keep them separate. Other aspects of transcription could have been taken into account instead. For instance, the model is still quite rudimental when it comes to the mapping between a mark on the document and the corresponding alphabetic character. When is an allograph an allograph and when does it become a different character? Are minuscules and majuscules the same characters or not? We know that sometimes we transcribe minuscules as majuscules and vice versa. What is the phenomenon according to which we distinguish 'u' and 'v' within a manuscript that uses them interchangeably? Do ligatures create new characters? The answer to all these questions is 'it depends', and it depends, principally, on the purpose of the transcription, but also from the type of document, where it was written and when, the level of understanding of the editors, their beliefs, the critical hypothesis, and so on. However, these answers need to be integrated into the model. To be fair, some of these have been considered, and in particular Caton's modality seems quite promising, but more could be done.

Connected to the models just analyzed, but developed independently, is Paolo Monella's proposal of a 'sausserean' approach to the declaration and transcription of characters (2013). Monella observes how signs that look the same may mean different things in different documents. For instance, in a document that only makes use of periods and semicolons as punctuation marks, the semantic coverage of semicolons will be very different with respect to another document that makes use of semicolons, periods and commas; variations of this type are found in all Medieval manuscripts. With reference to Saussure, for

whom any sign has a meaning which is devised only within its own semiotic system, Monella claims the necessity of declaring the meaning of each of the signs deployed by each transcription and consequent edition, in order to make sure that the equivalence of the sign does not hide a dis-equivalence of meanings.

In the same respect, an interesting perspective is offered by the *Digital Resource and Database for Palaeography, Manuscript Studies and Diplomatic* (DigiPal) project. In its attempt to define the components of handwriting and their features, the project team DigiPal, and in particularly Peter Stokes, has published a series of blog posts in which a model for describing handwriting has been proposed. According to this:

> Starting with the most concrete, a *graph* is an instance of the letter as written on the page. In contrast, an *idiograph* is 'the way (or one of the ways) in which a given writer habitually writes' a given letter, and an *allograph* is a recognized way of writing the given letter, as shared by a group rather than being distinctive of any given individual […] A *grapheme* is a letter-entity as an idealized, abstract, discrete, unit of a writing system. Also useful is Unicode's *character* to refer to the intermediate level between grapheme (which by definition has no physical form) and allograph. Thus the grapheme <a> has (at least) two characters, 'capital' **A** and 'small' **a**. The second of these has many allographs, one of which is Insular **a**, this is normally found in the script known as Insular cursive minuscule (Stokes, 2012, p. 383).

This complex classification tries to answer the question: how is it that when we see any of the 'marks' below (fig. 11) we recognize all as being an 'a? They do not look the same, still we do not question the fact that, if we have the expertise and knowledge to recognize them, we will transcribe as letter 'a'.

a	A	[61]	[62]	[63]	**a**	𝔄	A	*a*	𝐴

FIGURE3.1: LETTER 'A'

Some of the 'a's are majuscules, some are minuscule, some belong to the same font, some come from medieval manuscripts, but at some level of abstraction they all are 'the first letter of the Latin alphabet'.

Transcribing is an editorial act that simplifies and regularizes the phenomenology of the page. It is never mechanical or objective, but, like any other modelling activity, is guided and interpretative. The process of digitization through transcription, followed by the process of encoding such transcriptions, accompanied by the expressiveness provided by the TEI which allows for a very deep analysis of the transcribed signs; all of these together have opened a new field of study. Transcription, normally thought to be a preliminary operation to the real editorial act, has revealed itself to be a complex, interpretative operation, the mechanics, implication and hermeneutics of which are not yet completely understood but promise to be fascinating and scholarly engaging.

3.3 Returning to the Document: the Success of Digital Documentary Editions

The attention devoted to transcription can be related to a newly increasing attention devoted to documentary editing, as transcription almost imperceptibly slides toward the edition itself, until the only difference between the two endeavours is on the private nature of transcription and the public statement made by the edition (Pierazzo 2011, p. 464). This attention to the documents can on the one hand be related to uptake of theories of the texts that puts documents at the heart of the editorial care, such as the social theory of the text elaborated by McGann and McKenzie, new philology and genetic criticisms, and on the other hand to the emergence of a new genre of publication: the digital scholarly edition.

---

[61] CCCC 178: p. 416, Stokes *et al*., 2011-13, Online (digipal.eu/digipal/page/120/?vector_id=OpenLayers.Feature.Vector_926).

[62] St. Gallen, Stiftsbibliothek, Cod. Sang. 189, p. 76 – *Eucherii Instructiones; Isidori liber differentiarum; S. Hieronymus super Daniel* (<http://www.e-codices.unifr.ch/en/list/one/csg/0189>). Reproduced with permission of the Stiftbibliothek St.Gallen.

[63] CCCC 12: 36r: Stokes *et al.*, 2011-13, Online (<http://digipal.eu/digipal/page/59/?vector_id=OpenLayers.Feature.Vector_350>).

Now that digital scholarly editions have been in existence for nearly twenty years, we have seen the relative success of digital resources aimed at presenting documentary editions, with much less success for those that present critical editions.[64] Influential models for documentary editions such as *The Walt Whitman Archive* (Folsom and Price, 1995-), *The William Blake Archive* (Eaves *et al*., 1996-), the *Electronic Beowulf* (Kiernan, 2011), and, more recently, the *Samuel Beckett Digital Manuscript Project* (Van Hulle and Nixon, 2011) are not counterbalanced by similarly influential digital critical editions, an exception being the electronic edition of the Canterbury Tales (Robinson 1996; Solopova 2000) and other publications of the *Canterbury Tales Project*. The reasons for this unbalanced production are numerous and diverse.

Although the TEI was developed based primarily on the principles of the OHCO model, for which only immaterial features have right of citizenship in the encoding (cf. Chapter 2), it has always supported documentary editions at a deep level of granularity, and this tendency has been seconded even more with the introduction in 2011 of a new approach to text transcription based on the documentary evidence. This new approach allows editors to transcribe and edit their text using spatial and material concepts (zones and surfaces) as a base for their work, breaking *de facto* with the idea implied by the OHCO that only structural and immaterial features determine 'what texts really are'. Nonetheless, the TEI has preserved the possibility of editing the text from a textual point of view, which could be done alongside the documentary encoding or on its own, and this is exactly what is being attempted by the edition of Goethe's *Faust* (Brüning, *et al*. 2013). The dual approach of the TEI recognizes that editors and therefore editions may have different purposes: some editions may aim to reconstruct the text, others to analyze the document in depth, and others again to learn from both perspectives. This evolution of the TEI has taken place in the context of the increasing fortunes of the documentary edition hinted above, which, from a textual scholarship point of view, is a very recent development. This, it is argued here, is strongly connected with the adoption of the digital medium. As late as 2007 Hans Walter Gabler called for renewed attention to be devoted to documentary editions, claiming that the digital environment was most suitable for this type of edition, and promising that this change would be worthwhile. This solicitation was particularly aimed at Anglo-American scholarly editors, who have traditionally considered documentary editing a 'lower' form of editorial endeavour, an indication of which is given by the fact that documentary editing is also called, in particular in the US, non-critical editing. Non-critical implies but does not necessarily mean *un*-critical, as we were harshly reminded by Kline and Perdue in 2008:

[64] By 'success' we mean the number of editions produced by each of these two approaches.

> The documentary editor's goal is not to supply the words or phrases of a vanished archetype but rather to preserve the nuances of a source that has survived the ravages of time. Documentary editing, although noncritical in terms of classical textual scholarship, is hardly an *un*critical endeavor. It demands as much intelligence, insight, and hard work as its critical counterpart, combined with a passionate determination to preserve for modern readers the nuances of evidence. (p. 3)

In reporting this passage, Eggert (2009) does not fail to notice both the 'energy' and 'the note of continuing defensiveness' of this statement (p. 183), clearly marking the feeling of being considered second-class editors. And this perception is not unjustified, as declarations such as the following one by Tanselle clearly demonstrate: 'without further discussion […], despite the usefulness of the reproduction and transcription of the texts of documents, the attempt to reconstruct the texts of a work is a more profound historical activity' (1989, p.38).

Only a few years after Gabler's wake up call, however, in 2013 two scholars expressed their concerns over the increase of documentary editing, in particular in a digital environment. Peter Robinson, while conceding that 'one can welcome this attention to documents as a long overdue correction to the millennia-long concentration of scholarly editors on the work rather than on the document', he nevertheless sees 'dangers', as 'if we make only digital documentary editions, we will distance ourselves and our editions from the readers'; furthermore, he maintains, 'such editions [digital documentary editions], with their narrow focus on editor and document, fall far short of achieving the potential of editions in the digital world' (Robinson, 2013, p. 127). Franz Fisher has presented a not dissimilar case: although not criticizing documentary editions *per se*, he argues that proper critical editions are much more worth pursuing than diplomatic editions of the sources: as he declares: 'all texts are equal, critical texts are more equal than others' (Fisher, 2013, p. 90).

The two episodes that in Robinson's view mark the (dangerous) rise of documentary editing are the introduction of the new encoding model of the TEI mentioned above, and the publication of the *Jane Austen's Fiction Manuscripts Digital Edition*, which follows a page-by-page transcription module rather than a more semantic, paragraph-by-paragraph one (Robinson, 2013, p. 126). Editors have always had to face the question of what to edit, texts or documents (or texts of works vs. texts of documents);[65] in print, they have solved this mostly by mostly choosing texts (that is, texts of works) over (texts of) documents, given the cost and the traditionally relatively smaller impact of diplomatic editions which are more

[65] The terminology has been established in Tanselle 1989 and readapted by Robinson 2013.

difficult to read than critical and reading editions. This choice has been almost inevitable for works for which many witnesses survive: who indeed, apart from their editors and possibly a couple of libraries with a very large budget, would be interested in buying seven hundred versions of Dante's *Commedia*, one for each of the surviving witnesses? And if only a handful of them are chosen, which ones would have been selected? And which one would have become the one to read in high school? Historically, in cases like these, the only practical solution has been to reconstruct a version of the text which corresponds more closely to the theoretical assumptions of the editor,[66] and to serialize the rejected variant readings in the apparatus. The result is then the provision of a clean, reading edition, where the variants are conveniently marginalized at the bottom of the page or at the end of the volume, but which, according to many, gives a false impression of stability and purity of texts that are otherwise 'rough' and complex.[67] While this was (and – in many respects – still is) the most practical solution for texts with a large tradition, however, it is not necessarily the best solution in other cases, such as, for example, when only one witness survives; however, a sort of critical edition has traditionally still been provided even for texts preserved in a single witness, with the provision of a clear reading text, with internal variations and apparent errors appropriately set aside, in order to avoid disturbing the reading experience. This is the case of editions of Jane Austen's *Juvenilia*, and other works she did not publish in her lifetime. These were first published by R.W. Chapman from the 1920's; the text Chapman printed is a clean, 'final' version, with all the authorial workflow that can be reconstructed through a deep investigation of the original source material relegated to notes at the end of the book. This has been the only way readers have known these particular works, at least until the online publication of *Jane Austen's Fiction Manuscripts Digital Editions* in 2010. In cases like this, it is arguable that the edition of the manuscripts as critical text (texts of works) is the only sensible way to edit the material. Certainly the provision of a reading text is of fundamental importance to the general readership. However the edition of the text as document is also justified on scholarly grounds, and the general public may also find it interesting. This is demonstrated by the fact that the publication of the digital edition has generated a huge debate around Jane Austen's authorial habits, with readers and scholars

[66] This could be, for instance, the oldest preserved, the best survived, the closed to the authorial intentions, the latest editions, etc.

[67] 'The smoothness of the clear reading text, a hall mark of critical editing, in effect denies us an immediate awareness of the actual roughness of the textual record, and textuality itself' (Bryant, 2002, p. 27). On the same advice is Zeller who deprecates the 'traditional misleading smooth typography of a clear text' (Zeller, 1991, p. 40).

alike expressing their disbelief in discovering that she was not the polished, organized writer that was presented through the edited pages.[68] The reception of Hans Walter Gabler's edition of *Ulysses* shows the same astonished reactions: by presenting the documentary evidence of the autograph manuscripts in a synoptic format, he shocked his audience, which was not used to being exposed to the messiness of the authorial work.[69] Indeed, the same patterns hold for many literary works: even when the survival of only one or a handful of witnesses could have made the edition of a documentary edition viable and enriching, this has rarely been done. And when it has been done, it has been often attacked as 'dangerous': Paul Eggert beautifully accounts for the polemic lead by Fredson Bowers, champion of the authorial intention and critical editing, against *variorum editions* and documentary editions, an argument continued by Thomas Tanselle at the expense of historical documentary editors (Eggert, 2009, pp. 180-84).[70] The battle between critical editing and documentary editing[71] has been fought on many fronts: disciplinary, theoretical, methodological, and financial. The point here is that the attitude towards one or other editorial framework (documentary or critical) has always been coloured by strong theoretical positions: the question has very rarely been which editorial framework was best for the type of document under consideration,[72] but rather which was the theoretical point of view of the editor. Historically, this conflict has tended to pitch historians against literary scholars,[73] with the former more likely to produce documentary and diplomatic editions rather than critical ones, possibly obeying the unwritten rule that the role of an editor-historian is to make available historical sources 'as they are'. In the US this editorial approach was actively encouraged by the NHPRC (National Historical Publications & Record Commission) (Kline and Perdue, 2008, p. 12). From a theoretical point of view, advocates of documentary editions accused critical editors of hybridising their texts from many sources, building texts that never existed and hence altering the evidence from the

[68] See <http://www.bbc.co.uk/news/entertainment-arts-11610489>.

[69] See the story beautifully told by Paul Eggert, *Securing the Past*, pp. 164-178.

[70] See also Kline and Perdue, 2008, pp. 1-34.

[71] Or the battle of the 'intentionalists' against the 'materialists', to use the terminology by Bryant, 2002.

[72] By 'best' it is here meant which editorial framework is more able to respond to the editors' research questions. Fredson Bowers, in spite of being a strong champion of critical eclectic editing seems to hint toward this direction, as in his essay of 1991 he declares that it might be advisable in some cases to provide different editions for different types of readers, with documentary editions offered to scholars, and critical editions for the general public (Bowers, 1991, p. 55 and p. 61).

[73] 'For the sake of convenience, they [i.e. documentary and critical editing] came to be known as historical and literary editing, a division many regarded as unrealistic and unfortunate' (Kline and Perdue, 2009, p. 11).

past;[74] this attitude also characterizes the theoretical positions of the New Philology, the arguments of which are based in turn on fundamental critiques advanced by Joseph Bédier (1928) regarding the Lachmannian method (Greetham, 1994; Cerquiglini, 1999). Another aspect of this conflict has focused on methodological issues, and in particular the rigour and accountability of criteria for transcription, with literary scholars accusing historians of having a cavalier attitude towards their transcription criteria. Such episodes were 'a wake up call' for documentary editors, who started a period of re-examination of their methodological approaches and procedures.[75]

Once again, the discussion has covered what readers want: a reading edition that critically combines the variants of a possibly extensive textual tradition, or a documentary edition (or editions), possibly accompanied by many facsimiles to allow inspection of the original documents by themselves - if they ever wanted such a thing, as Vanhoutte (2009) asks. In this latter case the documentary editions have generally been considered unsuitable for the wider public, unreadable, or both. Genetic editing and genetic editions could also be included here: indeed, genetic editions resemble documentary editions a great deal, with the addition that the former also try to embed information about the different phases of writing and rewriting of the manuscript. The obscure, intricate symbols that necessarily characterize such editions in print are perhaps the principal reason for their cold reception by the academic community.[76] Economic factors have also been considered: the cost of producing a documentary edition has often been seen as unjustified by the small number of prospective readers except in the case of very important, unique documentary sources.[77] In the US, however, documentary and diplomatic editions have been used to a great extent for the publication of historical documents of the Fathers of the Nation (Jefferson's and

---

[74] The criticism to critical editor is mostly implicit in Kline and Perdue (2009) (cf., for instance, the mention to the 'vanished archetype' at p. 3), however it is mostly the principal argument of German editorial theories (cf. Shillingsburg, 2006, pp. 173-188 for an account of the debate).

[75] This criticism lies at the heart of the polemic between Tanselle and the ADI (Kline and Perdue 2008, pp. 19-22).

[76] Grésillon, 1994, pp. 195-202. A particularly good (or bad) example of this is represented by the genetic edition of *Hérodias* by Gustave Flaubert, edited by Giovanni Bonaccorso et al. (1991); this edition encompasses nine different types of arrows to mark the location of interlinear and marginal insertions belonging to four different revision campaigns. But see also Hunter 2007, pp. 118-120, for complaints about similar issues in documentary editions of early modern texts.

[77] See for instance the diplomatic edition of the *Domesday Book* by Abraham Farley and Henry Ellis (1783).

Washington's, for instance), but this fact, instead of encouraging the diffusion of such an editorial formula, has increased the theoretical and disciplinary divide, with literary scholars consistently producing critical editions based on the editorial framework of the copy-text and historians championing documentary editions. A documentary edition of works of literature in the Anglo-American world seems to be rarer than, say, in France, Germany or Italy,[78] and this applies also to authors for whom we have works transmitted by only one witness, such as, for instance, the *Juvenilia* or the unfinished novels of Jane Austen, as discussed above. French-style genetic editions or German-style *Historisch-Kritische Ausgaben* are even more rare.[79] One can see, for instance, the edition of Austen's *Sir Charles Grandison* edited by Brian Southam, whose editorial choice of publishing as a type-facsimile has been rejected by Michael Hunter (2007, p. 75): 'For, painstaking as it is, it fails to replicate all features of the original – not only different handwritings and letterforms, but also ink blots, different methods of striking through words, or exact details of the layout, for which only a pictorial facsimile suffices'.

In Eggert's view the cultural hegemony exerted by the copy-text theory (the so-called Greg-Bowers-Tanselle line) can be held responsible for what he sees as a lack of interest in documentary editions by literary scholars. However, this methodological hegemony has been tolerated with increasing irritation by the scholarly community; the most famous reactions to this editorial approach are represented by Jerome McGann's theory of the social text and the rise in favour of New (or 'material') Philology. Both these theories focus on the importance of the material support: the 'bibliographical codes' of McGann (1991) can be paired, in a way, with the *codex* of Cerquiglini (1999) (Driscoll, 2010). These contributions were elaborated in the same period as digital editions started to be produced, so it might not be a coincidence that practitioners of innovative textual theories have looked to an innovative publishing medium to convey their editorial products. Richard J. Finneran notes that the advent of new technologies 'coincided with a fundamental shift in textual theory, away from the notion of a single-text "definitive edition"', remarking that while 'a traditional print edition is able to accommodate this new thinking in textual theory either awkwardly or not at all, digital technology is its necessary and inevitable realization' (1996,

[78] An exception is some of the editions sponsored by the Early English Text Society (EETS) which focus on Old and Middle English texts, covering works that have literary, linguistic and historical interest.

[79] One of the rare exceptions is represented, by for instance, Hans Walter Gabler's edition of *Ulysses*, which has been mentioned before, which is not, strictly speaking a *Historisch-Kritische Ausgaben*, nor a genetic edition, but an adaptation of their methodologies to a specific case. It is not surprising to note that such an edition was conceived in a country with a more favourable attitude toward this type of edition.

p. x). Deegan and Sutherland (2009) take a similar line, with the caveat that they describe the renewed interest in more complex considerations of textual transmission, consumption and dissemination as not being 'computer-dependent' but 'computer-convergent' (p. 64).

There are many factors that have made documentary editions more intellectually interesting and affordable in a digital environment. The availability and relatively low price of digital facsimiles as well as their ease of publication on the web have been often mentioned as one of the main reasons for the success of digital documentary editions. This new affordance has generated some over-enthusiastic statements, like, for instance, the one from Kevin Kiernan who declares that an imaged-based edition 'subsumes the purpose of a diplomatic edition' (Kiernan, 2006, p. 266). The same opinion is shared by Hunter, who would rather have seen a facsimile edition instead of the ultra-diplomatic edition of Austen's *Sir Charles Grandison* (Hunter, 2007, p. 75). These claims have already been refuted elsewhere (Pierazzo, 2011, p. 472-73), but it is relevant here to highlight one crucial feature of diplomatic editions, namely that historically they were used as a substitute for facsimile editions when the latter were impractical or too expensive. Kiernan's and Hunter's declarations imply that there is no need to produce a diplomatic edition once digital images are made available. Time has demonstrated that this assumption is wrong. In fact, the availability and the improved publishing opportunities offered by the different digital formats (CD-ROMs, DVDs and the Web) and the subsequent publication online of millions of digital facsimiles, has by no means made diplomatic and documentary editions redundant: quite the contrary. Far from being subsumed by image-based editions, digital documentary editions have *de facto* incorporated them, with the most common format for the digital documentary editions being represented by the side-by-side layout, with the diplomatic (or ultra-diplomatic) edition juxtaposed with its digital facsimile. To be fair, Kiernan suggests that the image-based edition is not an image-only edition, but that the facsimile occupies a central role in the provision of documentary editions, possibly accompanied by a transcription. This format engages the editor in a sort of competition with the facsimile where, as noted by Kathryn Sutherland, 'the editor is continually on trial, open to account and correction' (Sutherland and Pierazzo, 2011, p. 202). The dialectic relationship between the diplomatic edition and the facsimile representation, while demanding extreme editorial rigour, engages the users in close inspection of the transcriptions/translations enacted by the editor in a sort of imaginary, involving, competition.[80] The diplomatic edition alongside the facsimile provides

[80] On the concept of transcription as a form of translation: 'transcription of a primary textual source cannot be regarded as an act of substitution, but as a series of acts of translation from one semiotic system (that of the

the reader with a simplified, mediated, interpretation of 'what's on the page'; therefore, instead of being made redundant by the presence of the image, it represents a sort of map for the understanding and navigation of that image. Just as a map of any given area of the earth provides the user with a way to move around in an unfamiliar place, the diplomatic edition provides the reader with the tools to decipher the linguistic and bibliographic codes which inform the document.[81] Therefore, while the provision of high quality digital facsimiles has effectively made redundant the provision of diplomatic editions as a substitute for facsimile editions – since facsimile editions can now be provided easily –they have also revealed that diplomatic editions (and especially digital documentary editions) are scholarly tools and not surrogates, and are worthwhile precisely as scholarly, interpretative tools. Consequently, the more digital facsimiles of primary sources we offer, seemingly the greater the need for digital documentary editions.

The implicit recognition of this type of functionality has helped considerably the establishment of digital documentary editing among scholarly practices, but this alone is perhaps not enough to explain their sudden success. The provision of digital documentary editions based on text encoding, which enables the possibility of presenting the users with many outputs from the same source of encoded files, has made digital documentary editions a very powerful and versatile tool. The possibility of enriching an edition with more information than is needed for each of the outputs that will manifest it is considered a defining characteristic of paradigmatic editions (Chapter 1). This opportunity, while not solely found in text encoding-based editions, is certainly a feature of such editions, and certainly represents one of the reasons for their success. Text encoding has in fact enabled editors to have their cake and eat it: features that were once to be normalized without mercy to produce reading, critical (or syntagmatic) editions, can now be retained and simply switched on and off at leisure to please different types of audience. These functions have attracted the attention of scholars for whom documents and documentary evidences have a central role, namely New (or Material) Philology and Genetic Critics: scholars that recognize themselves as belonging to these editorial orientations were in fact among the first to embrace the digital medium (Driscoll, 2010; Lebrave, 2008). As paradigmatic editions are able to store effectively an indefinite amount of information, it therefore seems possible to remain faithful to the document (retaining for instance, odd spellings, abbreviations, scribal

---

primary source) to another semiotic system (that of the computer). Like all acts of translation, it must be seen as fundamentally incomplete and fundamentally interpretative' (Robinson and Solopova, 1993, p. 21).

[81] 'Bibliographical codes' is an expression created by Jerome McGann (1989). As McGann mainly refer to printed books, the expression 'codicological codes' has been proposed for application to manuscripts (Pierazzo 2013).

errors etc.), as well as to make concessions to the reader who is then given the opportunity to choose which version to read. This is the case, for instance, in the LangScape project, where the passages describing land bounds contained within the corpus of Anglo-Saxon charters are presented as semi-diplomatic, edited or linguistically glossed texts.[82] In this sense paradigmatic editions could also be considered 'fully digital' in the sense expressed by Patrick Sahle, who has defined digital editions – as opposed to digitized edition – as editions that cannot be printed without loss of information (Sahle, 2008). And indeed printing paradigmatic editions will mean losing their richness, as their content will be only statically serialized, one output after the other, instead of embedded within one another.

Documentary editions have been at the centre of a long lasting theoretical debate, as they have been identified as the ideal editorial framework of those scholars who reject the creation of critical editions based on multiple witnesses. However, *per se*, they are simply an editorial methodology which has no particular theoretical implication. This is demonstrated by fact that documentary editions have been serving different theoretical positions, for instance the 'best-text' method,[83] the social edition, historical edition, genetic edition and material philology, without demanding any one of these. So documentary editions do not constitute or do not aim to constitute a theory of digital editions:[84] they happen to have been chosen as their editorial format by scholars who hold a theoretical position, but not necessarily the same one. It is not surprising then that the elaboration of the new model offered by the TEI for the encoding of documents (that allows for a page-by-page and zone-by-zone transcription, as well as the embedding of information about the process of writing) represents the result of collaboration between scholars holding many of the above mentioned theoretical positions, as well as from different disciplinary areas.[85]

---

[82] See website and Stokes and Pierazzo (2009), pp. 203-38.

[83] As derived by Joseph Bédier's famous criticism of the lachmanian method (Bédier, 1928, pp. 161-196, 321-356).

[84] This is what Peter Robinson seems to fear the most, namely that the general adoption of documentary editions will translate into a theory of editing: 'The dominance of the document model of textual editing in the digital realm suggests that a theory of digital editions is emerging, based on page-by-page transcription of individual documents, which asserts that a digital edition should concentrate on documents alone' (Robinson, 2013, p. 126).

[85] The people who contributed to the elaboration of the new model are: Anne Bohnenkamp (Goethe Universität), John Bryant (Hofstra University), Aurèle Crasson (ITEM), Jean-Daniel Fekete (INRIA), Daniel Ferrer (ITEM), Hans Walter Gabler (Ludwig-Maximilians Universität München), Axel Gellhaus (Institut für Germanistische und Allgemeine Literaturwissenschaft der RWTH Aachen), Almuth Grésillon (ITEM), Claus Huitfeldt (University of

The reasons for the digital diffusion of documentary editions are then varied and diverse, but, ultimately their success is probably connected to the realisation that they can offer far more scholarly value than the 'non-critical' label and their supposed surrogate function suggests. Furthermore, digital documentary editions seem able to answer old and new scholarly questions that have never found a proper answer, in particular in print-based editions. Manuscripts and documents are information-rich sites and 'putting the text back in context' is bound to reveal much more about the text than if we considered documents as simple supports.[86] The adoption of documentary editions can fulfil the desire to close a gap in textual scholarship, now that the technology supports it and makes it a worthwhile pursuit.

The growing attention of the scholarly community to documentary editing has also highlighted the need for a new theoretical understanding of this format. Considered often the mere recording of 'what is on the page', documentary editing has not been given as much theoretical attention as critical editing, with the exception of the animated discussion over objectivity in editing. In particular an animated debate has focused on the reflection of what it means to record what is on the page and how and if this should be kept separate from what is on the page actually means. This discourse is intertwined with the role of digital facsimiles within documentary editing as a way of objectively representing (or not) the real 'thing'. Both topics will be discussed in the following chapter.

Bergen), Dirk van Hulle (University of Antwerp), Jean-Louis Lebrave (ITEM), Wolfgang Lukas, Kenneth M. Price (University of Lincoln, Nebraska), Kathryn Sutherland (St. Anne's College, Oxford University). Members of the working group were Fotis Iannidis (University of Würzburg), Gregor Middell, Moritz Wissenbach (University of Würzburg), Malte Rehbein (University of Passau), Lou Burnard (University of Oxford), Paolo D'Iorio (ITEM) and Elena Pierazzo (King's College London).

[86] See Pierazzo and Stokes (2010) for a list of examples where the documentary edition represents a scholarly necessity more than a simple choice of convenience.

4. What's on the page? Objectivity and Interpretation in Scholarly Editing

4.1 Record and Interpretation

It seems clear, as demonstrated by the developing theories of transcription and the analysis of the rising fortunes of documentary editions in digital form, that the pursuit of digital documentary editions, which allow for the representation of the source document to a level of detail unimaginable in print, opens up a new series of theoretical issues. One of them is the necessity or possibility of separating the record of what is on the page from the interpretation that accompanies the availability of digital facsimiles, a discussion that is not born with digital, but has been amplified by it.

As we have seen, paradigmatic editions allow for the storage of several layers of information, and this, it might seem, enables the ideal separation between the record of what is on the page, and the interpretation of its meaning. This capability seems to suggest the theoretical framework originally outlined by Hans Zeller, which has strongly influenced editors in continental Europe in particular (Zeller, 1995). In his vision, in order to make 'securer and more objective' editions (p. 19), it is essential to operate a 'methodological separation of the record from the treatment of the record', where the record (*Befund*) is 'the extant material encountered by the editor' (p. 21), or to use our terminology, the documents that compose the tradition of a work, and the treatment of the record is the interpretation (*Deutung*), or the editorial work of crafting the (critical) text. The recording of what is on the pages should ideally be done without any interpretation, which is to say in the most objective way possible, and this is because "[t]he edited text is supposed to provide the *basis* for interpretation of the work in question. Therefore, the edition *itself* should be as free from interpretation as possible" (Huitfeldt, 2006, p. 194). The call for more rigorous methodological behaviour exhibited by the editor, who is not to be seen as a superior being able to provide superior readings and interpretations of the authorial intention (Zeller, 1995, p. 23), but as scrupulous and accountable recorder, becomes a somewhat rigid distinction between objective and subjective editorial practices, a risk which is present in Tanselle's assessment of Zeller's position: '[t]he search for properly 'scientific' method has been perhaps the dominant thread running through the history of textual criticism… Too often, however, rigor of method has been equated with the minimization of human judgment: instead the two must be carefully distinguished' (Tanselle, 1995, pp. 18-19).

Objective editing is defined by Zeller as 'merely purifying an authorized text of textual faults', and this only when the latter is unambiguous (Zeller, 1995, p. 38). But to what

level of objectivity can editors aspire? And does the absolute definition of objectivity allow for the definition of levels? Tanselle is very clear: while a methodological rigour is something to aspire to, this should be clearly disconnected from objectivity. In fact, 'a coherent rationale of approach is properly a desideratum of textual scholarship, but any rationale of critical editing that seeks to limit (rather than to systematize) the role of judgment is not coherent, since by definition critical editing exists to draw strengths of human judgment as a mean of correcting the defects of documentary texts' (Tanselle, 1995, p. 19). However, what Tanselle really has in mind is that objectivity should have nothing to do with *critical* editing. In fact he then declares that the role of documentary editing is editing sources without altering them 'except for the compromises entailed in presenting them in a new physical form'. So while 'correcting defects of documentary texts' is a defining characteristic of critical editing, representing without alteration is the purpose of documentary editing. Without alterations means objectively, as the only exceptions to this rule are due to the change of medium.[87] It also means that the only output allowed to documentary editions is one that imitates the source itself. These principles are also at the heart of the 'System of Manuscript Transcription' written by Tanselle in collaboration with Vander Meulen (1999, p. 201):

> [B]y "transcription" we mean the effort to report – insofar as typography allows – precisely what the textual inscription of a manuscript consists of. Obviously a transcription cannot exactly reproduce the relative precision or carelessness with which handwritten letters are formed, or their relative sizes, or the amount of space between words and lines; but it can aim to record every ink or pencil marking of textual significance on the manuscript -- all letters, punctuation, superscripts, canceled matter, lines linking or excising passages, and so on.

Judgement is conceded, but only when it is required to disambiguate complex situations, as a programme of systematic alterations is unacceptable (pp. 201-202). As we see, objectivity that was chased out of the door of critical editing has come back through the window of documentary editing.

Transcriptions (and diplomatic editions that derive from them) have often been described as a record of what is on the page, and these have been opposed to the editorial interpretations offered by critical editions. However, the fact that an editor can truly transcribe what is on the page without making any assumption about meaning is arguable. Robinson (2013) convincingly demonstrates how unrealistic the attempt of transcribing what

[87] The changes due to the medium could be quite considerable, according to McLuhan (1964), but this is perspective totally absent from Tanselle's argument.

is on the page is by reporting and discussing a reading taken from a manuscript of Dante's *Comedia*. This, as reported by Bordalejo features the sequence 'durao' with a dot under the letter 'a' (Bordalejo, 2010, Online). He argues that even in the decision of transcribing what is on the page – i.e. "duṛao" –

> The transcriber has made a series of decisions. The first is that it is a text in Italian, written by an Italian scribe around 1340 and so using letter forms and conventions characteristics of Italian vernacular manuscripts of that period. This determines that the decision that the first letter be transcribed as "d" (and not as an "o", as it might have been in some scripts), the second as "u" (although in other contexts the same two minims might be transcribed as "n") […] Note that once that the first letter is identified as "d", in Italian the second letter must be "u", not "n". (Robison, 2013, p. 116)

A similarly instructive example is offered by a reading from Cod. Sang. 189 preserved at the Library of St Gallen in Switzerland, containing a series of Patristic works. At p. 76, line 5 one finds the following sequence:

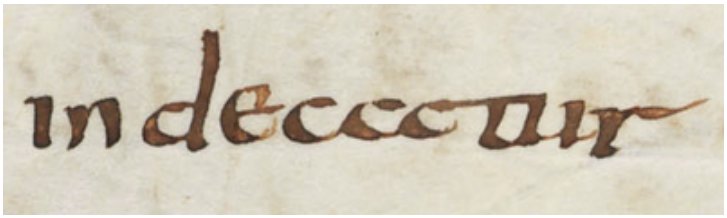


FIGURE 4.13: COD. SANG. 189, P. 76[88]

the reading of which can only be 'indecatur'. However, if we isolate the sequence 'ca' from its context, who would be able to tell that the three signs are actually two letters, a 'c' and an 'a' and not, for instance 'ccc' or 'ac'?

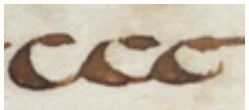


FIGURE 4.2: COD. SANG. 189, P. 76 (2)

It is only because we know that there is no word that contains three letters 'c' in a row (unless it represents the number '300' in Latin notation), and also because we know Latin and therefore we read (interpret) what surrounds the sequence, and also because we know

[88] St. Gallen, Stiftsbibliothek, Cod. Sang. 189, p. 76 – *Eucherii Instructiones; Isidori liber differentiarum; S. Hieronymus super Daniel* (http://www.e-codices.unifr.ch/en/list/one/csg/0189). Reproduced from *e-codices – Virtual Manuscript Library of Switzerland* with permission of the Stiftbibliothek St.Gallen.

that there is a special type of 'a' (the 'cc-*a*'),[89] that we can say that the reading is indeed 'indecatur'. Is this an objective reading? Or is 'indeccctur' an objective reading? What makes a reading objective? And is it influenced by the competences of the readers? Where do we draw the line between competency and interpretation?

4.2 Objectivity in Textual Editing

In their book of 2010 Daston and Galison draw a history of objectivity, choosing as a case study the production of images in scientific atlases; the fact already that the concept of objectivity (and 'subjectivity', its counterpart) has been 'invented' at some point (namely between the Eighteenth and the Nineteenth Century)[90] undermines its absolute value. In their book they examine the history of scientific image making and the evolution of scientific atlases as a publication genre, and they use them as tool for the analysis of the evolution and the moralization of the concept of objectivity. In their exploration they outline three types of epistemic virtues connected to objectivity and consequent forms of scientific observations and representation: 'truth-to-nature', 'objectivity' (mechanical or structural) and 'trained judgment'. In their examination they show how the scientific community has progressively moved away from the brutal mechanical 'truth' offered by photography – which they demonstrate to be less than objective and not able to represent the real truth beyond the mere appearance – towards the trained judgment operated by scholars, considered on the one hand a step away from objectivity, but on the other hand a step closer to the purpose of their research. A similar argument can also be made perhaps for editorial endeavours aspiring to be objective. But what are these types of epistemic virtues, how do they work and which of them is the type of virtue that is consciously or unconsciously pursued in textual scholarship and in particular in documentary editing?

The first type of epistemic virtue examined by Daston and Gallison is true-to-nature. According to this approach the researcher chooses the best, unblemished specimens of a particular entity (a bird, a stone, a flower) and then conflates them into an ideal representation. The value of such an artefact is that it stands for all others, and its ideal representation makes all members of the group recognisable as members of the group itself. It is clear as true-to-nature is a not a form of objectivity, but a pre-objectivity epistemic virtue

[89] For the cc-*a* in caroline minuscule, see Bischoff, 1990, (pp. 112-15). Such letter form can also be found in other early medieval scripts for which see, again Bischoff, 1990, (pp. 106 and 110).

[90] The first to use the opposition objectivity/subjectivity is Kant, but he uses them as an opposition universal/particular more than reality/mind.

(its practice can be traced back to the Seventeenth century), but the practice and value of which has not ceased to exist because of the 'invention' of objectivity.

If we transpose this virtue to the environment of textual scholarship we will find some analogies with both the theory of the best manuscript and of eclectic critical editing. In the case of the former, it is the act of selecting the best specimens that creates the connection. For the latter is the act of conflating features of different exemplars into a single ideal representative that can be considered analogous to the act of creating a critical edition from different witnesses. It is worth noting here that by eclectic critical editing we mean most editorial models that allow for combinations of readings for the preparation of a single edited text, the rationale of which is not relevant for our argument, as long as there is one. In both cases the edited text 'stands' for the work in one way or another: in the case of the best manuscripts because it represents its best historical embodiment; in the case of eclectic (critical) editing because it represents the best effort of the editor in reconstructing the text of the work.

The second typology analysed by Daston and Galison is mechanical objectivity, according to which it is better to provide the mechanical representation of many imperfect specimens instead of only one which is idealized and artificially constructed. Such specimens in their imperfection have the advantage of being real, and not idealized reconstructed representations. Historically, the introduction of this approach coincides with the invention of photography which was able to provide mechanically produced, and therefore allegedly objective, representations of the reality. We could then connect this type of objectivity with the provision of facsimile editions, the practice of separating record from interpretation, and New Philology for different reasons. While the connection with facsimile editions seems self-explanatory because of their reliance on photography, the reasons for connecting mechanical objectivity with the other two approaches needs some explanation. In the case of the distinction between record and interpretation championed by Zeller, the connection is that mechanical objectivity emphasizes the denial of the interpreting self: in order to know the true, the 'blind sight' of the researcher has to become objective. In the case of New Philology it is the lack of choice among the specimens that allows building the relation, because imperfections (errors) of the exemplars are irrelevant: they exist and therefore they are truth, and truer than any corrected version. In this case each representation stands for itself and the concept of work as we and many other scholars have defined it is irrelevant, at least at the stage of the representation. Texts built according to the principle of mechanical objectivity may be put in relation with other texts under the umbrella of the work, but they do not have to be.

A subtype of mechanical objectivity is so-called structural objectivity, according to which the mathematical, analytical abstraction of the scientific laws that are the basis of the existence of specimens is an even a more objective representation than the one provided by mechanical representation alone. This type of objectivity is also defined as 'objectivity without images' (Daston and Galison, 2010, p. 253). Sight can be deceptive, but mathematics is not, the reasoning goes, therefore the abstraction of mathematical laws and logical relationships among elements provide a scientific (objective) representation of this reality. In this case the connection with textual scholarship's practices is weaker than for the other cases, but one might think about stemmatology and phylogenetics, with their attempts to provide a logical representation of the extent witnesses. In the case of the Lachamnian method, furthermore, the connection can be also extended to the role of *recension* and the mechanical choices it provides to the editor.

The last type of epistemic virtue outlined by Daston and Galison is trained judgement. According to this approach the expert is allowed to smooth the raw data provided by nature for the purpose of study. Such smoothing has to be done within rigorous methodological limits by a competent expert, able to mediate the available data to the users. However, the unblemished data should also be provided to the user who will be called upon to judge the work of the expert. This last approach can be connected to all textual scholarship practices that emphasize the importance of providing access to the data to the readers, for instance within a critical apparatus. However the closest match is perhaps provided by those digital editions that provide access to the (XML) sources as well as to the edited texts.

Daston and Galison summarize the characteristics of the considered epistemic virtues with the following table (p. 371, slightly revised for this publication):

Epistemic virtue	From before 1820 **True-to-nature**	After 1820 **Mechanical objectivity**	After 1920 **Trained judgment**
Persona	Sage	Manufacturer	Expert
Representation	Reasoned	Mechanical	Interpreted
Practice	Selection, synthesis	Automated transfer	Pattern recognition

Ontology	Universals	Particulars	Families

TABLE 3.1: EPISTEMIC VIRTUES IN TIME (DASTON AND GALISON, 2010, P. 371)

The relevance of the application of Daston and Galison findings to textual scholarship relies on the centrality of representation in our discipline and on the fact the 'search for properly "scientific" methods has been perhaps the dominant thread running through the history of textual criticism' (Tanselle, 1995a, p. 18). The reason behind the quest for a scientific method is that interpretation pervades every aspect of the editorial work, from the realisation that 'a set of marks on the page … are text' (Bordalejo, Online, as cited by Robinson 2013, p. 116), to the moment the editor rejects a variant reading as erroneous and supplies a conjectural alternative. This fact is not unknown to Zeller himself, who is not shy about raising the 'subjective nature of editing' (Zeller, 1991, p. 19). There is also interpretation in the way an edition is laid out and presented: Zeller again reports the case of a German edition of Heine that in order to follow the classic genre aesthetics, ignores the 'blending of forms and genres' sought by the author (ibid.). There is interpretation also in the way digital photographs of manuscript pages are taken: the privileging of the page over the double page or the gathering as a digitization unit represents one interpretation of what manuscripts are and are made of. However, while it is true that there are different degrees of interpretation in transcribing a particular set of minims as 'u' and not as 'n', or in supplying a conjectural reading, it is also true that one cannot establish the exact point where the interpretation level is zero, and therefore objective, and when it is non-zero, and therefore subjective. Claus Huitfeldt recognizes how 'representation and interpretation' are 'points located at the opposite ends of a continuum' and the role of the editor is then 'to find somewhere along this continuum suitably clear demarcation lines, which allow us to decide, in particular cases and classes of cases, what is interpretation and what is representation' (Huitfledt, 2006, p. 195). According to this definition, then, the demarcation between representation (objectivity) and interpretation (subjectivity) is the result of a conscious editorial decision: a definition of objectivity which is quite unusual, but perhaps more respectful of the reality of editorial work. However, if we declare that we must decide what is objective, the objectivity will itself become the result of a subjective act (namely the decision), and therefore potentially indistinguishable from subjectivity itself. This argument is at the core of another type of objectivity not included, except perhaps implicitly, in the overview proposed by Daston and Galison, namely the so-called 'social agreement', the first articulation of which can be traced, once again to Kant: 'If the judgement is valid for

everyone, provided only he is in possession of reason, its ground is objectively sufficient' (Kant, 1965, p. 645). A similar argument is made by Huitfeldt: 'we may conclude that there is such a thing as objectivity of interpretation: the vast majority of decisions we make in this realm are decisions on which all (or most) competent readers agree or seem likely to agree' (Huitfeldt, 2006, p. 196). This sounds a very pragmatic (and civilized) form of objectivity: however one has to wonder how such an agreement can supposedly be assessed. As a full consultation with all or most competent readers is out of question, it would be probably a matter of self-assessment: as editors are competent readers they are able to imagine what their colleagues would say. But, again, how will competency be assessed? Will a formal qualification such as, say, a PhD provide this, or experience, or both? And will editors be able to assess their own competency? Will they be able to recognize if they are not competent enough? It seems that this form of objectivity is a rather subjective one. As already maintained in a previous contribution, it is arguable that the notion of objectivity is helpful or relevant when it comes to scholarly editing (Pierazzo, 2011). Any edition, documentary or critical, is an interpretative attempt by an editor to provide readers with a good explanation of the textual transmission of a particular work. And it is interpretative all the way down, from the reading and transcribing to the final publication, as the centrality of the Interpretant as core agency in the conceptual model seen in chapter 2 suggests. As Gianfranco Contini put it, the edition of a text 'is the most economic working hypothesis accounting for the data we posses' (Contini, 1986), or, again: 'a scholarly edition is, as every scientific act, a mere working hypothesis' (Contini, 1974, p. 369).[91] This understated definition of the edition has represented an ideological framework for Italian textual scholarship for a couple of generations; it does not leave any space to claims of objectivity nor to the universality of an *urtext* edition, but it rather presents the editorial work as an honest, time-specific achievement, and, as such, inscribed within the scientific tradition. Furthermore, this working hypothesis is bound to change over time as '[e]ach generation seems to hope for the best answer and finds the effort of the older generations to be inadequate. Soon it will be our turn to have failed' (Shillingsburg, 2006, p. 153); along the same lines Sutherland argues: 'all editorial theories imply the authority to represent or speak as the text; and all are ultimately revealed as temporal and temporary protocols for interpretation' (Sutherland, 2013, p. 48). Then how can the claim for objectivity be reconciled with an ever-evolving editorial hypothesis?

[91] 'l'ipotesi di lavoro più economica che tenga conto dei dati' (Contini, 1986); and 'un'edizione critica è, come ogni atto scientifico, una mera ipotesi di lavoro, la più soddisfacente (ossia economica) che colleghi in sistema i dati' (Contini, 1974, p. 369). Notice that, as most Italian editors, Contini does not consider critical and documentary as ontologically different types of edition.

Despite these difficulties, Zeller's framework of the 'objectification of editing' has found vigorous supporters, in particular in the German-speaking world. Hans Walter Gabler's theory which considers exogenous to the text anything that is not directly deductible from the document – including even the author – can be seen as a reflection of the fundamental categories of *Befund* (record) and *Deutung* (interpretation) (Gabler, 2012). The recent introduction of documentary encoding into the TEI has also fuelled this long discussion, as it has been seen as responsive to the needs of scholars for which a separation between record and interpretation is a fundamental editorial principle. Brüning *et al*. (2012) have in fact assumed this theoretical distinction as one of the main use cases for the introduction of the new documentary encoding, lamenting however the way the TEI has then failed to indicate clearly to which of the two levels each elements available for transcription should ideally belong. This also seems to be the interpretation of Robinson (2013), as he swiftly moves from criticising Gabler's documentary theory to disapproving of the new TEI module for documentary editing. And, indeed, if a scholar pursues an editorial work that is theoretically grounded on the clear distinction between record and interpretation, such a scholar may find the distinction between text and document provided by the TEI a useful tool for that purpose; however, the idea that the TEI encourages the provision of objective encoding contradicts the explicit views expressed in the TEI *Guidelines*:

> 'In these Guidelines, no hard and fast distinction is drawn between 'objective' and 'subjective' information or between 'representation' and 'interpretation'. These distinctions, though widely made and often useful in narrow, well-defined contexts, are perhaps best interpreted as distinctions between issues on which there is a scholarly consensus and issues where no such consensus exists. Such consensus has been, and no doubt will be, subject to change. (TEI Consortium 2013, § iv.1)

In fact, the TEI does not state why one should use the documentary approach with respect to the textual approach, and this is in line with the general aims of the TEI *Guidelines*, namely enabling scholars to do what they want. And, indeed, interpretation and subjectivity seem to be inseparable from the concept of encoding, as the discussion in section 5 will demonstrate.

4.3 The Digital Facsimile and the 'Real Thing'

The debate over the opportunity to provide documentary editions rather, say, than critical editions, is intertwined with the debate over objective editing. It is no surprise that both these discourses have become more sensitive as editing becomes digital; the computer is in fact

able to give false confidence to its users of being a 'totally mimetic space unshaped by the constraints of its own medium' (Deegan and Sutherland, 2009, p. 67). Being machines, computers seem to have an 'objective' aura, suggesting that the very fact that we work with computers and present our edition digitally makes such editions more objective and scientific. This perception is also supported by the diffusion of digital facsimiles, which represent in an almost iconic way the aspiration toward mechanical objectivity and the alleged necessity of producing documentary editions that look like the sources they represent. Twycross (2008, p. 23) declares:

> Digitization has ushered in a new age of manuscript studies. We can now view the image of a folio on a computer screen in colour in a detail which would have been unimaginable as little 15 years ago. We can read it; compare it with other texts in other libraries; transcribe it; even, if it is faded or illegible, restore it. We need never see the document itself (though would not recommend it).

The final aside ('though would not recommend it') only partially mitigates the previous declaration that the direct inspection of the document has become redundant, thanks to the provision of the digital image. According to this vision, digital photographs of documents seem able to provide an unmediated access to the real thing, and this is to be preferred to the 'scholarly hearsay' of texts (Nell Smith, 2004, p. 309). Sutherland (2009, pp.17-18) has argued against this vision that presents on the one hand the editorial work as an undesired intermediation, and on the other hand the digital artefact as if it were a substitute for the real object and the digital environment as free from constraints; she defines the faith in the unmediated representation provided by computers as 'screen essentialism' and compares it to the faith of the print editor who is convinced that it is possible to eliminate errors endemic to print without introducing new errors; she argues that 'both ways of thinking need to be recognized for their timeliness rather their essential truth' (Sutherland, 2013, p. 55). Sutherland is not the only one however to be worried about the misconception of digital facsimiles.

A digital facsimile, as any other surrogate, is a representation of some of the facts of a document, but not all of them. For instance, a digital page of a *codex* will represent the shape of the inscribed marks very well, their collocation on the surface, and their visual relationship with the other signs. If the image has been taken with the opportune settings, even colours and some aspects of the materiality of the surface may be reliably represented. However, other features such as the perception of the dimension of the page, its position in the manuscript, the roughness of the surface, its smell, may more difficult if not impossible to capture, not to mention its relationship with the facing page, and the quire. While the real

thing is a multi-sensual object, the facsimile selects only one of the sensorial levels, namely its visual aspect. The act of selecting facts from the infinite set offered by a document has been considered in Chapter 2 as a way to define models derived from the document itself; digital facsimiles with their selectiveness are not exceptions: they are equally definable as models of the documents conjured by the intentionality of an Interpretant in the same way that texts are, but with the difference that they are in turn virtual-physical objects, therefore they also are documents, even if only instrumental ones. Kralemann and Lattmann define images as simplified icons, which stand for the object they model only because of their visual similitude, which is considered the most basic form of modelling (Kralemann and Lattmann, 2013, p. 3406). The main purpose of the model presented by digital facsimiles is to facilitate access to the content of a document, while any other use (codicological, palaeographical and iconographical analysis, for instance) may be more limited: the fact that the digitization process is conducted page-by-page may in fact prevent an iconographic analysis of manuscripts where the decorations are deployed over double pages; digital facsimiles are not suitable for investigating quire structures, pricking and ruling, the state of preservation, or features of the binding.

The fact that digital facsimiles 'look like' the original object, or better, they reproduce the visual appearance of some portions of the real object, has made them particularly suitable for teaching purposes, when the real object is not within easy reach. However convenient, the use of surrogate for both teaching and research has been stigmatized by scholars worried about the limitations of any reproduction and possible misunderstandings that can arise from it. This is not a recent concern, either. From the late Eighties Tanselle has been very vocal against the risks of 'Xerox-land' (Tanselle, 1989b, p. 25), but his remarks echo earlier concerns over the provision of microfilms (Jackson, 1941). Tanselle's criticism is mainly directed toward the use of photocopies in textual scholarship, but some of his points directly transfer to digital facsimiles: 'There is no way that reproductions – regardless of what technology is developed in the future – can ever be the equal of originals as documentary evidence, for there is no way of getting around the fact that they are one step (at least) removed from those originals' (p. 38). The problem, of course, is that one may incorrectly consider reproductions as substitutes for the documents, which in the case of digital surrogates seems quite a high risk, as Nell Smith (2004), for instance, clearly demonstrates. According to Deegan and Sutherland, in fact, computers have given the false conviction that they can provide 'access to the historical versions of texts as real things; and that digital facsimiles are faithful or trustworthy equivalents of originals' (2009, p. 65). In their opinion digital images have an 'exaggerated verisimilitude' able to deceive even expert readers, a risk that is common to many digital resources. Scholars

therefore have to be vigilant in order not to be misled by the virtual materiality and reification enacted by the computers, but must consider them for what they are, namely models (and very useful ones, as it happens) with a specific and limited purpose. Even stronger (with hints of the supernatural) is the argument made by Elaine Treharne, who declares that only in person and in front of the real object one can really understand the document as a whole: '[w]ithout a multi-sensual embodied experience of the material artefact, we experience only the transcendent, the partial; and we only ever grasp a fragmented and limiting understanding of the book's intrinsic aura and of its human ghosts who have wondered through before us' (2013, p. 15). But, really, how high is the risk of misunderstanding the digital surrogate for the real thing? According to Kichuk, such risk is high as she reports the enthusiasm of scholars using the EEBO (Early English Book Online) collection (2007, p. 296). And it is not hard to believe that for some colleagues for whom the accessibility of the real thing is a distant dream, or a very rare event, the availability of digital facsimiles with their verisimilitude (exaggerated or not) may look like little less than a miracle. Yet, we should not exaggerate this risk either. One would expect that scholars are not that likely to mistake a surrogate for the real thing, nor is going in to the reading room of a contemporary library, with modern lightning, magnifying glasses and so on, able to reproduce the exact same reading experience of a contemporary of the document. Demonizing tools because they are used improperly and so produce bad scholarship is not a good idea. Digital facsimiles are models of documents, simplified by definition, able to answer some specific research questions, not all of them, in the same way that an edition is a model of a work. In both cases the intentionality and the interpretation of the editor who creates them represents a very specific point of view that needs to be considered and assessed when such models are handled.

The arguments made by Treharne (2013) with her reference to the 'aura' build, more or less explicitly, on the famous criticism made by Walter Benjamin (1932) regarding technically reproduced works of arts. He maintained that mechanically reproducing a work of art (and he was thinking of photography) will make the artefact loose its 'aura', thereby vilifying it. However fascinating and influential his argument is, we can now say that he was wrong. In spite of seeing endless reproductions of the *Mona Lisa* the queues to see the original at the Louvre are interminable and growing. The same happens for the Book of Kells: in spite of having its decoration on fridge magnets as well as presented in TV programmes and reproduced in many, many books, yet it is one of the most visited cultural objects in Ireland. I say 'in spite of', but perhaps we should say: 'because of'. In fact, the seemingly infinite reproduction of the same object, rather than tearing the aura, seems instead to reinforce it: because we can access cheap reproductions so easily, the 'real thing'

has acquired even more cultural relevance; because it is reproduced so much in every format, the original has acquired even a greater value. It seems that our society, that has been defined as the society of the image and representation, has also almost idolized the value of the real-life experience. People can 'see' their favourite singer or actor in endless videos and photographs in all possible media, but to have a short glimpse of the idol in flesh is worth hours of wait in inclement weather.

To conclude, if the same outcome we notice for the *Mona Lisa* applies to manuscripts (and there are no reasons why this should not be the case, and indeed the Book of Kells suggests that it does) it is probable that the digitization of manuscripts, while being exceptionally helpful for some type of scholarship, will not damage the aura or the understanding of the 'real thing'. However it is essential to remember that some types of scholarship is not all types of scholarship, and that a model is only useful when it is used for the purpose for which it has been built. Images are re-presentations of the real thing, and as Daston and Galison maintained, the prefix re- is essential here: 'representative images may purify, perfect, and smooth to get at being, at "what it is". But they may not create out the whole cloth, crossing over from nature to art' (2010, p. 282). Furthermore digital facsimiles, as with any form of photography, can hardly be considered faithful or objective. As mentioned above, the agent that selects, positions the document, adjusts lightening, filters, angles is a human one, and, as with all humans, is guided by judgement and subjectivity. In the chapter devoted to 'mechanical objectivity' Daston and Galison account for the long journey toward 'true' objectivity of photography, which, no matter how sophisticated and fastidious its set up is, will never be able to achieve the purity of mechanical objectivity.[92] When it comes to human beings, there are no mechanical tasks that can be said to be purely mechanical. But does it matter?

4.4 A Faithful Representation

Digital facsimiles are then not replacements for the real thing, the document, but can they replace documentary editions? In other words, because they mimic so well, even if selectively, the appearance of the sources, much better than any documentary edition will ever do, perhaps then they can be considered a substitution for them. This seems to be the opinion of Hunter (2007), for instance, not to mention Nell Smith (2004). But, as argued elsewhere, the vision of facsimile as substitutions for documentary editions or, historically, documentary editions as surrogates where facsimile edition were too expensive has been

[92] Cf. also the chapter 'Truth and the Digital Image' in Terras, 2008.

proved wrong. In fact documentary editions are more and more disenfranchised from the role of surrogate, demonstrating the scholarly value of their endeavours as tools for analysing specific features of the document and pursuing research goals, in other words presenting models of the text (cf. Chapter 3).

Documentary editions are supposed to represent documents faithfully, a statement which has to be understood to mean 'visually similar' (Tarte, 2011, p. 7). As we have seen above, in documentary editing the only concessions for not representing source documents as precisely as possible are made because of the changes in the medium and technology adopted by the remediation (Tanselle, 1995). In fact, all editorial formats that traditionally embody documentary editions attempt some level of visual resemblance: semi-diplomatic, diplomatic, ultra-diplomatic, type-facsimile editions, all are defined and distinguished from one other by their level of resemblance to the source document. However, it is arguable that looking like the source is the only possible and meaningful embodiment of a documentary edition. There are in fact aspects of documents that can be better studied if resemblance is not considered a priority. For instance, if the purpose of the documentary edition is the analysis of the process of composing and the stratification of the writing, the formats that highlight the features that make these aspects more evident may be preferable. As endorsed by the types of activities that go under the epistemic virtue of trained judgment, a programme of conscious alterations can be undertaken in order to study a particular object; similarly, contrary to Tanselle's argument, such an approach can prove very useful for documentary editions. Editions are models of documents or works, and as such they only include some subset of facts from the many dimensions that the Interpretants recognize as relevant to their purpose of research, among which the visual appearance should only figure if relevant for the scholarly discourse, and not as an unquestioned requirement.[93] An example of this can be seen in the edition of the working manuscript of the libretto of *Tosca* edited by Gabriella Biagi-Ravenni. The manuscript presents at least five different hands: the two librettisti, Luigi Illica and Giuseppe Giacosa, Giacomo Puccini himself, the publisher/producer Giulio Ricordi, a professional scribe plus some cut-outs from some printed proofs. In her edition Biagi-Ravenni associates each hand with a different colour in order to distinguish them; in doing so she deliberately eschews accurate reproduction of the colours of inks in the manuscript in order to present the information more relevant to her purposes, namely who wrote which sequence of text. Another example is given by the Proust Prototype, where the background colour of the zones (editorially established) signifies the incertitude in the reconstruction of the temporal sequence of writing. In both cases, *Tosca* and Proust Prototype, a symbolism that is absent

[93] For a discussion about Dimensions and Facts derived from Documents, cf. Chapter 2.

from the documents themselves has been adopted in order to highlight some features that editors decided were important; both editions still represent the documents, but the points of view adopted by the editors have led them to deliberately alter the visual appearance of some of the features selected from the infinite presented by the document. In fact 'representation is always an exercise in portraiture, albeit not necessarily one in mimesis' (Daston and Galison, 2010, p. 382). As a matter of fact any form of representation is an act of alteration of the object being represented: no matter how accurate the edition will ever be, the gap between the editorial model embodied by the edition and the original document cannot be filled. And it should not be filled: the gap is created on purpose to enable a discrete analysis of the object itself. So perhaps, instead of fighting against the inevitable losses that such representations entail, it may prove more useful to use the inevitable transformations necessitated by the remediation as a cognitive tool. Simplification is a virtue of models, not a weakness, as it allows seeing more clearly aspects that may be lost in the continuity and complexity of reality.

The provision of digital facsimiles alongside (or embedded within) the edition guarantee even more freedom to the editor. Sutherland has noted how the facsimile presented alongside the edition is likely to engage the editor in a sort of competition with the reader, who is then able to scrutinize the minutiae of the editorial work (Sutherland and Pierazzo, 2011, p. 202). But what if the editor consciously rejected the competition? It is in fact simply impossible, no matter how hard one tries, to reproduce the document exactly as it is, so why not decide to consciously and purposefully alter the appearance of the document in order to analyse less evident features? While the provision of the facsimile can fulfil the purpose of providing a visual representation of the document, the edition may concentrate on something less visually obvious and more critically relevant, such as, for instance, the process, the time, the different authorial agencies, and so on.

The search for objectivity in editing is most likely a fruitless quest. Perhaps a more useful approach is one of pursuing accountability instead. Accountability is at the base of any rigorous scholarly method and provides other interested parties with the tools for verifying the work done, testing it and then agreeing or disagreeing with it. Best practices in digital editing encourage one to make available one's archive and one's editorial 'dirty laundry' together with the edited texts (Bodard and Garcés 2009; Cummings 2009; Boot 2009), a practice that has the twofold advantage of improving scholarship and making editorial work fully accountable. Furthermore, aspiring to be objective in the editorial work may simply not be a desirable goal at all: the role of editors is to handle difficult primary sources and traditions because they understand them better than most other readers, and therefore they are

asked to make some sense out of them. And this requires interpretation. One is here tempted to ask: what is wrong with interpretation?

4.5 Markup and Interpretation

Interpretation is also considered a significant problem within the digital editorial community, with the adoption of markup, in particular TEI markup, being attacked as embedding unwanted interpretation into texts from at least the early days of the TEI (Sperberg-McQueen, 1991). In fact, since markup is a way of expressing one's interpretation and understanding of a text, its use for cultural heritage texts has been criticized in the belief that if one embeds markup and interpretation into a text, then this text cannot be reused by other people, affecting therefore the possibility of exchanging and interchanging encoded files. It has then been maintained that one should not encode texts at all or, if really necessary, at least keep this encoding well separated from the text (Eggert, 2005 and 2006; Schmidt, 2010 and 2012). According to this vision, in fact markup becomes an obstacle to interpretation: 'The text is the thing being interpreted, not the markup. Once embedded, markup obscures and biases what a new scholar, who didn't carry out the initial markup, can see'; therefore 'cultural heritage texts should be in a plain text format' (Schmidt, 2010, p. 349). However, a text without some markup is not a text without interpretation, but a text where the interpretation is deceptively hidden by commonly shared writing conventions, or simply silently embedded. The idea that a text without markup is an un-interpreted text is simply wrong. In this realm where there is no meaningful objectivity, as we have demonstrated, texts must necessarily be an interpretation. In fact, even the notion of a text free of markup is misleading, if not false: markup is not only what lies between the angle brackets, but it also includes all the paragraphematic symbols and conventions that we are now used to taking for granted as 'natural' parts of the text but which are the result of millennia of evolution of writing conventions. Capital letters, spacing, punctuation, dashes, brackets, all of these conventions contribute to the legibility of the text, but in most cases they are not born with the text in which we read them, unless we are considering very recent texts. For cultural heritage texts, they are the result of editorial interventions, and they embody the interpretation (the model) of specific editors. Charlotte Rouché traces a concise but nevertheless deeply insightful history of markup, from antiquity to current days, and from *scriptio continua* to XML, discussing how the introduction of markers such as word spacing and paragraphematic signs were progressively (but not linearly) adopted to support different

forms of reading and literacies.[94] She then discusses the introduction of scholarly markup, such as, for instance, the Leiden conventions,[95] which smoothly bring us to text encoding, where markup serves both as an aid to reading (by the human as well as by the computer) and as scholarly convention (Rouché 2012).

If we look at the text of Shakespeare as presented by any modern edition, for instance, the entire paragraphematic system is the product of one or more editors and therefore represents an interpretation of the text. The fact that such a system is part of any (printed) edition of a text and therefore seems 'natural' to a modern reader does not affect its interpretative nature, a fact that is demonstrated, for instance, by the fact that no two editions present the same punctuation. In several contributions Dino Buzzetti has drawn attention to the ambiguous nature of markup (by which he intends the XML type of markup) which can be at once part of the text (when, for instance, it substitutes a feature of the texts such as a punctuation mark) and at the same time expresses an argument about the text (when, for instance it declares a string to be the name of a person), concluding that markup is, if not dangerous, at least problematic (Buzzetti, 2002 and 2009; Buzzetti and McGann, 2006). Nevertheless, the ambiguity is the one of language itself and is inevitable as our writing system has incorporated functions of markup which are historical and critical constructions. Claiming that a comma is an inherent part of the text may therefore be arguable from an historical point of view: if the text has been written in Antiquity, the Middle Ages, the Early Modern or Modern periods, for instance, it is very likely that that comma was not originally part of the text at all and has been deliberately inserted in the text by some editor at some point as a way to make the text more readable or clear, according to her/his interpretations of the passage.

A second and perhaps more dangerous implication in the assumption that a text without (XML-type) markup is a text without interpretation is the idea that a text objectively can and does exist outside the negotiation and cultural dialectic of editorial mediation. However, a text, any text, is the result of an act of selection operated by an Interpretant; it is a model, and therefore it is always the result of a culturally complex compromise and not an objective entity that can exist independently from interpretation. To consider an extreme case, in what we conventionally call the *Odyssey* every single word, its location, its

---

[94] On the 'invention' of punctuation the most important contribution is of course Parkes (1992), often cited by Rouché (2012). Schmidt thinks instead that markup and paragraphematic system are two distinct entities and should not be conflated (2010, pp. 338).

[95] On the Leiden conventions and their significance for editorial practice in Classics, cf., for example, Dow (1969) and Panciera (1991). The 'conversion' of the Leiden convention into XML-TEI is at the base of the creation of the EpiDoc encoding format (Elliot et al. 2011-2013).

belonging to a structure such as the verse, for instance, is the result of centuries of editorial conventions and compromises, as its most ancient witnesses are a few hundred years more recent than the supposed creation date, and so only fascinating but mainly unsupported hypotheses can be used to postulate what happened in that temporal gap. As argued above, a text without markup is not a text without interpretation, but rather a text where the interpretation is deceptively hidden by commonly shared writing conventions. But when it comes to texts, we should oppose the ontological category of objectivity with the scholarly category of accountability, so if we want to be 'scholarly' in our endeavours then we need to provide our texts with explicit statements about interpretation. And again, markup can do exactly this, namely making explicit and accountable editorial decisions.

And again the question arises: what is wrong with interpretation? The principal activity of scholars in the humanities is to interpret historical artefacts, however when it comes to representation of texts this seems to become a problematic activity. But, as Sperberg-McQueen maintains, 'if our understanding of texts is worth sharing in critical essays, it is worth sharing in our markup of electronic texts' (1991, p. 36). Editors are the people that know the most about the text being edited: they may be wrong, but their opinion is worth having, in particular if such an opinion is appropriately supported by evidence and accurately documented. The TEI, with all its limits, has been able to provide editors with the tools to unmask the deceptions of clean texts, exposing the 'wires' of the cultural compromise that is editing, documenting their decisions and, why not? their trained judgement. Since the TEI has become an international standard masterfully documented in itself, it also allows for the interpretation to be transparently declared and is therefore controllable and accountable, as is required by scholarly endeavours. And if someone would like to disagree, nothing prevents the provision of another (encoded) interpretation, in the same way that scholars disagree with their colleagues and respond to them by the provision of other essays and editions. It seems ironic that while the goal of 'definitiveness' is being abandoned in textual scholarship (Shillingsburg, 2006, p. 154), this is still considered a requirement for digital texts.

Having argued against the assumption that it is possible and desirable to edit without alteration, and that a documentary edition has to look like the document it edits, it seems now time for a new definition of documentary editions, their goals and their rationale. We could then define a documentary edition as the edition of a primary source with the aim of analyzing and representing it for a particular purpose of research. Whether such an edition is effective, scholarly and in general a worthy endeavour will be judged by if and how the

editor has fulfilled the purpose of such research. This will allow also for an honest transmission of the text preserved in the document through time: if we make the inevitable editorial interpretation accountable, it will be possible for the reader to assess the value and trustworthiness of the edition.

5. Work and Workflow of Digital Scholarly Editions

The previous chapters presented and discussed theoretical aspects of editing, the implications of some aspects of which are specifically connected with the conception and preparation of scholarly editions in a digital framework. The following chapters will turn steadily toward digital editing, including technical issues and choices. This will not, however, becoming a 'how-to' handbook for digital editing: as with the previous chapters the focus is still on the methodological questions generated by the digital medium.

The first set of questions to be examined here focus on the work and workflow deployed in the production of digital scholarly editions. How are such editions prepared? Is the work necessary to its production substantially different from that required by printed editions? Certainly in both cases the editors are required to obtain access to some sort of primary sources, perhaps securing reproductions from libraries, then the text of the document needs to be transcribed, and if there is more than one extant witness, then these need to be collated and their differences accounted for and possibly organized in a *stemma codicum*. Then the edited text needs to be crafted, annotated and polished for publication. All these operations and phases remain unchanged with the shift to the new medium. However the way they are undertaken as well as the emphasis on some steps over others are considerably different and imply a radical re-thinking of the editorial work as well as the role of the editor.

The use of computers in the preparation of a scholarly edition can assume very different roles, and what editors expect to gain from using them varies enormously. To some extent, all editions are produced with the help of computers nowadays, in the sense that they will all at the very least be typed within a word processor, the transcription will probably be done from a digital photograph and the edition, once completed, will be sent to the publisher electronically. However, extending Patrick Sahle's useful distinction, this type of work more than a digital workflow could perhaps be called a 'digitized' workflow (Sahle, 2008, Online), where computers are tools and have no (or are not recognized to have) hermeneutical value. At the other extreme, there are editions that not only are produced with the help of computers, but they also adopt computational processes and methodologies in order to produce highly interactive digital editions; if we stick with our distinction, we could call the latter a 'digital' workflow. O'Donnell reflects along similar lines when he claims the right of editions with no digital outputs also to be called digital if they adopt a digital methodology in the editorial work, arguing that digital should not be reserved to the type of output alone, but that the whole workflow should be considered (O'Donnell, 2008).

In 1996 Peter Shillingsburg described a workflow based on computers and called it 'computer assisted scholarly editing'. In his vision, computers can be very usefully employed in most of the phases of the editorial work and in particular in alleviating the so-called 'idiot work' constituted mostly by collating, typesetting and proofreading (Shillingsburg, 1996, p. 134 and 139). However, while Shillingsburg is certainly more digitally literate than the editors who use a word processor and emails as the only digital props in their workflow, 'computer assisted scholarly editing' as he presents it is not very dissimilar from this in the sense that computers are viewed as a convenient tool able to alleviate the editorial work, fitting in the normal workflow without questioning or substantially changing it.

A few years later Gregory Crane *et al*. prefer talking about 'ePhilology', describing the digital editor as someone who moves in between the 'algorithm-heavy, knowledge-light approaches of computer science and the wholly manual practices of the traditional editing', finally stating that 'all philological inquiry […] is now a special case of corpus linguistics' (Crane *et al*., 2008). The distance between ePhilology and computer assisted scholarly editing could not be further. Crane *et al*. present the new workflow as requiring an editor with augmented computational skills and the editorial work as a computational-heavy algorithmic process, with both requirements thought to be necessary in order to fully unlock the potentials of digital editing. The framework described by Crane *et al*. is really exciting to some, embedding as it does the promise for new and advanced research. However, this change of habits and acquisition of skills is not easy or painless for the editor.

5.1 Modelling the Edition

Tim McLoughlin describes well the disorienting sensation of being in a totally new territory when he decided to approach a different type of workflow, a digital one. According to him, in moving from print to digital the most important thing that happens is that the purpose of the edition is transformed, with the final product changing from a text to be read to an 'object' to be used (McLoughlin, 2010). This change forces the editor to undertake unexpected tasks, most of which are likely to be unwelcome, or, as Bree and McLaverty put it, non relevant for the editor, as they maintain that, while working at a digital edition, 'the scholar is also likely to have to acquire skills which may well be useful and interesting in themselves but which stray a long way from what could be regarded as making a direct contribution to scholarly research' (Bree and McLaverty, 2009, p. 127). At the same time, the new editorial model is full of surprises, such as the flexibility and openness of the encoding schemas, and the need to guess what users are likely to want *to do* with the 'material', as opposed to 'read texts'.

These are all changes that lead McLoughlin to describe this as '[n]ew territory indeed' (2010, p. 46). In his response to McLoughlin's concerns, Rehbein outlines the steps that are necessary in order to achieve what he calls 'the transition from classical to digital thinking'. In his account, Rehbein describes how this transition necessarily should go from thinking how the text will look on the printed page to a functional and semantic model of the text. Digital thinking asks the editor to think in a more abstract way, and to move 'from output-driven to input- and user-driven design' (Rehbein, 2010, p. 62). It also calls for the application of a functional model which he has called the 'why-what-how' approach, according to which, before you can decide *how* you do something, you need to be sure you know *what* you are going to do, but this could be meaningless unless you have clearly established *why* you are going to do it. In other words, Rehbein is saying that before starting to edit a text, you have to model it, and by modelling we intend here the analytical process of establishing the kind and purpose of the edition, its implied community of users and what features best represent their various needs (McCarty, 2005, pp. 20-72), applying the type of theoretical abstraction which lies at the heart of chapters 2 and 3. Although this is not a new activity – editors regularly determined the type and purpose of their editions long before the advent of computers – nevertheless modelling has acquired a new centrality in digital editing, the reasons for which have been covered in the previous chapters. In addition, it is worth noting how in traditional editing many decisions in term of what to offer to the readers and in which form are in fact not the choice of the editors preparing the scholarly edition, but are determined by the publishing house, with its stylesheet and publication format. In this respect the model of the text in most cases is not determined by the editor of a traditional print volume, but instead by the publishers (or the series editors). What is also different between modelling for print and modelling for digital is the way such a model is expressed: a declaration of intents and editorial practices in traditional editorial work, some kind of computer-friendly formalism in the case of digital editing.

A method that has proved to be successful for modelling digital scholarly editions is to start from the end, namely designing the final product – the edition, the *what* according to Rehbein taxonomy – and then working backwards, trying to understand which methods and tools will be required in order to achieve the desired output (Sutherland and Pierazzo, 2011, p. 195). This approach is thought to work for several reasons, firstly because it allows scholars not used to working in digital environments to concentrate on the things they would like to do with the edition, rather than trying to think how to produce the edition, an understanding that may only come to them at the end of the editorial work. Secondly, it allows scholars to drift away from the paradigm of the page: if they are offered the possibility of designing a digital product right at the beginning of the editorial process, it will be easier

to think in digital terms, and not to transpose on the new medium their expectations nurtured within a print environment. Finally, this approach helps scholars to concentrate on the research questions that the edition will have to answer, or the *why*. From experience, perhaps the best way to approach the editorial work is to use visual aids, designing a so-called 'wireframe', which is a skeletal drawing of a fully functional final product (the website) (fig. 13).

<table>
<tr><td colspan="2">Banner</td></tr>
<tr><td colspan="2">Home | About | Search | Browse | Contact</td></tr>
<tr><td>Contextual menu</td><td>Search:

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Typi non habent claritatem insitam; est usus legentis in iis qui facit eorum claritatem. Investigationes demonstraverunt lectores legere me lius quod ii legunt saepius. Claritas est etiam processus dynamicus, qui sequitur mutationem consuetudium lectorum. Mirum est notare quam littera gothica, quam nunc putamus parum claram, anteposuerit litterarum formas humanitatis per seacula quarta decima et quinta decima. Eodem modo typi, qui nunc nobis videntur parum clari, fiant sollemnes in futurum.</td></tr>
<tr><td colspan="2">Footer information</td></tr>
</table>

FIGURE5.1: A SIMPLE WIREFRAME

In spite of the traditional mistrust of Humanities scholars towards visual representations (Jessop 2008a, p. 47) the graphic support of a wireframe has proved to be a fundamental thinking aid across many digital humanities projects developed at the Department of Digital Humanities (King's College London), a way to embody visions and imagine the future, not to mention the advantage from a computing point of view of having ones mind clearly made up before the development of the real website. Indeed Martyn Jessop is right when he declares that the 'greatest value' of digital visualisation is for 'synthesis and modelling' (2008b, p. 288). According to this top-down approach, a functional analysis of digital editions allows the determination of the type of editorial model that will be necessary to deploy in order to achieve the desired output; in other words, determining what a digital edition should do will determine the theoretical model that should lie at its base.

According to Buzzetti, a digital edition is both an 'image', a representation of the text, and also 'textual data to be processed' (2009, p. 47). In fact, in order to be processable, data must be modelled in a way that is suited to the purpose for which it has been created. In other words, the model of an edition must foresee all the possible questions editors and end users are likely to pose, and must embed the knowledge necessary to answer such questions. In a work of 2011 I have listed a series of features that an editor may want to include in her/his transcription in the creation of a documentary edition (the *what* of Rehbein), and a list of parameters that that will help decide their inclusion or not in the final edition (the *why* of Rehbein). Although the list of features has been modelled with a digital documentary edition in mind, the parameters apply to any type of digital editions, and are:

1. **The purpose of the edition**: which are the scholarly reasons for which an edition of a particular work or text is necessary; which gaps in the previous editions, if any, is the editor to fill? Which is the theoretical framework that guides the editorial work?
2. **The needs of the users**: to which type of user, present and future, is the edition aimed, and what is the particular subset of user likely to need in order to achieve their aims?
3. **The nature of the documents**: are the documents to be edited clean, scribal manuscripts? Early printed material? Draft manuscripts? Which are the extension and the relevance of the non-verbal content? Is there more than one witness or many?
4. **The capabilities of the publishing technology**: which is the best technology, if any, able to fulfill the requirement of the edition? Which is its life expectancy? Is it sustainable?

5. **Costs and time**: to what degree are the desired outcomes actually achievable within the timeframe allocated by the funding available? Which features will require more resources than are actually worth in terms of desired output? What can be postponed to a successive phase? Which are the priorities?

The article (Pierazzo, 2011) started from the acknowledgment that in the print world the limits of the model were established by the limits of representations offered by print. This position, implicit in many editions, is made explicit by Vander Meulen and Tanselle: 'by ''transcription'' we mean the effort to report—insofar as typography allows—precisely what the textual inscription of a manuscript consists of' (Vander Meulen and Tanselle, 1999, p. 201). A similar consideration can be made for critical editions based on multiple witnesses, with the format, the inclusiveness, and the location of the *apparatus criticus* depending on the constraints of the publishing house. However, the improved representations and storage capabilities offered by digital publishing systems have on the one hand lifted such limits, but on the other hand made necessary the introduction of other criteria for choosing which features to include in editions, and this in turn should now be based on scholarly considerations rather on constraints of the medium. For that reason, of the aforementioned parameters, the purpose of the edition has been recognized as the most relevant criterion for the production of a model for an edition. This is perhaps one of the most important transformative opportunities offered by the adoption of a digital framework for a scholarly edition, that is the possibility for scholars to establish criteria and models based more on scholarly concerns than on other pragmatic parameters such as, for instance, the limits of the publishing system. Editorial practice has been so largely informed by print technology that in most cases editors have accepted publishing conventions as if they were scholarly rules. Two examples of this come to mind: the first is taken again from the diptych of articles written by McLoughlin and Rehbein in *Computerphilologie* (McLoughlin 2010; Rehbein 2010). In his contribution McLoughlin laments the sense of disorientation given by the lack of an 'agreed style-sheet' in digital editing and by the general openness of the TEI *Guidelines*, which allow one to choose the appropriate level of encoding in one's edition and to create new elements when the required ones where not available (pp. 45-6). In his side of the debate, Rehbein outlines the discussions which led to the modelling of the digital edition of the correspondence of James Barry, edited by Tim McLoughlin. This process can be summarized emblematically in the encoding of an uncertain date. Here the 'common editorial practice' of putting the uncertain date in square brackets (according to an 'agreed style-sheet') is considered by Rehbein an example of 'output-driven editing', and is compared with the text encoding alternative ('input- and user-driven editing') of encoding the reasons for such uncertainty, and the weight of the uncertainty itself. Rehbein argues for the superiority of the

latter because it describes the editorial principles rather than the output (Rehbein, 2010, pp. 60-61). Despite Rehbein's convincing argument, McLoughlin answers by lamenting the verbosity of the TEI encoding and the increased workload it brings, noting that his 'experience has been that an editor might need an additional 30% or more time for the project' (McLoughlin, 2010, p. 46).

The second example is offered by the *critical apparatus*, a requirement for any critical edition based on multiple witnesses, and yet its legendary density and unfriendliness makes it the emblem of what does not work with printed editions.[96] The point here is that, while the information collected in the apparatus needs to be there, the apparatus itself does not have to be so cryptic or 'arcane' (Shillingsburg, 1996, p. 119), and once the space constraints dictated by print publication are lifted, the apparatus can actually become the place where the scholarship of the editor is proudly highlighted. Sperberg-McQueen observes, in fact, that a multi-purpose edition may be able to display the apparatus to different users in different ways, or even not present it at all if desirable. He also argues that this will help put the apparatus back at the centre of the scholarly discourse, as the practice of relegating it to the bottom of the book has made it simply unreadable without buying two copies (Sperberg-McQueen, 2009, p. 33). Furthermore, the format of traditional critical apparatuses, in spite being well established in the scholarly community, are not alien to ambiguities or inconsistencies, as spotted by Andrews and Macé: 'of the several critical apparatuses analyzed during the course of our research, each of them contained some error or ambiguity that required the editor to be consulted for clarification' (Andrews and Macé, 2013, p. 506). Clearly the critical apparatus is an area where improvement is indeed possible and necessary.

5.2 Computing the Edition: the Two Paths and the Third Way

The role of computers in textual editing take two different approaches:

1. Computers have to take care of the difficult, boring tasks that litter the editorial path to the production of an edition. Editors, then, seek tools that will assist their work; the tools themselves and their heuristics are not questioned, as long as they do what they are 'told'.
2. Editors interact deeply with computational methods, shaping their work using markup language, using advanced analytical methods and building the tools they need to

[96] For a more detailed discussion of the critical apparatus and what is wrong with it, cf. Chapter 6.

process their texts. The tools and their heuristics are considered an integral part of the scholarly endeavour and so they should be published and open to scholarly examination along with the editions.

The first attitude is strictly connected with the 'computer as electronic research assistant' attitude mentioned in chapter one: computers can be assigned mechanical work under the direct control of the editor. They will do the tasks they are assigned without mistakes (as long as the instructions are clear), leaving the editor in control of the work and in control of the computer. At the same time, the editor will not interfere with the details of the methods used by the computer: the computer will simply do what it is told. We could perhaps call this approach 'computer assisted philology (or textual scholarship)'.[97] The editorial work remains conceptually the same, and the computer is used in various ways to speed up or make such work more precise.

But what is the editorial work and where can computers be safely employed? In a contribution of 2000 Parker asks what computers are for in editing and answers with a list of tasks for which the use of computers can be of help. In his list, Parker includes:

1. Collating witnesses;
2. Allowing for the alteration of a base text without having to revise a complicated *apparatus criticus*;
3. Analyzing manuscript relationships;
4. Selecting the most significant witnesses;
5. Producing a publishable edition;
6. Enabling collaboration;
7. Allowing the reuse of data for different purposes;
8. Making possible a wide range of presentations.

This task-oriented list does not allow for the discovery of unexpected results, unless accidentally.[98] Computers are seen as way to help in 'build[ing] on the achievements of the past while attempting to avoid its failures' (p. 33).

---

[97] The term 'philology' (as well as its translations in various mainly European languages) means different things in different countries and in different disciplinary contexts. It is used here, following the Italian tradition, interchangeably with 'textual scholarship'.

[98] Incidentally and rather oddly, one may notice that transcription is not part of this list, which is an area that has received a great deal of interest in the field, with a large number of contributions, as seen in Chapter 3.

In a similar intervention of 2006, Andrea Bozzi describes a 'workstation' able to support the editorial work in all its stages. He imagines a user-friendly software suite 'since the users are not necessarily experts in computer science tools'. This workstation should be able to support the editor not only in the preparation of the edition, but also in core activities of textual criticism. He then defines the computer-assisted editorial work and textual criticism as 'computational philology'. His specifications for the workstations include:

1. Support for transcription by handling digital images
2. Support in the creation of the critical apparatus
3. Handling bibliography
4. Connecting and organizing witnesses
5. Creating indexes and concordances
6. Mapping transcription and images

In his view, the workstation, once populated with the data provided by the editor, could be made available to the public, becoming a digital edition.

While some of the proposals advanced by Parker and Bozzi are indeed tempting, the problem with 'computer-assisted philology' is threefold. Firstly the lack of engagement with the technology is risky and limiting at the same time. It is risky because it lacks the deep understanding of the ways data is handled, the actual process that is undertaken and its implicit assumptions, with potential misunderstanding of both the meaning and the relevance of the results. In this respect the enlightening contribution of Sculley and Pasanek of 2008 marks a landmark for the awareness and caution with which we ought to handle computational methods. In their article, they expose the implicit assumptions at the core of several data-mining algorithms, reaching the conclusion that failing to understand and engage with such assumptions, biases and limitations will invalidate any result produced computationally. In fact, interdisciplinary work and the reliance on the tools that someone else has developed and of which we lack the necessary understanding comes at a risk as, 'coming from disparate backgrounds, we tend to believe that our own assumptions are shared assumptions' (p. 409). An approach then that relies on sophisticated pieces of software such the one required by editors that want to take advantage of the technology without engaging with its principle can be hazardous as well as limiting. Once again Sculley and Pasanek, having discussed at length the risks behind the blind trust of methods which scholars do not control, then state how

> [T]he virtue of automated analysis is not the ready delivery of objective truth, but instead the more profound virtue of bringing us up short, of disturbing us in our

> preconceptions and our basic assumptions so that we can exist, if only for a moment, in uncertainties, mysteries and doubts (p. 423).

But it is not just the lack of understanding of the processing implication that makes this approach difficult: it is also the lack of appreciation of the transformative hermeneutics of the deployment of the computer. Willard McCarty expresses this well in a contribution of 1991:

> In the early stages of a new technology, people tend to think that its purpose is merely to replace and improve on something they already know. The promise of the new is thought to be quantitative: the new things will do the old job faster, more efficiently, and more cheaply. Tools, however, are perceptual agents. A new tool is not just a bigger lever and a more secure fulcrum, rather a new way of conceptualising the world.

It is the new conceptualization of the world produced by the use of computational methodologies (the 'tools') that make digital editing something 'else' with respect to traditional editing.

The third reason why computer-assisted philology is problematic is sustainability. In order to build tools that textual scholars may use for their editorial work, developers need to be able to foresee the tasks that scholars wish to perform in other words; developers have to model the editors' work. However, this is easier said than done. While it is (relatively) simple to model the way one scholar works, or even a small group, it is much more complicated to model the working methods of a large community of textual scholars, due to the different theoretical approaches to their workflow, the different types of editorial product they aim to produce, and their national and disciplinary habits and idiosyncrasies. To make scholars across disciplines agree on a set of standards seems far from achievable, as bitterly admitted by Tara Andrews when she declares how 'flexibility and customizability is currently much more important to textual scholars than the sort of standardization that would allow for true progress toward digital critical editions' (Andrews, 2013, p. 63). The only exception to this is given by the TEI which is indeed flexible and customizable, a fact that is seen as a problem by developers of tools and as a blessing by (some) scholars because it accommodates many diverse editorial habits.[99] It is then possible to develop tool suites for the benefit of a scholar or of a project, but these are rarely usable outside the environment that has developed them, making them unsustainable.

---

[99] The flexibility of the TEI is not a blessing for everybody: the criticism of Tim McLoughlin discussed above is a clear example of that.

Andrews, (2013) provides a list of areas of editorial work where the use of a computer can be fruitfully employed. She revised the list provided by Peter Robinson in 2004 who in turn was trying to convince 'digital skeptics' that the application of digital methods still imply a large use of manual work.[100] Andrews' list includes:

1. Transcription
2. Analysis, including collation, stemmatic and phylogenetic analysis, stylistic analysis, data mining.
3. Editions, that is the establishment of the text and the building of the critical apparatus
4. Publication.

While on the surface this list seems similar to the ones seen before, Andrews' list presents some substantial differences from Parker's and Bozzi's as she engages directly with the computational methodology, stressing that the 'deeper value of digital philology […] is that it should allow not only for innovative means of publication and display, but also innovative working methods and unexpected results' (p. 66). This attitude locates her in the second approach mentioned above, namely editors that allow the transformative act of computational methodology to question their working habits and assumptions. We could then call this second approach, following Andrews, 'digital philology', emphasizing with that term how the digital medium shapes and informs the philological work. According to this approach, the editor should be able to control the data-entry and the software in a deeper way than as implied in computer-assisted philology seen above. Editors should acquire encoding and programming skills, in order to prepare, analyze and publish their data in the exact way they have intended. This is exactly what Andrews reports she has done as she outlines the steps that have lead to the digital publication of the *Excerpts of the Chronicles of Matthew of Odessa* (Andrews, 2012). However fascinating, this approach has been frowned upon as too technical and outside the realm of editorial work. For many, the use of computational techniques implies a division of labour with the editor doing the 'intellectual' part of the work and some 'assistant' doing the practical work of, for instance, text encoding.[101] However, supporters of text encoding have long argued that the evolution of the editor into an encoder is a natural one, as encoding is not the mere mechanical application of tags, but a highly interpretational and intellectual activity (Sperberg-McQueen, 1991). It is interesting to

[100] Incidentally, it is worth notice how Robinson (2004) calls 'science' the use of computational methods, and 'art' the manual intervention of the textual scholar, a terminology which, even if it relates to the contribution of Housman of 1921, can give the wrong impression of a false reliance on the scientificity of computing, as lamented by Sculley and Pasanek (2008).

[101] See for instance, Eggert (2009b), p. 66 and Bree and McLaverty (2009), p. 127.

note that these claims about the intellectual and critical nature of text encoding echo very closely the ones made by traditional textual scholars when accused of being only 'practitioners of the text' by literary critics; emblematic in this sense are the declarations of Greetham (1994):

> [I]t must be emphasized that editing and textual scholarship are not simply technical skills, which once learned can be easily transferred from one field or period to another, without the editor's having developed any historical training in the new area. Editing depends upon textual scholarship, but textual scholarship is not merely method or technique; it is judgment and criticism, evaluation and discrimination, encompassing historical and cultural learning as well. (p. 5)

Editing is technique as well as understanding, and it always has been. The acquisition of computational skills should then not been seen as outside the editorial practice but rather a simple evolution and not a revolution. This is the reason why, for instance, the largest number of TEI practitioners are indeed people who deal with manuscripts, and most likely are editors (Burghart and Rehbein, 2012, § 10).

But what are, exactly, the skills required for this approach, and is it realistic to expect that all editors will acquire them? The first requirement is to prepare the text for processing, which means, in many workflows, to transcribe and encode it according to a computational formalism.[102] This first step requires transforming the editor into an encoder; more specifically, in most cases, into an XML-TEI encoder. Learning this is not a trivial matter, as demonstrated by Tim McLoughlin experience's (McLoughlin 2010), who, however, complains more about the different logic of text encoding versus traditional practices, than about the technicalities of XML-TEI itself. Similar results have been uncovered by Burghart and Rehbein (2012) by means of a survey conducted among encoders of manuscripts. According to their survey, there is a noticeable gap between the moment someone becomes aware of the TEI and the moment she/he start using it (§ 28, 32-35). The same survey shows also how the difficulties reported by McLoughlin are not an isolated case, with the TEI's many choices and options being one of the biggest obstacles in the learning process (§ 34, § 58, § 67), which is like saying that the complicated aspect of the TEI is to use it as a modelling tool. Shillingsburg is even more radical when he complains about the 'irritating distraction' of XML and TEI (2006, p. 115). Again, Rehbein and Fritze (2012) report that the most complicated aspect for the participants in their courses on digital editing was the

[102] For modern texts transmitted by print it may be more convenient to OCR the source and then correct it manually.

theoretical aspect of encoding, including modelling, while the practical was much more clear (2012, pp. 72-73). This is in fact a finding that might come as a surprise to some: one might have expected that editors would find the actual technical encoding more difficult than the theoretical aspects of it, as scholars in the humanities are used to handling abstract notions. These results confirm then two assumptions made before, namely that textual scholarship is also a technical activity and therefore the application of text encoding comes as a natural evolution of the editorial work; and that modelling in order to enable computational processing represents a novelty in the humanities, able to challenge the way scholars think about their work.

But encoding is just the beginning of the journey: the reference model provided by Rehbein and Fritze (2012, p. 52) shows a complex, yet pedagogically simplified workflow of digital editing.

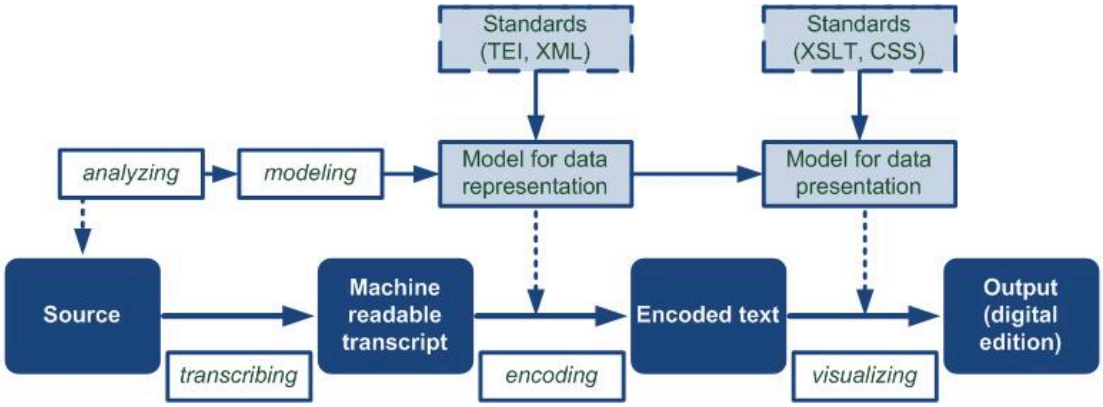


FIGURE 45.2: WORKFLOW OF EDITING (REHBEIN AND FRITZE, 2012, P. 52)

Once the text is encoded, it will then need to be processed somehow, and therefore the editor will be required to have a certain level of familiarity with one or more programming languages; if the encoding has been done in XML-TEI, then probably the programming language will be XSLT, which is able to take an XML file as an input and produce many types of outputs. Most digital editions will present some if not all their results on the web, therefore HTML, and CSS and probably JavaScript should be added to the list of necessary skills, not to mention principles of Web design, in order to produce usable and pleasant websites. Most editions will not only offer the edited texts to their readers, but also some other material and elaborated data such as indexes of people, places, subjects and so on. For this task, structured data formats such as relational databases or ontologies may also be appropriate. And the list could be enlarged to include skills as system administration, metadata formats, querying languages and so on. All of this is on top of the 'core'

competencies that one is to expect from a scholarly editor such as textual scholarship, codicology, palaeography, historical linguistics, literary criticism, and so on. Is this a realistic proposition? It is hard to see how it could be. Burdick *et al*. (2012, pp. 132-3) provide an even longer list of 'core competencies in processes and methods' which are required, in various degree and combinations, in the creation of most digital resources, and significantly group them in three subsets:

- Technical: Familiarity with datatypes and file formats; Database knowledge; XML structured data; Metadata standards; Scripting languages; GIS platforms and spatial data; Virtual simulation tools; Existing and emerging platforms for content management and authoring; Interface design as knowledge modelling; Game engines; Design for mobility and diversity; Custom-built vs. off-the-shelf.
- Intellectual: Cross-cultural communication; Generative imagination; Iterative and lateral thinking.
- Administrative: Intellectual property; Institutional circumstances; Sustainability, funding, and preservation.

While there are scholars who have achieved such an impressive skillset, it also seems evident that they are setting the threshold very high and that it is not likely that this profile will become very common in the foreseeable future, if at all. To be fair, Burdick *et al*. did not envisage it to be the skillset of an individual, yet if the editor has to accomplish the creation of a high quality digital edition, that list is not far from the actual skills required by the task. For sure the list could be cut down to a more sizable dimension if we locate our ideal editor within a centre for Digital Humanities, or if we allow her or him to have enough money (a grant, most likely) to be able to afford to hire someone (or, more likely, a team) with the required technical skillset. However, this possibility is only likely in some privileged cases, and is not available to the large majority of scholars who might want to prepare and publish their editions digitally. In the course provided by Rehbein and Fritze the infrastructural and transformational part was achieved via a pre-existing tool called 'Scalable Architecture for Digital Editions' or SADE (Czmiel, 2008)[103], but even with an infrastructure already in place there were a few problems of setup in order to connect the textual editor to the database. This may sound trivial, but solving it requires relatively high level of experience in hacking tools and web services (Rehbein and Fritze 2012, pp. 64-66, p. 71), an expertise which is often beyond the grasp of most editors.

[103] SADE is now available from a different website with respect to the one included in Czmiel (2008): <http://www.bbaw.de/telota/software/sade/sade-1>.

Editing is often a solitary work which is performed by scholars on their own or with limited funding. Conversely, digital editing is typically either a very expensive activity or it requires such an impressive amount of technical literacy to make it unrealistic. Peter Shillingsburg remarks how '[c]reating an electronic edition is not a one-person operation; it requires skills rarely if ever found in any one person' (Shillingsburg, 2005, p. 94). In his paragraph meaningfully entitled 'It takes a village', Shillingsburg examines how the adoption of a digital framework may lead to the conviction that 'it is the textual critic's duty, in the electronic age, to become an expert in electronic matters, perhaps for the same reason some editors became type compositors – they do what they have to so in the absence of the support that would provide them with the necessary team' (ibid.); nevertheless, he maintains that the workload is so daunting that a team is required anyway.

Both approaches, computer-assisted philology and digital philology, have limits and advantages. While computer-assisted philology enables editors with a limited computational literacy and working on their own to take advantage of the rich opportunities offered by computers, it is also very difficult in practice as such a tool-suite does not yet exist, due to the difficulties of modelling for so many different behaviours that is required by the disciplinary and contextual differences of the editorial work. Digital philology, on the other hand, while giving to the editor – and enforcing – the full control over the digital workflow, also requires an impressive technical skillset which, again, either makes it scarcely available outside institutional frameworks such as specialized centres of Digital Humanities, or requires considerable funding to build up a heterogeneous team, with the challenges that come with it. Is there any other possibility? Is there a Third Way? Andrews uses this expression to refer to a different and truly digital textual scholarship, able to challenge all the assumptions and the heuristics of our discipline, embracing the digital paradigm, freed from the page constraints. While this call seems more than worth answering,[104] the system that she describes is not the third way that is sought here, as her approach seems more a radicalization of what we have defined as the second approach (digital philology) rather than a new approach in term of work and workflow.

The third way could perhaps be found instead in what I have defined as the 'bricks approach', which attempts to build small tools rather than big workstations, in such a way that with a small amount of configuration they could be combined and customized to serve a

[104] Cf. Chapter 9.

specific editorial work.[105] The idea behind this approach is rather simple: while it is impossible to develop a single framework that will satisfy all the possible use cases that characterize the variegated panorama of textual scholarship, it seems relatively easier to highlight 'microtasks' that are undertaken in most editorial enterprises and build tools that support them; it will then be up to the user to combine and customize the tools in a fashion able to fit the specific workflow. Van Zundert and Boot claim that this approach is a defining factor of the 'Digital Edition 2.0', and that it should characterize not only the workflow but also the publication stage: 'Editions should make use of services (e.g. a collation or indexing engine) that are already published on the web, for reasons of both technological and scientific efficiency and reliability. This implies that software/services will be highly distributed' (Van Zundert and Boot, 2012, p.144). The authors propose a further step in the third way, namely the cloud service approach: services such as collation, text analysis, and image manipulation will be web-based and reside somewhere in the cloud, and the edition will simply access them 'on the fly' and on demand. It goes without saying that this approach requires that the data format should be compliant with some agreed sets of standards; it also presents other practical and theoretical implications which will be discussed in chapter 7.

This third way tries then to combine the advantages offered by the two approaches seen above, namely a reasonable requirement in term of technical literacy along with sustainable technical development. However, at the present time the tools available that follow this approach still require a high level of computational skill to be customized,[106] and they are therefore *de facto* not available to people without access to IT support or not in possession of advanced digital and hacking skills, but this, hopefully may only be the result of the fact that this is a relatively new approach that requires more time to mature .

5.3 The Role of Standards in Textual Editing: the Text Encoding Initiative

In the previous chapters as well earlier in this one, the data format developed and distributed by the Text Encoding Initiative has been mentioned a number of times as one of the most

[105] This is the approach that has been pursued by Interedition, an EU-funded project that has aimed at developing a shared methodological basis for textual scholarship and that favours the interchange of tools. The list of tools developed by the project is available at <https://github.com/interedition>.

[106] The experience of Rehbein and Fritze mentioned above presents this exact problem: a tool easy to use in theory, but which glitches resulted in unforeseen complications (Rehbein and Fritze, 2012, p. 71).

important for digital textual editing. It is time now to analyze the real impact and significance of the use of the TEI and other international standards in scholarly editing.

The Text Encoding Initiative is a consortium devoted to the development and maintenance of the TEI *Guidelines*, which offer a set of recommendations for textual markup (Cummings, 2008, p. 451). At the time of writing, the TEI *Guidelines* are expressed in XML (eXtensible Markup Language), a standard developed and maintained by the World Wide Web Consortium (W3C). The inception of the TEI is commonly associated with the so-called Poughkeepsie Principles of 1987, a short document of nine points that states the needs for a shared vocabulary and format for text encoding and then provides directions on the development of the Guidelines (Cummings, 2008, p. 453). Since its first usable version in 1990,[107] the TEI has not only offered guidance in the encoding of textual features, but has developed a substantial international research framework. Fotis Jannidis (2009) has in fact defined the TEI as three different things at once: an organization, a research community, and a set of concepts and tags. The main difference between the TEI and other markup formats (such as, for instance, DocBook and LaTeX) is in fact its emphasis on research and scholarly applications. The TEI aims to provide not only a best-practice format for the encoding of texts, but also a support for research and scholarly endeavours. As stated by the TEI's Goals and Mission, the TEI Guidelines 'seek to support discipline-specific analysis and research approaches'[108] and they recognize the existence of a TEI research community as one of the main assets of the TEI itself. However, this last statement has been questioned by Jannidis (2009, p. 257) who, after having qualitatively examined the contributions of people to the TEI mailing list and having found that most of the contributions are produced by a handful of people, he concludes that 'the TEI is not a research community yet, but it certainly is a community'. However, since Jannidis's analysis, the community around the TEI has demonstrated its research vocation more strongly: for instance the TEI has launched a peer-reviewed open-access journal (*Journal of the Text Encoding Initiative*, first issue published in 2011)[109] and its annual meetings have become more prominently scholarly conferences, with selected papers published within the journal.

As maintained by Renear (2004) and Cummings (2008), the main significance of the TEI is perhaps represented by the fact that it not only offers support for the encoding of texts, but that it also provides a tool for a better understanding of texts and their features. The TEI has in fact produced a taxonomy of features (represented by the more than 500 elements and

[107] TEI Consortium (2013), § iv.2.

[108] Cf. <http://www.tei-c.org/About/mission.xml>.

[109] See the journal website: <http://jtei.revues.org>.

attributes) which has provided a common vocabulary to describe textual features across borders and disciplines. For the editorial community, the adoption of the TEI has been seen by many as a way to describe the text in all its observable features, to a level of detail unimaginable before (Driscoll, 2006). The possibility offered, for instance, to encode abbreviations and their expansion, or the original and the modern spelling at once has opened new perspectives to textual scholarship. These options allow one to migrate the documentation of the editorial intervention from the summative account that is normally provided in the criteria of the edition to the text most often in an introduction, to the exact place where the intervention takes place, giving the editor the opportunity to verify each intervention at leisure, and offering the same possibility to the reader, providing a much fuller and more transparent documentation of the interpretative decisions undertaken by the editor. This is one of the aspects indicated by Bodard and Garcés as best practice to qualify an edition as 'open source' and properly scholarly (2009), not to mention the requirement of the accountability of the editorial work. It will not come as a surprise, then, that a large part of the TEI community is indeed represented by textual scholars.[110] In the month of July 2013, for instance, 54 of the of 148 messages sent to the TEI List were on topics directly connected to editorial issues, and 43 more had topics of editorial interest, demonstrating how lively the editorial debate is within the community.[111]

The TEI *Guidelines* are the *de facto* standard for textual encoding in the humanities and encoding one's texts as TEI files is now considered almost mandatory for many public funding bodies. The reasons for this requirement is that TEI files, being based on an open standards such XML and being so well documented and shared in the scholarly community, are easier to preserve than other file formats, making them (in theory) sustainable for many years to come.[112]

In spite the many advantages offered by encoding with the TEI, there has been a steady series of criticism of this format and the philosophy it represents in the past few years. The reservations expressed are threefold: access, flexibility, and overlapping hierarchies. Firstly, the TEI Guidelines are a formidable publication: the latest version (v. 2.5.0) counts

---

[110] The TEI Special Interest Group on Manuscripts is by far the largest of the TEI SIGs, counting over hundred and fifty subscriptions to the mailing list. Certainly not all of them are textual scholars, but a large part is. To them one has probably to add the members of the SIG on Correspondence as well.

[111] The data analysis is mine. The TEI-L online archives are freely accessible on the web: http://listserv.brown.edu/archives/cgi-bin/wa?A0=tei-l

[112] The same cannot be said for their interfaces, for which see chapters 7 and 8.

no less than 1605 pages in A4 format.[113] For both beginners and developers to find one's way in such a big publication is quite off-putting (Burghart and Rehbein, 2012). Simplified versions of the TEI, such as TEI Lite, do not solve the issue as simplification often means that some of the real advantages of using the TEI are left out, namely its support for advanced research. Teaching materials and tutorials are abundant, but only a handful of them are thought for self-learning;[114] furthermore they stop at a level that is too low for most purposes, before the TEI can show its true potential. Many scholars find learning the TEI to be unappealing, especially because they soon realize that the TEI is not enough: to produce any meaningful output, one must also become proficient in a number of programming and markup languages, as mentioned above. The transformation of the editor into an encoder, while it can be seen as a way to improve one's capability of representing the material to be edited, it can also be an unwelcome and unsettling change, as McLouglin (2010) has demonstrated.

Secondly, the TEI is very rich and flexible, offering more than one way to encode what may appear to be the same phenomenon, with only slightly different nuances of meaning associated with the different encodings. While this offers scholars the possibility of being as true as possible to their research aims, it makes extremely difficult the development of generic tools that could help to reduce the steep learning curve required by learners of the TEI. Furthermore, such an abundance of resources is perceived as confusing and overwhelming, in particular during the early stages of learning the TEI (Burghart and Rehbein, 2012).

Thirdly, because the TEI is based on XML, it cannot offer a full and convenient support for features that requires non-hierarchical structures, namely the possibility of encoding so-called overlapping hierarchies. According to XML, for instance, if there is a sequence such as the following fictional example

~~deleted and~~ <u>~~underlined~~ sentence</u>

where some words in a text have been deleted and another sequence of words has been underlined, starting from the middle of the sequence marked for deletion and continuing after the end of the deleted sequence, these two phenomena can not be marked simultaneously as the underlining, starting as it does from within the deletion, has to be interrupted before ending the deletion, and then perhaps restarted afterward, even if there is a single stroke in the source document, as for the following example:

---

[113] http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf

[114] A remarkable exception is offered by the *TEI By Example* project: Van den Branden *et al.* (2010)

In the source: ~~deleted and~~ ~~underlined~~ sentence

In the markup: ~~deleted and~~ ~~underlined~~ sentence

The same problem arises if one wishes to encode verses and syntactic boundaries, and in many other cases. The necessity of structuring every XML file within a single ordered hierarchy lies at the base of the development of the OHCO model (De Rose *et al*., 1990), discussed at length in Chapter 2. Cummings (2008) lists a series of possible approaches that may help to circumvent the problem (p. 463), but the reality is that the problem has no real solution. For many encoding projects, with very dense markup the problem of overlapping hierarchies is a serious drawback.

One possible approach to this problem might be the adoption of a stand-off annotation, as suggested by a few scholars (Cummings, 2009; Eggert, 2005; Schmidt, 2010). According to this approach, one is to keep a 'low density' XML source file (say, a text with only minimal markup) and to keep all complex and conflicting layers of markup in separate files, which are to be instantiated on-request (or *Just-In-Time* as in Berrie, 2006) by the user (Bansky, 2010). This solution presents the advantage of keeping the XML markup at a manageable size while allowing for complex and deep annotations: these may overlap, but because they are stored separately in different layers, and only instantiated on demand, they do not conflict. Such a solution seemed so much more suitable for cultural heritage texts that it produced understandable enthusiasm which, in some cases, has been coloured by strong ideological beliefs. In particular, Schmidt (2010) maintains that embedding XML markup into a text is practically and methodologically wrong, while using it as stand-off annotation allows for a better management of knowledge and for interoperability. In his vision, in fact, embedded markup makes it impossible to share files as they are inevitably blemished by editorial interpretation. In his later intervention of 2012, Schmidt radicalizes his position and proposes the use of RDF-OWL (a web ontology language) as a standoff technology; this technology has the advantage of being non-hierarchical and therefore overcoming the limits of XML seen above. This is not a new idea: back in 2005 some early experiments demonstrated the potentials of such an approach (Tumariello *et al*., 2005), but also its drawbacks, the biggest of all being the lack of editorial tools and support at the scale of the individual scholar. While XML has a very simple syntax that can be controlled with the naked eye (at least to a certain level of complexity) and therefore can be entered and edited quite simply by typing it in, a standoff approach, whether based on XML or on any other type of annotation syntax, requires a complicate software architecture. If the task of learning XML

and the TEI is considered demanding by editors, it is not surprising that RDF is proving even more demanding and often discouragingly so, given the almost total absence of ready-to-use solutions. However, this difficulty is not limited to RDF but is true for any stand-off implementation: the few tools available are, at the moment of writing, either too complex or are centrally controlled and not customizable by the user, in the sense that one can only mark up these features which are predetermined by the software producer.[115] From a processing point of view, the loss of standard tools for handling the encoded material is also problematic and therefore requires a considerable investment to be able to develop equivalents to what is already available for XML.[116] An increasing use of stand-off annotation will certainly increase the demand for ready-to-use out-of-the-box tools on behalf of the editors and, in the short to medium term will also have the consequence of connecting even more editorial endeavours to computer centres able to handle the complexity of the data representation. One has to conclude then that, while stand-off markup seems to be a very promising long-term perspective, it seems still that it is not yet ready for general use. To become so it will have to be able to become more user friendly and available to people unable to obtain IT support from their institutions.

While a stand-off approach represents a desirable evolution from a practical point of view, from an ideological point of view, however, the use of stand-off markup can come dangerously close to the idealization of text as a stable, pure object without interpretation. In fact, as mentioned above,[117] one of the major criticisms of embedded markup is that it represents an interpretation of the text and therefore is unsuitable for sharing and interoperability. The underlying assumption is that a text 'stripped out' of its markup is a non-interpreted text (Schmidt 2010 and 2012; Eggert 2005 and Berrie 2006), which of

---

[115] At the time of writing, an exception seems to be the CATMA project, which allows the creation of one's tagset and is quite easy to use. However, its purpose is not to create digital editions, but to analyze texts from a critical analysis point of view, therefore it only proposes very basic visualization features. Encoded files can be exported as TEI but using a data structure that is almost unprocesseable. In fact, if users encode without any hierarchical preoccupation, exporting that in XML will result in a highly fragmented file which is very hard to manage with standard tools.

[116] Schmidt mention this latter issue but dismisses it in a quite curious way, by saying that once you do not have XML you do not need languages such as XSLT or xPath anymore (Schmidt 2012, pp. 141-2). This is certainly true, but it ignores the fact that they must be substituted with something that gives the same ease in handling the data. As a matter of fact none of the tools available to handle stand-off is yet able to stand the comparison with the XML family of languages in term of richness and easiness of use.

[117] See Chapter 4, p. 00.

course, it is not: an allegedly 'plain text' (that is a text without visible codes) is not a text without interpretation, but a text where interpretation is conveyed by writing conventions and implicit assumptions, as demonstrated in the previous chapter.

Stand-off markup may represent a practical and clever solution to a serious problem, the one of overlapping hierarchies, and enables the co-existence of different levels of interpretation, but it has, or should have, nothing to do with interpretation or lack thereof. Even when looking at a clean, reading text, we cannot forget its essentially interpretative nature. Scholars may decide to agree on a specific version of the text, but this can only be seen for what it is, namely a pragmatic compromise, a working hypothesis, or, to use a terminology that characterizes this book, a model of a document or of a work.

5.4. Editorial Work, Collaborative Work

One of the most striking changes determined by the provision of scholarly editions in digital format is the necessity for the editor to collaborate. Peter Shillingsburg talks about the 'village' that is required in order to produce a digital edition and the support of a specialized centre for Digital Humanities has been invoked previously as a possible way to overcome the need to acquire advanced computational skills in order to process and publish one's edition. The extent of the change required to the working habits of editors in order to be able to collaborate cannot be underestimated, and it is not necessarily painless, as pointed out by McCarty (2012, p. 2): ' "Collaboration" is a problematic, and should be a contested, term. It is often used nowadays to denote what Peter Galison has called a transcendental virtue (2004:380), that is, a good without qualification'. However we all know too well that to establish a real collaboration requires some effort.

McCarty notices how '[s]ocially the history of computing in the humanities has involved a long struggle to establish computing practitioners and non-technical scholars as equals in research' (McCarty, 2012, p. 3). This is actually what characterizes a particular form of collaboration required by the creation of digital editions and of digital resources in general, namely the collaboration between the 'techies' and the 'non-techies'. In the past, when the work of editing was supported in the delivery stage by publishers, we used to call it 'division of labour', while we call (or we should call) the work with computer engineers and developers 'collaboration'. The fact that we have changed the way we call this sharing of work with other people denotes a change in the perception of the qualitative nature of such sharing, as in some cases we have moved from collaboration to interdisciplinarity.

For centuries, editorial work has been more or less the province of a single scholar,[118] able to control and supervise all stages of the process up to the point when the edition was delivered to a publisher; the publisher was then in charge of typesetting, preparing the layout, copy-editing the manuscript and so on. With the advent of digital publication, the scholar is forced to cooperate with other experts from the very beginning. Texts need to be digitized, checked, annotated, encoded and transformed; facsimile images need to be annotated, elaborated and transformed into suitable formats. If the edition is to have an electronic output, then an interface must be designed, including both functionality and graphics; search engines and browsing facilities must be designed and deployed; infrastructural support is needed, such as networking and maintenance of servers; and so on. Editors are also asked to be involved in tasks traditionally the province of a publishing house; again these tasks are now to be performed in cooperation with other experts, most probably from other disciplines. Kenneth Price pushes the boundaries of collaboration even further by suggesting how collaborative project teams must also collaborate with 'librarians, archivists, graduate students, undergraduate students, academic administrators, funding agencies, and private donors' (Price, 2009, p. 437).

The experience of collaboration is not always positive. Lynne Siemens (2009) has provided a good description of the tensions one can find within the team of a digital humanities project, the main difficulty being one of reciprocal understanding across disciplines. The need to negotiate concepts like authorship and responsibility are also likely to become an issue (Siemens et al. 2009). However, collaboration can be also a very rewarding experience. The experience of working on the *Jane Austen's Fiction Manuscripts* digital edition, for instance, has been a fulfilling journey and as a team we experienced little of the incomprehension and misunderstanding outlined by Siemens; perhaps the reason for this is because both sides of the team, domain experts and digital experts, spoke the same language, the language of scholarly editorial practice (Sutherland and Pierazzo, 2012). A recent *festschrift* publication in honour of Harold Short (Deegan and McCarty 2012) entitled *Collaborative Research in the Digital Humanities* has demonstrated how fecund and faceted collaboration can be and how impactful and widespread are its effects. Sculley and Pasanek (2008) also remark how 'in collaboration, we discover an opportunity to examine our practices' and that results need to be 'simultaneously meaningful in two disciplines' (p. 421).

Collaboration, then, is not only a necessity of digital scholarly editing, but also an opportunity. The availability of the Internet and of web-based communication tools have

[118] A fact that, according to McCarty, is likely to make humanities scholars 'vulnerable to forgetting the social dimension of knowledge' (McCarty 2012, p. 4).

enabled distance teamwork to a level unimaginable only a couple of decades ago. These factors open up new forms of collaboration, but also create issues which were previously almost unknown in the humanities such as the need to manage the project workflow very closely, storage and access to project data and forms of internal communication.[119] The task of coordinating the work of several people, all of them managing the same batch of data in various ways, possibly working from scattered locations, with different working habits and different skills, is by no means to be underestimated. An article of 2010 outlines the challenges and the opportunities offered by an international distributed team working on the edition of the correspondence of Giacomo Puccini (Pierazzo, 2010). While the communication channels offered by the Internet have made possible as never before the creation of a team , the dislocation and collaboration among people coming from different disciplines have determined the choice of technologies and the editorial workflow. For instance, it was clear that the further the distance among the project team members, the more user-friendly the technical solution would have to be.

Most of the discussions held in this chapter have assumed that a fundamental change in the workflow lies at the base of the production of digital editions, namely the end of the relationship between academics and publishers. This relationship was inaugurated at the end of the Fifteenth Century with some famous collaborations (the choice of words here is not accidental) such as the one, for instance, between Aldus Manutius, prince of printers, and Pietro Bembo, humanist and accomplished textual scholars of classical and modern languages (Lowry, 1979). Through the centuries this relationship has been at times fruitful, but also problematic; the constraints of the printed page and economic considerations regarding the commercial viability of certain editorial products have encouraged editors to engage with the digital from the earliest hour, in the belief that the digital would offer them the freedom of pursuing their scholarly goals in more depth. This hope has proven to be well founded, with some remarkable works of scholarship produced; however, the so-called freedom has come at a price, namely the retraining needed in order to acquire new and ever-expanding skillsets and to completely reconsider the way the editorial work is done, its

[119] See Kamp (2008, p. 78): '[…] according to observers, research activities within the scientific community as a whole are increasingly determined by the character of a (disciplinary as well as trans-disciplinary) "team work" among the scientists of globally scattered institutes. So, the academic and scientific communication in the internet brings about new and promising possibilities, but it generates new desiderata regarding cognitive authorship, copyright, and reputation, among others'.

method and its purpose. As maintained before, this change is likely to represent the best thing that could have happened to textual scholarship, although not an easy one.

Yet, however changed, the possibility of a renewed relationship with the publishing world is not dead. In the next chapter publishing methodologies and the role of publisher in digital scholarly edition will be investigated.

6. The Publication of Digital Scholarly Editions

What happens once the edition is ready for the public? If the editors expect to publish their edition in print then the answer is quite simple, or allegedly so: they will send their edition to the publisher, then there are a few operations that requires a certain interaction between the publishing house and the editors, such as responding to queries, revising the text after copy-editing and peer reviewing, proof reading, and then the edition gets published. A convenient division of labour allows for the editor to concentrate on the establishment of the text and the provision of the paratextual material, while all the technical work is undertaken by the publisher, more often than not in conjunction with a printer. Such a workflow is described by Bree and McLaverty (2009) as being valid and useful:

> The traditional publisher-scholar partnership for specialist academic work in print form, where scholars provide the material, the scholar and the publisher discuss how it can be published, and the publisher publishes it, seemed to us to remain both valid and useful in practical terms. (pp. 128-9)

This is of course an idealized situation that in practice allows for a lot of variation. For instance, in some cases editors are required to provide 'camera ready copy', meaning that editors must conduct some essentials tasks on their own such as copyediting and page formatting, and the publisher only produces the physical copies of the book and provides distribution channels. In other cases, editors are involved in the design of the covers, or other design activity. It should also not be forgotten that the division of labour which we consider to be 'the right shape of things' is only a recent acquisition. In the Early Modern period editors used to live with and be paid by publishers. Some of them were publishers themselves. This is the case, for instance, for Anton Francesco Doni, a Florentine intellectual of the Sixteenth century; in his complex and picturesque life he was resident editor for at least two publishers in Florence and Venice, author of a considerable number of unconventional literary works, as well as a publisher on his own. Contemporaries reports the wonder of seeing him writing and editing in the middle of the noise of the print.[120]

[120] 'Poi vi dico, e ve lo affermo da vero cristiano, che le sue opere [of Doni] son composte da lui mentre che le si stampano, e se venisse quattro feste fra la settimana non avanzerebbe gli stampatori di due righe, anzi dà la copia apena che basta e bene spesso fra i romori delle stampe: perché manca, gli bisogna scrivere; e quanto manco tempo ha, allora scrive con più attenzione, sí che non solamente voi, ma tutto il mondo si maraviglia': Doni (1998, p. 47).

During the Early Modern period the editorial work was more or less the exclusive domain of publishing houses, and indeed this actually identified them, as demonstrated by the fact that in several languages such as, for instance, Italian, Spanish and French, the word 'editor' also means publisher.[121] In the modern period we have seen the gradual differentiation of the editorial and publishing functions, with printing becoming a separate activity and editing being split between copyediting, which remains within the publishing house, and scholarly editing, which has migrated to universities.[122] Tanselle (1995) in mapping editorial activities distinguishes between Historical and Nonhistorical editing, where the latter is the activity undertaken in publishing houses with the purpose of improving 'literary excellence or marketability' (pp. 13-14). In the early days of print the two activities were closely related and performed seamlessly by the same person.

Editing is then not only *ars*, but also *techne*, and it has always been that way (Robinson, 2004, citing Housman 1921). However it is certainly true that the publication of an edition in digital form has forced editors into much more *techne* than was customary in the recent past, undertaking unusual tasks and becoming literate in many computational formalisms, which, as clearly stated by Bree and McLaverty, are felt to be a long way off from 'making a direct contribution to scholarly research' (2009). In the previous chapter the evolution of the editor as encoder has been examined; here we will examine the evolution (or the devolution, if we consider the not so distant past) of the editor as publisher. The two concepts are strictly related, as encoding requires the editor to decide which output to provide for the encoded texts. Encoding, providing an output and publishing the outputted files are all considered part of the same conceptual node in the definition of the rationale for digital documentary editions, as scholarship and meaning can be found in both the encoded texts and in the scripts that actualize them (Pierazzo, 2011). Markup and encoding not only determine what some specific features of texts are and what they do, but they also imply (or make explicit) the way the text should be displayed. They then assume some of the functions of the publishing workflow, as remarked by Deegan and Sutherland:

> Encoding provides some of the mediating functions represented in book production by a range of publishing and printing-house practices – copy-editing, formatting, text composition, setting in type and printing – all of which impose interpretative values, and over all of which the literary editor has little control. (2009, p. 79)

---

[121] 'Editore' in Italian means both editor and publisher; the same goes for 'editeur' in French and 'editor' in Spanish.

[122] This is obviously a simplified picture: the editorial/publishing world is much more complex than this, as the existence of University Presses demonstrate.

The authors recognize how encoding deeply changes the way we understand and handle texts: 'the computer as an inscription technology redistributes the range of specialist activities that contributes to the textual production in book form' (ibid.). This redistribution is still under way, with editors, publishers and centres for Digital Humanities all attempting to find their role and their specific contribution in relation to the new medium. But it is also the definition of the format of this publication that is still under way. Many discussions revolve around which features digital editions should have and what differentiates them from a digital archive or a digital library; whether they should or should not have a print counterpart, and whether it should look like a book or not.

Since the advent of the web, the role of publishers as intermediaries between editors and the public (the users) has been questioned, generating what Bhaskar calls the phenomenon of disintermediation, namely the 'unbundling of publishers from the literary value chain' (2013, p. 61). The relative simplicity of web publication via HTML, made even easier by the use of WYSIWIG software for the production of websites, has fuelled the perception that the cultural mediation offered by publishers is no longer necessary. The demand for advanced tools for computer-assisted philology that allow scholars not only to process the text, but also to enable online publication (or print-alike) leads in the same direction. A comment by Martin Mueller on the TEI mailing list of 18 January 2013 has fuelled long discussion on the usability of the TEI, then he alleged that the TEI is unsuitable for an English graduate student with no IT support to put 'on the Web some 17th century poems she has transcribed from a manuscript'.[123] The implication is that, in order to be successful, the TEI should support self-publication. This expectation assumes that not only publishers, but also web designers, web graphics, and programmers are really not necessary to modern digital editions, as the editor is supposed to control the entire workflow, from the manuscript to the web thanks to computers. Is this a reasonable expectation, and, if not, then what is the role of publishers in this environment?

The production of editions for internal use has always been part of academic life, such as the production of editions as part of doctoral dissertations which were meant to be circulated in a restricted circle until judged good enough to be made public. At that point, and only at that point, if ever, were publishers involved in the workflow. With the Internet things have changed dramatically (or are expected to have) in that respect as well: a digital edition

[123] See the message at the following address: <http://listserv.brown.edu/archives/cgi-bin/wa?A2=ind1301&L=tei-l&F=&S=&P=57972>.

requires in most case a digital environment like the one provided by the Web, which is public in most cases. Private circulation can be guaranteed by locking the content with a password or by using private intranets, but the workflow and the technical requirements of publishing in an internal network or in the public sphere are more or less the same. This is probably the framework that Mueller had in mind with his provocative post: tools for self-publication and the support necessary to produce an edition good enough to be read, in the same way that most word processors are able to produce printable pages and digital documents that can be read on the screen. How realistic is this provision of computer-assisted tools for self-publication? Here, as in many other cases, the answer is 'it depends', with variable factors such as the complexity of the edition, the layout, and the expected functionalities. The fact that not many tools like this currently exist suggests that it is actually not that easy.

6.1 Work in Progress

The previous discussion mentioned another crucial point regarding the evolution of the modes of publication in a digital environment, and it is the concept of 'good-enough' which in turn, as we will see, also draws in the connected concept of the open-ended edition. What is meant here is that, while a print book has been through many quality checks justified not only by academic requirements but also by the more pragmatic consideration that once published it cannot be modified, a web publication can be updated and modified at leisure, and therefore on-going publications, with materials that can be modified at any point, have become the norm. For instance, up to October 2013 one could see the warning message on the *Henry III Fine Rolls Project* website (Carpenter *et al*., 2009-2013): as the Internet publication precedes the print publication that completes the range of deliverables of that project, users are warned that using a particular subset of the available material (i.e. the subset that has not yet been fixed in print) is done at their own risk

> The translation of the rolls presented here for [a particular range of rolls] is still in progress and is subject to change (including the numbering of the entries).
> Anyone wishing to cite entries in these later rolls are urged to contact the project team to ascertain the latest state of progress.[124]

[124] The funded project has now reached its end and the quoted sentence has been substituted by a similar disclaimer: 'Although the project has now finished, users are encouraged to report any errors and comments through our feedback form as updates to the website will continue on a periodic basis'. <http://www.finerollshenry3.org.uk/content/calendar/calendar.html>

More details about the relationship between print and digital for this particular project are discussed below; here it will suffice to point out how a rolling publication has become part of the digital environment. Furthermore, incomplete editions or work-in-progress editions have also become part of the editorial scene. The digital edition of the *Codex Sinaiticus*,[125] for instance is far from being completed: each web page of the edition is intended to present the facsimile of the manuscript on the left, the transcription on the top-right and translation on bottom-right, however the translations are missing from all pages.[126] This is just one example of a situation that has become standard and follows changed editorial frameworks. In fact, substantial digital editions are likely to be supported by public funding which is normally bestowed for an average periods of about three years; however, editorial enterprises notoriously take a very long time. According to a recent publication by Budrick *et al*. (2012) 'project-based scholarship exemplifies contemporary Digital Humanities principles', with one of its defining characteristic being the fact that it is 'iterative and ongoing, rather than fixed or final' (p. 130). Project-based scholarship also requires 'creating visibility within a scholarly community and among potential users and future contributors' (ibid.), and therefore publishing a bit at a time, showing that something is happening, building a community of users, inviting feedback and discussions is considered a constitutive part of digital scholarship. This evolution can also be seen in the light of the diffusion of the habit of making available pre-print papers within an academic or disciplinary repository. This tendency, connected with the concept of Open Access (for which see Chapter 7), is having a huge impact on scholarship, as argued by Pritchard (2008), for example. The habit of accessing 'good enough' pieces of scholarship is become part of scholarly work, with more and more higher education institutions encouraging, if not requiring, their employees to store pre-prints on institutional repositories. The growing success of an academically-flavoured social networks such as Academia.edu or the Research Gate is part of the picture.[127] The scholarly debate, in common with every aspect of our life, has become faster: discussions and arguments are done in 'real time' on the web, before and in parallel to traditional channels such as published articles and reviews which normally take months if not years to circulate. Speed, easiness and accessibility and a wider impact are characteristics of online self-publication, a form of dissemination that becomes more

---

[125] Cf. <http://www.codexsinaiticus.org/en/>. The *Codex Sinaiticus* project has been developed in a partnership between the British Library, the National Library of Russia, Leipzig University Library, St. Catherine Monastery.

[126] Website checked on 20th August 2014.

[127] At the moment of writing, Academia.edu counts more than 5,000,000 users.

common by the day and that promises to change many of our established expectations about the way scholarship is accessed and produced.[128]

One can see here how these values contrast with the traditional expectations of textual scholarship, the purpose of which is thought to be the production of durable, accurate and reliable editions of texts. How can on on-going publication, updatable and updated, be considered reliable? How can a changing edition be quoted? When does a digital edition becomes good and not just good enough? The concept of work-in-progress publication has been seen as liberating by many, as outputs can be circulated much earlier than in the past, with a potentially greater and more timely impact on future scholarship. However, this has produced a series of concerns on many fronts, not least its impact on career building and the academic and scholarly evaluation of the revisions. In fact, if digital editions can be published while still not ready, more work is expected after publication, which will probably not be visible at first sight. How can one justify, post-publication work, in particular when funds have run out? And even if the editor has decided to wait until the work has reached a mature stage, when all the checks and controls have been made before publishing online, how can one resist the temptation of updating when the inevitable mistake will be found? This on-going process of correcting seriously undermines another important scholarly 'infrastructure', namely post-publication reviews. Some of the most important scholarly contributions in many disciplines stem from reviews of someone's publications. One can mention here, for instance, Carlo Dionisotti's contribution on the practice of 'vulgarising' (i.e. translating into vernacular Italian) works of classical antiquity in fifteenth-century Italy, which stemmed out a review of *The Classical Heritage and its Beneficiaries* by R. R. Bolgar (1954),[129] or the article by David Dumville 'The Beowulf Manuscript and How Not to Date It' (1998) which takes its basis from a review of Kevin Kiernan's edition of *Beowulf*. Digital editions, and in general digital publications, allow their authors to correct them after publication; this makes the practice of reviewing almost meaningless. An unfavourable review, such as the two mentioned above, both famous for their harshness, can encourage the author of the criticized resource to simply fix mistakes and change those features that have met with criticism by a particular reviewer, without necessarily leaving any trace of what was there before. The result will be that the review will become obsolete if not pure slander. Bodard (2008) states it quite clearly:

---

[128] Cf., for instance, Lawrence (2001), a study on the impact of online availability conducted when this phenomenon was just beginning, where the increase of citation for online papers with respect to printed ones was estimated to be over 280%.

[129] See the chapter: 'Tradizione classica e volgarizzamenti' in Dionisotti (1967).

> A more significant problem arises when one considers the status of a work that has to interact with, and be reviewed within, the world of peer reviewed, cited, traceable, and replicable scholarship. If a later work of secondary scholarship cites an online project such as *Inscriptions of Aphrodisias*, perhaps arguing with an interpretation of the evidence or adding new data to the historical debate, then that citation needs to be followed back to the source by a reader. If the editor of the cited project reads this new article, agrees with the arguments, spots an error, or otherwise sees a need for change, and simply updates the original site, then the reader following the reference back from the secondary source will no longer see the text upon which the scholar was commenting. This is clearly an unacceptable state of affairs, even if the original text can be recovered from an Internet archive or cache in some form (§ 32).

This change is not necessarily a bad thing, and reviews could be replaced with a more friendly feedback process that accompanies the evolution of the scholarly edition in a more collaborative way. It is a change, though, upon which the community has not yet reflected enough, and one of the factors that feeds into a general mistrust about digital publications from a large portion of the academic community. Chapter 8 will continue the discussion on this topic, in particular assessing the trustworthiness of on-going publications.

6.2 The Shape of the Digital

One fundamental difference between print and digital publication is that while the way a printed text will look in the hands of the reader is more or less predictable,[130] this is not so certain when it comes to websites. Any web designer knows that, in spite of declarations from software developers that they all follow the same web standards, it is a painful and almost prohibitive task to make the website look more or less the same on all Internet browsers. This is so even before one considers the release of new versions of browsers, an event that happens at an alarmingly fast rate and which may change the way the text is displayed completely. Such changes may also generate new bugs and problems in pre-existing websites, and the resources to fix those problems may or may not be available at any particular moment.

Furthermore, with the diffusion of such small portable devices as smart-phones, netbooks and e-readers, users can face very different experiences in accessing the digital

[130] A century (if not more) of scholarship in analytic bibliography has taught us that this affirmation is not always true, but we ask the readers to concede our simplification for the sake of argument.

resource, sometimes with the risk of misunderstanding it entirely. For example, a transcription may have been designed only to be read alongside an image, but many portable devices are unable to display both image and text on the screen at a readable size. So far, print culture has produced objects that are stable from many points of view: they have in-built durability; they cannot be changed overnight; and they look the same in the hands of different readers in different countries and at different times. By contrast, digital culture has still a long way to go in all these respects: longevity, reliability and stability are still distant goals.

Publishing digitally may also mean several things at once, as there are many forms of digital publishing. It might even not mean to produce a digital output at all, but to use computational methodologies to produce printed books, as argued by Daniel O'Donnell (2008). He maintains how the word 'digital' is taken to mean at least three different senses:

1. A display format (i.e. 'screen')
2. A processing and organizational technology (i.e. markup languages, database software, etc.)
3. A way of understanding texts and culture encouraged by the use of computers (namely, cyber-culture).

In his opinion, sense 1 is almost accidental, and what qualifies an edition as digital is the application of digital methodology, as intended by sense 2. While O'Donnell is certainly right in advocating for the recognition of the digital nature of the workflow and not only of the final product, it is also true that only by applying a digital methodology to produce a digital output, perhaps alongside a print publication, that the true potential of the computer can be exploited, taking advantage of the interactivity and the enhanced representations offered by the digital environment.

Digital outputs can assume several shapes and may be hosted by a variety of supports. Pioneer digital editions were mostly published in support such as tapes, floppy discs, CD-ROMs and DVDs. This is the case, for instance, for the *Canterbury Tales* series of CD-ROMs, *Electronic Beowulf*, and many others. This tendency is drifting away, and for some very good reasons. The publication on supports like CD-ROMs is in a sense very convenient. Because it results in a physical, tangible object that can be sold, purchased and catalogued more or less in the same way as a printed book, it fits an established workflow for libraries and can also be quoted more easily. However, it cannot be updated. The rapid evolution of computers from both a software and hardware point of view has determined a rapid obsolescence of editions recorded on digital supports. O'Donnell (2008b) gives an account of the accessibility of editions on CD-ROMs less than ten years old. His findings are

discouraging: most of the editions he examined were completely unusable only a few years after their publication, a result that seriously undermines the whole concept of producing digital editions. However, while the general picture was very bleak, there were also some bright spots. O'Donnell found that the editions that relied on some open technologies such as HTML, for instance, were still readable with some limitations, even if they were on CD-ROMs. Sustainability is then the biggest drawback for publication on CD-ROM or DVD. This form of publication has not been completely abandoned (for instance the edition of Dante's *Commedia* edited by Prue Shaw was published as DVD in 2010), but is much less common than it used to be.

The most common framework for publishing digital editions today is the Web. The Web offers not only a highly flexible environment to accommodate the peculiarities and sophistication of digital editions, but it is also more sustainable than CD-ROMs and DVDs will ever be, as it can be easily updated to face evolution of software such as web browsers, plug-ins and fonts. For instance the *Jane Austen's Fiction Manuscripts: A Digital Edition* was first published in 2010, but then updated in 2012 in order to accommodate, among other things, the change of size of the font used, which had somehow lead to a relocation of some of the interlinear insertions (Sutherland, 2010). Yet, this updatability and adaptability of web publications comes at a price of fragility and lack of embodiment. A web publication is easy to put online, but is also easy to take down, willingly or not. Web hosts change ownership, location, and anybody that has ever been involved in an upgrade of a computer server knows how many things can go wrong during an operation which should in principle be a routine one. Another complication with web publications is how to quote from them. The scholarly method is based on the fact that it will be possible always to trace quotations and references in order to verify assertions and conclusions. Print culture has elaborated several ways to facilitate this fundamental activity, such as the standardisation of imprints, the elaboration of referencing systems and the creation of canonical references for classical authors. In contrast, web culture is struggling with this. The lack of a formal reference tool such as page numbers, for instance, is challenging the use of digital resources. Referencing systems give only partial support for web resources; for instance the Harvard and MLA referencing systems requires a web page to be cited by author, when available, but authorship of a web resource is very rarely available in the same way it is for a book. Who is the author of the *Jane Austen's Fiction Manuscripts: A Digital Edition*? Jane Austen? Kathryn Sutherland, the lead editor of the text? José Miguel Vieira, the lead web developer? The project team as a whole?[131] In this

[131] It is interesting to note that while the digital recognize this as an issue and question itself with the authorship and credit of digital artefacts, this does not happen in print culture.

specific case the website offers a clear format for citation, where the principal intellectual responsibility has been recognized to be Kathryn Sutherland's; however, not all websites provide guidance for citations, leaving their users to choose any format they wish or as required by their style sheet.[132] Another common request of referencing systems is the date of access of the website in question, but this is only partially meaningful: if the function of a good reference is to be able to trace back a quotation, then knowing that a particular resource was available at a given moment in time does not offer any consolation or help if that resource is not accessible any more. However, the date may help in locating a particular version of a website if it is located in an Internet archive, such as, for instance, the Way Back Machine.[133] The same applies to the provision of the URL of a web resource, which is also subject to change and which has now been recognized as non-essential for the MLA referencing system, as people that want to trace a digital resource may more effectively use a search engine to locate it rather than clicking on a web address which may have become obsolete.

The sense of instability offered by web publications is perhaps one of the biggest drawbacks of this form of publication and it is perhaps at the heart of many and diversified attempts at dual publication in print and on the web, as we will see shortly.

As common as web publication may seem to be for scholarly editions, this is not the most common format of digital publication any more. The explosion of the eBook and tablet market of the past few years has revolutionized the digital landscape again. Amazon, for example, announced that their US sales of eBooks in 2010 surpassed that of printed paperbacks by 15 per cent (Amazon, 2011). After many years of the eBook market seeming on the point of evolving from niche into mass (see, for instance the optimistic evaluation of Hillesund in 2001), once this finally happened its consequences have been disruptive for the publishing world, and in particular for commercial bookselling. One of the main factors that has contributed to the establishment of eBooks is the introduction of reading devices such as the Kindle connected to online stores such as Amazon. Following the same business model provided by Apple with the iPod as device and iTunes as virtual store, which has rapidly revolutionized the music market, the Kindle has provided not only a reading support, but also the convenience that has made the loss of physicality affordable. In fact, one can buy a book

---

[132] No guidance, for instance, is provided for citing the *Codex Synaiticus* project. Here it has been decided to quote as editors Juan Garcés, project curator in London, and Zeki Mustafa Dogan, Technical Co-ordinator in Leipzig, following the indications found in the *Contact* page <http://www.codexsinaiticus.org/en/contact.aspx>.

[133] The Wayback Machine can be found at <http://archive.org/web/web.php>

and begin reading on the Kindle in a minute or so, at any given time and from any given location, provided that there is an Internet connection. As for quality, digital music in the form of Mp3 is much lower quality than the one provided by music CDs, yet this factor is not a major consideration for most consumers: the switch to digital portable music has been quick and apparently irreversible. The same seems to be happening with eBooks: while printed books still provide an aura of quality (whether deserved or not remains to be ascertained), the convenience offered by eReaders and tablets seems to overcome much resistance, at least judging by the increase in sales in the past few years. The same can be said for the loss of tangibility of books: the tactile experience of the printed book, the sense of ownership it generates, its capability of hosting memories of the readers and readings, all of this is lost in eBooks; however, this loss seem something that the users of eReaders are able to cope with without too much regret, again by judging by the sales figures.

The eBook revolution has only slightly influenced the publication of digital scholarly editions. One of the reasons why editors choose to produce digital scholarly editions is the interactivity and advanced research functionalities offered by the digital medium; however eBooks only provide a digital, resizable version of print publications, with, in most cases, the only interactivity being offered by the haptic metaphor of turning the pages; as such they present themselves as 'digitized' books, and not 'digital' books, to borrow once again the handy distinction by Sahle (2008). It is not a surprise that eBooks are only occasionally chosen as the ideal output format for scholarly editions. And if they are offered as eBooks, these are normally alternative formats offered by publishers of printed editions in PDF, and not as resizable eBooks, give the fact that complex page layouts are not controllable in such formats (as they are in ePubs and Mobi, for instance).

Nevertheless, tablets do offer a format that seems suitable for hosting digital scholarly editions: not via the eBook, but the app. Apps (mobile applications) are software that run on mobile devices such as smart phones and tablets, and they hold a considerable place in the mobile market.[134] They are relatively easy to develop and offer captivating, haptic interactivity which looks like a promising environment for the provision of scholarly editions with an eye to the general public. The digital publisher Touch Press, the mission of which is 'to create new kinds of books that re-invent the reading experience'[135] has published

[134] See the report on *The state of the Tablet Market* according to which in 2013 apps have generated a global revenue of $8.8 billions, counting for 35 per cent of all mobile market revenue, with the perspective of growing at a vertiginous speed (http://tabtimes.com/resources/the-state-of-the-tablet-market).

[135] See <http://www.touchpress.com/about/>.

several apps featuring editions not very dissimilar to scholarly editions produced by academics on the web. In particular the apps for Eliot's *The Waste Land* and Shakespeare's *Sonnets* were launched in 2011 and 2012 respectively with substantial success.[136] These multimedia apps offer images, audio and video, as well as the texts and commentary and represent a very interesting take on editions of literary works aimed at a non-specialist public (Arbuckle 2014). The choice of apps as an output format for scholarly editions offers some benefits, such as the possibility of taking advantage of very user-friendly interfaces and of integrating with a playful environment,[137] both of which can open new readership for scholarly editions and collaboration with those publishers who seem interested in investing on the mobile market, as the case of Touch Press demonstrates. On the down side, the sustainability of apps is questionable and potentially as poor as that of CD-ROMs. With the operating system on tablets being updated every six month on average, the stress on the producer of apps to keep up with this evolution may prove to be too difficult to sustain in the longer term, in particular for an academic product, which is normally funded only to produce the digital object, and not to keep it updated. Publishers, however, may be in a better position to cope with the cyclic need for updating. Touch Press, for instance, have turned the frequent release of new operating systems to an advantage, using this as an excuse to produce a new release of their works and therefore re-marketing them again. However, whether on the Web, an eBook, or an app, the problem remains that digital objects require constant attention after their initial release, and this is unlikely to change in the foreseeable future. It also raises many concerns regarding the sustainability of digital publications, and this in turn affects their perceived trustworthiness, as we will see in Chapter 8.

6.3 Dual and Hybrid Publications

This overview of possible digital outputs for digital scholarly editions has introduced into the discussion the key figure of the publishing world: the publishers. The relationship between digital editors and publishers has not been at its best for the past twenty years, and it shows. In 2011 Touch Press released *The Waste Land* app, which, in addition to T.S. Eliot's poem, includes multimedia content, rich annotations, and facsimiles of the manuscripts, but,

---

[136] In a private interview (16 September 2013) the founder and CEO of Touch Press, Max Whitby, has declared that since its first launch on 2011 the *Waste Land* app has sold 25,000 copies, with productions costs having been covered within six weeks from the launch; a comparable figure characterize the sales of the Shakespeare's *Sonnets*.

[137] Such environment also offers the potentials of being able to sustain the gamification of editions described in Chapter 1.

interestingly, no named editor. This absence is symptomatic of the increasing separation of publishers and academic scholars in the digital world.

In the Nineties, scholars celebrated the possibilities offered by the digital medium as a way to free themselves from the 'tyranny' of publishers who were seen as dictating limits on the format and quantity of material to be fitted into printed publications. As a result, most digital scholarly editions have been and are produced directly by universities, academic computing services or by private computing companies working under the direct supervision of scholars, but without the fundamental professional and marketing expertise that publishers can provide. In contrast, publishers of scholarly editions have often looked ambivalently at the digital medium, in most cases uncertain about the uptake of digital books and concerned about issues of copyright and piracy. For around two decades the situation was a stalemate: a few brave scholars ventured alone onto the web, fuelled by enthusiasm for the capabilities of the new medium and by the hope of securing major research funding, while publishers mainly kept printing on paper, more or less leaving the scholars to experiment with the new medium. The sudden arrival and rapid market domination of tablet devices has shaken this situation: the boom in sales of eBooks and apps seems to have pushed publishers of scholarly editions toward the digital market. An example of this is the launch of the Oxford Scholarly Editions Online project. It is interesting however to note here how 'Scholarly Editions Online' essentially means the provision of printed books in PDF format, a choice that looks rather reductive compared to the way digital scholarly editions have been conceived by Digital Humanities scholars, these are 'digitized' and not 'digital' editions.

This is a generalization, of course, and collaborations with publishers have always taken place. One example is the Samuel Beckett Digital manuscript project which is a partnership of the University of Antwerp, University of Reading, and the Harry Ransom Center at the University of Texas at Austin with the University Press Antwerp (UPA). At the time of writing the manuscripts of *The Unnamable*, *Stirring Still* and *what is the word* are available online via a subscription system that is managed by UPA, but more are scheduled to come. The partnership fulfils the double goal of providing a publications status to the digital publication on the one hand, and a financial infrastructure, the revenue of which is used to pay royalties to the Estate of Samuel Beckett for the permission to present the unpublished material, which is still under copyright.

Other types of collaboration concern dual publications, that is, the publication of a scholarly edition both on the web and in print. This form of publication has taken many forms and it seems hard to generalize about its purpose and level of success, therefore it might be more fruitful to examine some case studies. The first of these case studies is the

Cambridge Edition of the Works of Jonathan Swift. The editors in chief are Linda Bree and James McLaverty, who in 2009 gave an account of the rationale behind the hybrid publication of Swift's work. The edition has been published following the traditional setup for an endeavour of such proportions, with editors-in-chief and editors for each volume, a format which was felt to be more manageable than a huge Swift database (p. 128). But beside the eighteen volumes which are being published by the Cambridge University Press, a digital archive has been provided to support and fill the limitations of the print editions. The archive provides access to the witnesses used to prepare the critical editions, as well as supporting materials such as full documentation of accidentals in parallel and alternative versions. With this publication model there is no overlap of content between print and digital, but instead we find complementarity, and while the print edition remains the authoritative point of reference for scholarship, the digital archive (and not edition) provides the supporting material which cannot be supplied by the print.

The second case study is offered by the *Henry III Fine Rolls project*. The project encompasses a hybrid edition which includes a print edition published by Boydell & Brewer and an almost identical edition available online. The rationale for this type of publication was described by Ciula and Lopez (2009). The authors reflect on the significance of producing the edition in print from within the scholarly community to whom the edition is targeted: 'along with the materiality of the primary source, the materiality of their scholarly editions about the sources is central to the research process' (p. 131). The two mediums, print and digital, live side by side without one superseding the other or having a preeminent position, in contrast to the previous case study, the two components shape each other:

> These two publications are not necessarily perceived, or perhaps more importantly, used as separate resources with rigid boundaries between them. For a scholar interested in the historical record, the reading of the Fine Rolls edition as the seeking of information related to it is a comprehensive process that does not stop when the book is closed or the browser is shut. (p. 133)

Whether or not this was the intention of the authors, the quotation seems to suggest that of the two activities mentioned, namely 'reading' and 'seeking information', the former is assumed to happen via the printed book, and the latter via the digital edition. In this case, then the motivation for the dual publication is not to overcome the limitation of the print medium, but rather to offer different access to the material in order to suit different use cases, in addition to the attempt to provide a material support to scholarship, as stated above. But is it really the case that people read in print and seek information online? The scholars are not too sure about this and point at future research on the topic: 'we believe that further research

into the usage of these parallel publications will lead to a better understanding of scholarly needs' (p. 132).[138] However fecund dual publication may be – and Ciula and Lopez make a very good argument that it is – this model can cause some problems too. While the print volumes represent discrete units that cannot be modified or updated once they are published, the web content of the *Fine Rolls Henry III* can and indeed has been updated since. In an on-going project, new research on more recent material is likely to cast a different light on older material or to uncover previous mistakes. The correction of the digital content can happen easily, without even acknowledging it, while for print this is normally more complex. In this case an 'Addenda and Corrigenda' section is provided in each volume after the first one, and also posted online,[139] but these solutions can hardly hide the fact that the two editions are misaligned, with the web edition providing a much more updated version than the printed one.

The third case study is the *Cambridge Edition of the Works of Ben Jonson*. While this presents in many respects a similar intellectual and practical framework to the edition of Jonathan Swift examined above, the publication model differs substantially. Here, again, the digital version that is being prepared will include the material that was not possible to fit within the covers of the seven volumes published by Cambridge University Press, such as, for instance, '90 old-spelling texts, 550 contextual documents, 80 essays, several hundred high-quality images, and 100 music scores; it lists details of more than 1300 stage performances, and has a cross-linked bibliography of over 7000 items'.[140] However, unlike the Swift archive, this will also contain the full text published in print, interlinked with the old spelling version and the witnesses. Access to the digital edition will be managed by CUP, the revenue of which will contribute to the long-term maintenance of the digital edition. In this case, then, the print edition does not represent the only location for the authoritative version of the text, but rather a sort of concession to an established form of accessing scholarship. The two editions, the print and digital, have their own individuality and independence, while still building upon each other.

These three case studies do not cover the whole phenomenology of dual and hybrid publication. What they demonstrate, however, is the degree of experimentation in trying to combine the old with the new, in the attempt to renew the scholarly edition without losing the

---

[138] Cf. Chapter 6 on this very topic.

[139] Cf. Carpenter *et al.* (2009-2013), <http://www.finerollshenry3.org.uk/content/book/addenda_and_corrigenda.html>.

[140] See the 'About' section of *The Cambridge Edition of the Works of Ben Jonson Online* website <http://universitypublishingonline.org/cambridge/benjonson/>.

advantages offered by the print publication. The fact that several editions are attempting to reinvent publication strategies without breaking with the tradition can be also interpreted as a symptom of uneasiness with the digital medium. In spite of the many advantages offered by digital editions, it seems that the scholarly community still finds the drawbacks too compelling to be ignored.

The results of a survey undertaken by Porter (2013) seem to agree with this evaluation. In her article, Porter compares the results of two surveys: one she undertook in 2011, which involved 169 people, and the other in 2002 with 98 people; in both cases, all participants described themselves as 'Medievalists'. The comparison shows that although the use of journals and facsimiles in digital format has gone up considerably in the ten years between the two surveys,[141] the only type of publication that has gone down is the use and reliance on digital scholarly editions. Specifically the number of respondents who declared that they used mostly print editions rose from 44 per cent to 58 per cent, while the use of electronic edition only shows a modest raise, from 0 per cent in 2002 to a mere 7 per cent in 2011. Commenting on her findings, Porter suggests a disconnection between what the Digital Humanities believes a digital scholarly edition to be (namely a purpose-built digital resource) and what university administrators (and scholars, one is tempted to add) think it is, namely something like Google books; however she does not attempt to explain the apparently scarce use of digital editions. The confident use of other digital resources such as journals and facsimiles seems to suggest that the problem, if any, is not with the medium, but with the genre itself. In a book of 2009 Deegan and Sutherland discuss at length how print cannot be taken as a single monolithic reality, but that it should be differentiated into many typologies, each of which should be examined in its own account. In particular they discuss the shift in the medium of newspapers and news more generally, and also scholarly editions. In their attempt to evaluate the new medium critically they reassess the so-called limit of print (that is the limitation in space and information it can offer within a single publication) and argue that this is a virtue rather than a defect: 'among the advantages of limited access to textual variance may be an emphasis on evaluation over qualification, selective insight over information' (p. 71). They argue that if electronic editions have yet to demonstrate that they are worthwhile 'beyond the circle of expert who assemble them' (p. 69), one of the reasons is

[141] In 2002 over 60 per cent of respondents declared they used journals in 'print mostly' and over 30 per cent in 'print only'; in 2011 over 45 per cent declared they used 'electronic journals mostly' and about 35 per cent 'equally print and electronic', while 'print only' dropped to 0 per cent and 'print mostly' to around 15 per cent. Similar results are shown for the use of facsimiles, with 'electronic mostly' rising from below 5 per cent to over 40 per cent and 'electronic and print' from around 5 per cent to over 20 per cent.

that they offer too much unselected, unordered, overwhelming information. They refuse here to consider positively one of the alleged biggest advantages of digital publication, that is the almost unlimited availability of space, as, in their opinion, this openness is likely to mean an anarchic aggregation of material of undetermined quality, rather than new opportunities open for scholarship.

The varying quality of the digital resources is also seen as a detrimental factor, and Deegan and Sutherland echo Shilllingsburg's remark that 'texts on screens look remarkably alike, despite profound differences in quality' (2006, p. 87). Shillingsburg also questions the longevity of digital editions and wonders if 'the fact that an electronic text is searchable compensates for the fact that we cannot guarantee its continued existence ten or twenty years from now' (p. 29). Of all the arguments presented so far for the lack of success of digital editions among scholars (that is the provision of too much uncontrolled information, and editions of doubtful quality and longevity), the latter seems the most convincing and connects with some of the reasons seen above for offering dual and hybrid publications. Printed books provide a certain embodiment of the text and scholarship, a format that can be reliably quoted and retrieved. On the other hand digital resources seem to disappear overnight; many are just relocated, rather than deleted, yet the lack of stability of digital editions is perhaps one of the main reasons to distrust them.

The provision of a print edition alongside a digital edition may help to stabilize the scholarship, subtracting it from the elusive materiality offered by the digital artefacts. However, to do so requires some form of hybridization of the two formats. The case of *Fine Rolls Henry III* project shows some of these hybridizations, with traditional footnotes becoming version 'after-entry' notes in the printed, on the model of the web version (Ciula and Lopez 2009, pp. 135-6). In this vision the new medium 'does not supersede the old one, like the wireless did not replace the concert hall and the television did not replace the cinema, but joins the old medium in an often positive and invigorating interaction' (p. 131).

However, this 'media complementarity' as defined by Wulf Lucius (2008, p. 131) is seen as limiting by scholars within the digital humanities. In particular Patrick Sahle (2008) considers 'printable' editions as not truly digital; whether he admits any validity in the possibility of printing a scaled down version, it is not clear. For sure, though, the hybridisation of the two media is a double-edged sword. On the one hand, it provides a respectable, material support, a 'regulatory authority' (Deegan and Sutherland, 2009, p. 72) to the instability of the digital edition, but on the other hand it anchors the new medium to the possibility offered by the old, without allowing for more imaginative approaches. Tara Andrews recalls how her PhD thesis, which was produced digitally and fully encoded in TEI,

had to be converted and scaled down as her University regulations required her to produce a printed volume (Andrews 2013, pp. 69-70); she then concludes that the requirement of producing citable-by-page printable versions of our digital editions will seriously undermine the possibility of developing digital editions that take advantage of the full computational potentials offered by computers.

Deegan and Sutherland (2009) see in the hybrid edition a convenient format at least for the medium term; Sahle (2008) and Andrews (2013) advocate for editors to abandon the page paradigm and embrace the full potential of representation and computation offered by computers, and to produce truly digital editions that will be able to convince their readers that they are worthwhile. Both positions seem reasonable and pursuable, and they may not necessarily contradict each other. In embracing the digital, some editors will pursue a safer approach by providing print and digital version of their endeavours, while others will experiment with the medium, pursuing innovative publication formats. This has been called the age of the digital incunabula, whereby the coexistence of new and old, of innovative and traditional is to be expected. It is not yet clear which format digital editions will take in ten years, nor in which way we will use them, hence the call and the need for experimentation in many directions. Nevertheless, what does seem clear is that publishers will still play an important role in the production of digital scholarly editions. It will therefore be crucial in the years to come for scholarly editors to accept the challenge offered by the digital environment in order to be able to negotiate their needs with the publishing houses, otherwise the risk is that once again publication models will be developed in response to the needs of the production chain rather than to the newly discovered possibilities offered by the digital medium for the representation and analysis of texts.

6.4 How Much Information? Knowledge Sites and Open Source Publications

The question of what should be provided as part of the digital publication is also part of the discussion around digital editing. According to Peter Shillingsburg (2006, pp. 96-102), thanks to the expanded storage capabilities of the web, the unit of publication no longer needs to be the edition of a work, but it may, and should, take the shape of a knowledge site which should be modular and open-ended. It should include textual material, which should constitute the core of the site, plus contextual and analytical material, as well as offering the possibility of users to contribute to it. Print editions can also be accommodated within this framework, but only as offshoots, 'targeted to specific audiences or for specific uses such as reading or teaching as opposed to prolonged and detailed study' (p. 97); such prolonged and

detailed study should then, in his opinion, be conducted digitally. He argues that as all the stages of work in textual scholarship are digitized, it sounds like a regression to have to select only a small portion of the material produced during the work to produce a limited printed edition, and that it would be much more useful to publish everything - all the transcriptions, collations and annotations. A digital edition should then be extensive, not selective (a vision contrasted by Deegan and Sutherland, 2009, as seen above).

However, in spite of the extensiveness of the knowledge site conceived by Shillingsburg, there are some types of by-products of the editorial work that have not been included in his outline, and these are the editor's private notes (reading notes, sketches, outlines, for instance), the source files (XML, or whatever file format has been used to encode the texts) and the scripts to produce the editions. Nevertheless these materials have been considered integral parts of the editions by several scholars. In a contribution of 2009 Bodard and Garcés presented their rationale for what they have called an 'open source critical edition'. Borrowing the concept of open source from the world of software development, they argue that an edition can only be called scholarly if it provides its dataset, which will allow other scholars to reproduce the editorial work and to reach, possibly, different conclusions to the ones of the first editor. They maintain that scholarly method is based on the transparency of the methodology and the open availability of all the information accessible to the scholar who first worked on a particular argument. If an editorial endeavour claims to be 'scholarly', then it has to comply with the rules of transparency and reproducibility. The analogy with the concept of open source allows them to establish that open does not mean enabling plagiarism, quite the contrary. Software licensed under an open source agreement is covered by copyright, with some rights reserved (for instance, the right to reuse the source as it is without modifications); the provenance of the software has to be acknowledged under all circumstances. Contrary to common belief, open source does not mean for free either: as a matter of fact many tools released under an open source licence are available for a fee. Open source simply means that a tool is distributed together with the code that has generated it: the source is open to scrutiny (and sometimes modification) to all those who already have the tool.

Applied to digital editions, open source means that editors must distribute their source files together with their public interface; therefore for an XML-based edition, for instance, the availability of the HTML output is not enough and the XML-encoded files, together with encoding guidelines, schemas and other validation tools have to be available to the users. James Cummings also argues for the availability of the source files, but for different reasons. He maintains that TEI-based scholarship can only progress if scholars are

willing to expose their sources, a practice that it is more necessary than ever, in spite of the resistance manifested by many, with 'excuses' ranging 'from licensing concerns to delicate academic egos (for example the reluctance to show XML with so-called "tag abuse")' (Cummings, 2009, p. 316).

The inclusion of source files is an essential but not sufficient condition for Peter Boot. In his view the purpose of digital editions is 'to make the content of works available as potential nodes in a larger digital network that will include not just the sources but also the tools, the output and the intermediate product of scholarship' (2009, p. 203). In particular, he focuses on annotations and calls a body of annotations *mesotext*: '"Mesotext" because it is text that can be located somewhere in between the primary texts of scholarship (the sources that scholarship is based on), and the secondary texts, the books and articles that it produces.' Annotation is identified as one of the scholarly primitives by John Unsworth (2000) and as a fundamental component of a digital edition by Peter Robinson (2003), therefore it is claimed by Boot to be part of one's dissemination package. Mesotext is composed of three types of annotations: (1) reading notes, (2) a body of supporting evidence for a scholarly argument, and (3) a collection of observations. In his vision, however, annotations are organized and structured in a hypertextual architecture (p. 207: 'mesotext is an organized collection of such annotations'),[142] being much closer to Shillingsburg's Contextual Data (2006, p. 101), which are described as 'things that went without saying'.

Connected to but somewhat extending the idea of the open source edition is the claim that the scripts and the software developed to produce the edition are part of the edition as well. In some cases, and in particular when the edition is produced from a paradigmatic TEI encoding, part of the scholarship and of the understanding of the meaning of the encoding is stored within the scripts (usually XSLT files) that are used to produce the output (Pierazzo, 2011, pp. 474-75). Scripts, far from being simply technical components of the website, store scholarly information about the handling of features such as corrections of scribal errors, preservation of original spellings, layout and paginations and therefore can be rightfully included as an essential component of the deliverable of an edition.

From the discussion it follows that what constitutes an ideal digital edition is every single byte that was produced in connection to the production of the edition itself. Is this a sustainable model? Are users really interested in all the aforementioned material, or will they

[142] To support such an organization Boot suggests the use of various tools, in particular he mentions *Pliny*, developed by John Bradley, for which cf. Bradley (2008).

feel overwhelmed, as maintained by Deegan and Sutherland (2009)? An analysis of digital editions from the perspective of the user constitutes the core of the next chapter.

7. Using Digital Scholarly Editions

Recent contributions have drawn digital editors' attention to the end users of their endeavours and the ways they use or might use them. In particular, a recurrent question seems to be 'are digital editions texts to be read or objects to be used?' When examining the rationale of the dual publication of *Henry III Fine Rolls Project*, Ciula and Lopez seem to imply that both activities characterize the genre: 'For a scholar interested in the historical record, the reading of the Fine Rolls edition as the seeking of information related to it is a comprehensive process that does not stop when the book is closed or the browser is shut' (Ciula and Lopez 2009, p. 133). Tim McLoughlin (2010) points out how in the electronic medium we have became users who no longer read texts but 'browse', 'search' and 'navigate' digital 'objects' (p. 40). In reply, Rehbein (2010) insists that searching and browsing are not exclusive to digital devices but that they characterize our use of certain kinds of printed books; he gives the example of so-called access resources, such as dictionaries (pp. 63-64).

However, despite this body of opinion, very few studies have been undertaken to test whether our intuition about the use of digital and analogue objects is actually supported by evidence. Even more crucial is the fact that there seems to be only a vague notion of who we are preparing our editions for: are they editors, scholars, students, or the general public? And what are the requirements of these groups of users? Are they the same or does each of them approach the edition with different purposes?

7.1 Readers and Readings: the Audience of Scholarly Editions

A tautological definition of readers is 'people who read', however the simplicity of this definition is still able to help us focus on the crucial activity of reading, which is neither simple nor well understood, as indicated by the discussion between McLoughlin and Rehbein (2010). The former, in fact, complains about the switch of focus in digital between reading to browsing and searching, while the latter claims that these activities are forms of reading too. Rehbein is of course right: reading takes different forms according to the object of reading, the purpose of reading and the medium from which the reading is performed. Historians of reading often point out that 'far from being "natural", reading is a complex skill' and is 'characterised by the evolution of the document from a linear and uniform flow of text to a tabular organization' (Vandendorpe 2008, p. 203).

Borrowing Heyer's proposal, Vandendorpe suggests we differentiate modes of reading using as a metaphor the way our ancestors gathered food: grazing, browsing and

hunting. The first represents continuous reading, where the 'reader picks up everything'; the second – browsing – is 'scanning of large body of information with no particular target in mind', an activity which is also called 'skimming'; with the third – hunting – as seeking for a specific information (2008, p. 205). The latter form of reading is also called 'scanning'. Continuous reading mostly applies to prose and in particular to novels or essays, skimming to newspapers and scanning to dictionaries, amongst others. The last two modalities are also fairly common digital types of reading activities: they are not exclusive nor have been born with the web, as rightly pointed out by Rehbein, but it is clear that the use of the web has emphasized them to the point they may seem to be born digital.

However comprehensive, the list of types of readings given above is incomplete. In addition we could include the so-called 'close reading', that is the type of attentive reading we reserve for texts we wish to memorize and learn from. In fact, close reading is not reading, it is re-reading as in order to learn or memorize some content we have to read it a number of times; a good synonym could be 'study'. Continuous reading is normally linear: in order to grasp their meaning, their message, novels and essays are normally intended to be read from the first to the last word, following the order intended by their authors or editors. This may or may not be the type of reading that is appropriate for other literary genres, though: poetry, for example, does not need to be linear and readers are encouraged to linger over a verse, a sentence, a word at their leisure. Siemens *et al*. (2012b, pp. 167-8) distinguish two types of reading: '1) the type in which the economization of comprehension is the aim; and 2) the type in which engagement with the aesthetic artifact is the aim', where the former can encompasses Vandendorpe's browsing and hunting, and the latter the grazing plus the close reading.

Reading is a complex matter; it is at the core of many intellectual activities and is performed in many different ways. When we think of readers as targets for digital editions it is not clear, though, which readers we intend and, more to the point, which kind of reading we have in mind. Even print scholarly editions can hardly be thought to serve continuous reading in the way a paperback does. Edward Vanhoutte (2009) reports on how the issues of his own scholarly edition of *De Leeuw van Vlaenderen* was overshadowed by the fact that a very popular newspaper distributed for free 500,000 copies of a modernized version of the same book. From this account, he draws a few lessons and in particular how 'buyers of literature are not interested in a reliable text, but in any easily acquirable text', and that 'the audience for scholarly edition is small, specialized and mainly undefined' (pp. 101-102). However, the problem may have been not with the genre of publication in itself (as

understood by Vanhoutte), but in the misunderstanding of the type of reading that a scholarly edition supports with respect to other types of edition.

Let us make clear that here and elsewhere by scholarly edition we intend not only an edition that contains a text established by a scholar: to be called scholarly, in fact, it must include the supporting arguments for the establishment or explanation of a text, it must contain an account of the scholarly work, presented in a format that fulfils the expectation of the scholarly community to which the editor belongs. The definition of scholarly allows for an internal variation: critical editions, diplomatic and other types of documentary editions, as well as commented editions: all can be defined scholarly as long as they provide the scholarly rationale that supports the establishment or the interpretation of the text. The edition of *De Leewe van Vlaenderern* edited by Vanhoutte fulfils these requirements precisely since, besides the 'critically established non-modernized and "clear" reading text', it also contains a 'contextualizing essay […]; a bibliographic description of the witnesses; a genetic essay; an explication of the editorial principles; a list of emendations in the reading text with references to the corresponding readings in the most important witnesses', and so on (Vanhoutte 2009, p. 101). Other examples include the critical editions of classical texts published in the Oxford Classical Text series, for instance, which respond to the expectations of the community of scholars by providing all the paratextual material in Latin. There are, of course, types of editions that fall in between scholarly and reading editions: reading editions that aspire to be considered scholarly by adding an introduction by an established scholar, or scholarly editions that 'hide' their scholarship at the back of the volume or in a second volume in order to present a clear text; such editions have in fact tried to solve the problem of the two audiences outlined by Lavagnino (2009), but their success is arguable. In fact, if one accepts that an edition can only be called scholarly if it presents the supporting argument for the establishment or interpretation of the text, then, by definition, such editions are by their nature typically unappealing to the general public because these additions, even if hidden away, will make the book itself too dense, too difficult to read, too heavy to carry, too expensive, and so on.

Scholarly editions are hardly books to be read continuously and linearly: there are far too many footnotes, internal references, variant readings, critical apparatuses, introductory materials, back matters, bibliography, specialized indexes, and so on to be able to favour a type of reading that we could call 'leisurely reading', namely the type of reading one likes to do on a beach or on lazy Sundays afternoons, or on public transport. For this type of reading (type 2 of Siemens *et al.*, 2012b), a page that only contains the main text and a light book that it is not weighed down by front and back matters are much more suitable. Or at least, this is

what the publishing industry has allowed their readers to believe; a point we will return to shortly.

'Reading' a scholarly edition is a different matter: scholars will aim at this genre of publication if they are interested in something more than the mere wording on the page: the commentary, the tradition, the variants, the history of publication, the receptions, censorship, the genesis and all the important elements that make a scholarly edition worth seeking. Contrary to what Vanhoutte claims, the audience for scholarly editions is clearly defined, and it is constituted by all those scholars for whom the textual is not the only concern, but who vest a great interest in the paratextual materials as well. This latter point is somewhat recognized by Vanhoutte, and in fact he declares that 'not all professional academics are interested in the full capabilities of the scholarly edition' (2009, p. 103). And frankly, why should they be? Why should a scholar of Nineteenth English Literature or History, say, be interested in seeking a critical edition of Homer? In some cases scholars of other disciplines may find it interesting to approach other disciplines' scholarship for personal interest, but this may safely (and perhaps unfortunately) be assumed to be rare. Some scholarly editors may interest themselves with, say, DNA studies, but only in very rare cases would they wish to read and appreciate a scholarly contribution on the topic, seeking articles on scientific journals. The fact that scholarly editors handle works of literature that potentially have vast readerships should not allow us to overlook that their scholarly editions are highly specialized publications, which more or less correspond to scholarly papers on scientific journals and not to hardback publications on DNA targeted at the general public. Scholarly editions as we intend them are research endeavours and their value resides in the advancement they have made to the scholarship of a particular work, and not in the number of copies they have sold. This is a crucial point: scholarly editors cannot fool themselves about the impact and the popularity of their outputs, no matter how much the new religion of 'impact' may push them in pursuing such goal.

The attempt to make scholarly editions read and readable by non-editors has characterized some of the discussions of the past few years. John Lavagnino expresses the wish for an enlarged readership quite strongly: in his contribution of 2009 (but reporting on debates and discussion of 1997) he maintains: 'we [editors] all want our readers to look not only at the texts we have established but also at our textual commentaries and notes; we do not want those parts to be treated as being just for other editors' (p. 65). This desire, however, is likely to be unrealistic, as more or less recognized by Lavagnino himself; nevertheless he maintains that 'a scholarly editor is still expected to serve a larger community that may not – and, at present, usually does not – take any great interest in the discipline of editing' (ibid.).

Statements of this type, conversely, combine two different issues: while it may be correct to assume that scholarly editors are called to serve a large community, scholarly editions do not necessarily have to do the same: quite the contrary. Scholarly editors might (and should) be involved in the production of simplified editions, for instance so-called reading or public editions, providing a reliable, quotable and readable text for a large public of readers, without the scholarly paratext such as commentaries, critical apparatus, indices, for instance; however, this does not imply that all editions must fulfil an outreach purpose, in particular it does not imply that scholarly editions have to do it. As a matter of fact, providing a reliable text of a particular work of literature, history or philosophy is not the only aim of a scholarly edition, but only one of them. It would be unthinkable to publish a scholarly edition with only the 'clean' text without providing any supporting evidence or argument on how the editor has crafted the text as none of the expert readers will accept the conclusions of the editor, the text would be considered unreliable and the edition will not be considered scholarly; in the same way a scientific argument would not be acceptable without supporting data or the results of experiments and similar evidence. However, the evidences, accompanying materials and the formalism via which scholarly editions have to be provided are the reasons why they are not sought by people looking for some leisurely reading. What is maintained here is that this is not a drawback of scholarly editions, the purpose of which is to make a scholarly argument, and on the other hand a reading or public edition fits the purpose of being accessible to a larger audience and there is no reason why scholarly editors should not undertake the production of such products. With no commercial pressure to sell more copies, as the internet can be a much more suitable environment, we can perhaps envisage a future where digital scholarly editions and public editions can live side by side, are produced by scholars who are able to understand and appreciate the different types of scholarship that is required to produce these two types of editorial products. Like anything else, different genres of publications are aimed at different readers and different types of reading. Sutherland asks how do we evaluate editions: 'by what criteria do we now measure a "good" edition or recognize the value of its contributions – to the life of the literary work, to the community of readers, to the wider culture into which it inserts itself?' (Sutherland 2013, p. 58). The answer is probably 'all of the above', but not necessarily all at once.

7.2 Editorial Formats and the Nature of Texts

It is important here to establish a principle which could be of crucial significance for the future of scholarly editing. At the moment there are two main types of editions of texts: scholarly editions (in all its forms and varieties such as critical, diplomatic, semi-diplomatic,

and so on) and non-scholarly editions (reading, modernized, abridged, and so on). While scholarly editions must be provided with a discussion about the rationale for establishing the text, commentary, apparatus, description of sources, and similar paratextual supporting evidences, editions that are proposed for the wider public have almost none; and while the scholarly edition represents in an almost iconic way the fact that 'the text is a historical artifact' (Sperberg-McQueen 2009, p. 28), and in particular via its most striking feature, the *apparatus criticus*, the absence of any discussion about the historical nature of the text in public editions fuels a dangerous confidence regarding the stability of texts, as they are presented in an oversimplified way and all problematic areas are smoothed by the regularity of the printed page. In fact, if we accept that the absence of scholarly paratex is what makes the reading edition palatable for the wider public, the same absence is also responsible for many of the misunderstanding about what texts really are. This publishing format in fact has contributed to the creation of false impression that texts are only made of letters, punctuation marks and white spaces, and that these components are an uncontroversial and stable part of the texts and always were. This belief is also very treacherous as it denigrates the role of editors: if texts are uncontroversial, editors and their work are not that significant and possibly pointless or even dangerous, as they may disturb what is otherwise perceived as fixed. In fact, Fraistat and Flanders lament how 'the role of scholarly editor is too often assumed to be best performed when least visible' and then advocate for textual scholarship to be considered as constitutive of textual culture, and not mere curatorial (2013, p. 1). This kind of perception is also at the heart of fundamental misunderstandings over the nature of texts, represented by theories such as the OHCO model, for which text can be traced to one and only one hierarchy of structures,[143] or such as the conviction that, because plain text files does not contain markup, therefore they do not contain interpretation.[144] In this misunderstanding about the nature of texts, scholarly editors have a strong responsibility, as they have on the one hand made their scholarship more or less incommunicable,[145] and on the other hand they have allowed it to be relegated to the bottom of the page or, worse and more frequently, to the end of books. Obviously, publishers share this responsibility. But while publishers are commercial enterprises with profit as a target, scholarly editors are not, and, as argued in the previous section, they have the responsibility of serving a larger community beyond their own through the production of more accessible forms of their texts. In a sense while it seems delusional to expect a large readership for scholarly editions, a commitment toward making editions that are more accessible and that present textual variation as part of

[143] For which cf. Chapter 2.

[144] Cf. Chapter 4.

[145] Cf. for instance Shillingsburg (1996), in particular Chapter 10, and Lavagnino (2009, pp. 66-67).

the normal nature of texts seems a more achievable and serviceable goal. It is in fact possible to conceive a richer version of reading editions, which may be even called 'outreach editions', the goal of which will be not only to support reading, but also to make accessible some aspects of the history and transmission of the text. Natural sciences have developed quite a rich and engaging set of methods for the dissemination of their findings, and while it is possible to find publications, games and toys that deal with scientific content, very little of the kind can be found for texts and textuality. Could we conceive of the elaboration of a new form of publication targeted at the general public, or subset of it, that presents in a pleasant, easy and engaging way aspects of the history of the text alongside the text itself? Could we learn something from science communication?

The proposal for an outreach edition does not fulfil though all the possibilities of making textual scholarship more accessible. It has to be attempted wisely; the risk is that to make scholarly editions palatable for the larger public will require changing their nature to a great extent, depriving them of their defining function, namely to provide the scholarly community with a reliable text and with all the evidence and supporting material required to verify the textual reconstruction. Therefore, instead of denaturizing existing scholarly editions, the involvement of textual scholars in the provision of new editions for the wider public that are able to support leisure reading as well as to present some aspect of textual variance could be a much more culturally meaningful and responsible approach.

7.3 Readers and Users of Digital Scholarly Editions

With the advent of the digital medium, a number of scholars thought that the division of readership (specialist vs. general public) that has been so built diligently by publishing habits might have been reconciled in the new environment.[146] This belief does not seem to have been supported by evidence. The digital environment seems to be the ideal space for providing editorial products which appeal to the wider public and even to make scholarly editions that are more accessible: but these are two distinct activities, two distinct goals that deserve two (at least) distinct types of digital editions. On the other hand some encouraging data coming from the *Walt Whitman Archive* (30,000 visitors a month, coming from 'every inhabited continent and at all age levels') demonstrates that sometimes people actually use web-based scholarly editions and find it interesting to engage with a scholarly community (Price, 2013). This might be due to the role of intellectual 'Father of the Nation' attributed to

[146] Cf. for instance, Sperberg-McQueen: '[E]lectronic editions can serve multiple audiences much more conveniently than print editions can' (2009, p. 35).

Whitman, to the variety and richness of the content of the website, or because the *Whitman Archive* is a project of long duration. The debate is open here, really, and future years will show if the provision of freely accessible digital scholarly editions may actually contribute to the enlargement of the readership of such a specialized genre of publication, in spite of all the drawbacks that have been argued here.

Reading works of literature may have different purposes: one can read to enjoy the content, or for the purpose of studying some particular feature of the text; some may also be interested in the transmission history of the text. All these types of reading (and re-reading) may be supported by books, but, as we said before, maybe not by the same books. In the same way, a digital environment can support (and enhance) all of these types of readings, but not necessarily within the same edition nor even within the same digital environment. In the previous chapter different types of digital environments have been analyzed, and figures about the market penetration of eBooks have been discussed. From such figures it seems clear that the fear or distaste of reading in digital format has been largely overcome by the introduction of tablet computers, which seem to provide a pleasant enough environment for digital reading to take off at a dramatic speed, as witnessed since 2010. Nonetheless the fact that people read eBooks does not make them more likely to read digital scholarly editions, as the data presented by Porter (2013) over the acceptance of digital editions by the community of Medievalists demonstrates. The lack of stability of digital editions is likely to be a strong contributing factor to such reluctance, Another cause may be that existing digital editions fail to support the type of reading that characterizes scholarly editions, namely re-reading, studying and annotating, presenting evidences, supporting editorial arguments. What is claimed here is that digital editors are not addressing the needs of the readers/users of printed scholarly editions, nor are they really considering the needs of a wider public, failing to recognise their real target audience and their reading requirements.

But what are these requirements? An editor approaching a scholarly edition will look for documentation about criteria for transcription and edition of the text; the reconstruction of the tradition and the relationship between the witnesses as well as a way to verify the hypothesis expressed by the edition; easy access to the description of witnesses, with detailed codicological and palaeographical analysis, and possibly the reproduction of the witnesses as well; a way to compare variant readings and verify the claims made by the editors; finally they find a way to highlight, annotate and relate different parts of the edition, with the annotations remaining accessible at any subsequent access. All these activities – the account

of which is necessarily incomplete as almost no specific studies of the matter exist[147] – amount to a very special type of reading, which we could perhaps call 'editorial reading' or 'editorial analyzis'. Non-editor scholars, on the other hand, will probably mainly be interested in the text itself to which they will apply any possible form of reading, from scanning, to skimming, to close to distant, according to their interest in the work. If the edition provides an accurate codicological, palaeographical or linguistic analysis, then scholars of the respective disciplines may also be interested in that particular section of the edition. Most of them will not have any particular interest in the transmission of the text: even if editors think that scholars in other disciplines should be interested in text transmission because texts cannot be understood without an in-depth knowledge of the transmission itself, nevertheless it is rare that this happens in practice. For many non-editor scholars it seems that the accurate reconstruction of the tradition and account of the textual variation as well as of the editorial criteria followed by the editor may be reassuring (if it is there it means that the editor has done his/her work properly), but it is substantially an immaterial provision of information, due to the complexities and imprecision of most critical apparatuses (Vanhoutte 2009, p. 100; Shillingsburg 1996, pp. 118-19).

However the lack of uptake of digital scholarly editions might also be affected by the fact that we may have chosen the wrong type of digital. While tablets computer are taking over as reading supports, computer screens do not make pleasant reading interfaces, as demonstrated by several tests,[148] therefore providing our scholarly editions as websites may not be the best strategy if the goal is to encourage a larger readership. However the eBook revolution seems to have only superficially touched on scholarly editions, if at all. This is mostly due to the fact that eBooks seem able to support what we have defined as 'leisurely reading' well enough, while they are less appropriate for other uses.[149] But eBooks are not everything that tablets can offer: apps look much more promising. The fact that digital editions may become more acceptable in a format that is potentially more engaging and user-friendly with respect to web seems confirmed by the positive impact made by non-scholarly editions like the ones produced by TouchPress, for instance. These editions provide a good example of the 'outreach edition' defined above: they provide a reading text, which is

[147] An exception to this is Siemens et al. (2012b), which presents some preliminary findings of the INKE project.

[148] Cf., among others Nielsen (2009 and 2010); the same study suggests, however, that the screen of a tablet computer is a much more pleasant reading environment.

[149] A pilot project conducted in 2010 within six US universities over the exclusive use of tablets as reading support for some courses had mixed results. Cf. the results as reported by Princeton University at <http://www.princeton.edu/main/news/archive/S26/64/38E35/>.

performed by various celebrities (mostly actors), and also primary sources, like manuscripts and first editions, as well as commentaries. The target of these publications is not editors, and these are not scholarly editions, however they certainly have a high appeal among people passionate about literature for professional or personal reasons. But what about editors? Could a critical apparatus be made more engaging if offered in an interactive way in an environment that has some sort of ludic aura around it, or will editors think that it is does not look scholarly enough? This is indeed a risk: if the goal of a scholarly edition is to make a scholarly argument for the benefit of other scholars, then before engaging with the new environment some serious analysis of the acceptability of the new environment should be undertaken.[150]

7.4 Looking for New Models: the Critical Apparatus

The analysis presented above is not completely empirical, based as it is on personal, biased observation. In fact there is a distinctive lack of studies on the way scholars use scholarly editions and in the way intelligent website design may support their type of reading. Even specific studies on interface design for cultural heritage material (such as the otherwise very good one produced by Ruecker *et al*., 2011 which will be analysed later on in this chapter) tend not to consider the needs of the readers, but rather the potentials of the data first.

The lack of systematic, quantitative data on scholarly habits therefore makes the elaboration of an intelligent design of a useful, reader-friendly digital scholarly edition very complicated. Editorial habits and expectations cannot simply be transferred from print to the web, as the contribution of Rehbein (2010) regarding the principles of digital thinking demonstrates. The same concerns over non transferability of habits from print to digital have been expressed by Sperberg-McQueen when he complains about the fact that 'some parts of current theory are tainted by silent (and perhaps unconscious) assumptions based on the immutability of print' (2009, p. 35). However, in order to produce innovative interfaces, it is necessary to separate the research needs from the medium that has supported them, and this may prove to be particularly difficult for people educated within the sole dominance of a particular medium. This dependency on the page paradigm (Sahle, 2008) is made particularly evident by studying what can be called the defining feature of the critical edition, namely the

[150] This is the purpose of one of the research stream promoted by the European project DiXiT (Digital Scholarly Edition Initial Training Network); this particular aspect of the project, the results of which are expected by 2017, will be conducted by a collaboration of the University of Antwerp and King's College London. Cf. the website project website: <http://dixit.uni-koeln.de/home.html>.

critical apparatus et variorum or *apparatus criticus*, as demonstrated by the 'outstanding, and in some respects alarming' example (Sperberg-McQueen, 2009, n.17) offered by Tanselle (1972). In this contribution the content and the principles of building a *critical apparatus* are all expressed in terms of space and layout – what goes at the foot of the page and what at the end of the volume, for instance – more than the editorial principles according to which it is a good idea to provide some information. Shillingsburg pushes the question even further, by complaining about how budget restrictions have forced editors to shrink the provision of variants, and by claiming that the critical apparatus 'was designed to show the wrong thing' (1996, p. 118). In his opinion, in fact, the provision of single undifferentiated lists of variants has caused readers to vacillate 'between wanting […] to know the variant versions of the work […] and, on the other hand to forget the whole thing as more trouble than the effort of deciphering and sifting an apparatus was worth' (Ibid.). Lavagnino (2009) picks up on the same topic, complaining over the same issues, and wonders if there is the possibility of elaborating another model for presenting the data that goes beyond the bare access; he then proposes some ideas of how this model should look. In particular he calls for the provision of a narrative able to explain and make some sense of the data offered by the facts, *de facto* calling for an integration of textual scholarship and textual criticism. He also mentions some interesting models of print publications which have tried to present textual variation in a more interesting way, but he then admits that these have been embodied by massive editions, resulting therefore in models that cannot be easily replicated. For instance, in the case of Coleridge's poems edited by Mays in 2001, the attempt to present a reading text alongside an *editio variorum* has inflated the publication up to six volumes 'of what a conventional presentation could easily fit into one' (Lavagnino, 2009, p. 74); he then admits that 'this edition is not simple to use'.

The critical edition has been shaped by print, and the *apparatus criticus* as we know it responds to the scholarly circumstances of the time , that is to the need to condense the largest amount of information into the smallest space possible and for an audience which was made of an handful of erudite scholars. The format of the *apparatus criticus* has not changed in more than a century: in fact one has only to examine Lachmann's edition of *De Rerum Natura* of 1850 to discover that its apparatus looks more or less exactly like the one of any modern edition. Yet, both of the aforementioned circumstances have now changed considerably: firstly the digital environment does not present the same space constraint as the print edition, so while the need to provide exhaustive documentation of the editorial process remains (actually, it is enhanced by the accountability made affordable by the digital medium), the need to condense into the most economic form does not. The audience has also changed, as the number of potential readers has substantially – or perhaps only potentially, as

Porter (2013) seems to imply – increased over the past century, with the increase of higher education establishments.

An interesting attempt to overcome the limits of the printed apparatus is represented by the dynamic apparatus that features in some digital editions produced in collaboration with the CTB (Centrum voor Teksteditie en Bronnenstudie), a research centre of the Koninklijke Academie voor Nederlandse Taal- en Letterkunde (Royal Academy of Dutch Language and Literature), which was coordinated by Edward Vanhoutte until 2013. The editions produced there all feature a critical apparatus that changes according to which witness the reader chooses to display, as well as presenting a more 'official' one which is displayed along with the edited text (see Vanhoutte 2007).[151] This approach allows for a real appreciation of the dynamicity and instability of texts, as well as showcasing the role of editors, and this solution demonstrates that innovative solutions can be indeed found when one breaks free from the constraints of the 'page paradigm'. Stemmatology and phylogenetics have also attempted to provide a different approach to the apparatus, with the provision of dynamics stemmas that afford access to variants from the graphic representation of the relationship among the witnesses. In all these cases we can see a shift from the critical apparatus as a repository of variants (the 'graveyard of variants', as it has been called) to a showcase of textual variation. Further work and experimentation in this direction could indeed lead to a new reconceptualization of the phenomenon of textual variation and its relationship to editorially established texts.

7.5 Interface Design: Principles

The hypothesis that a different digital medium might improve the impact of scholarly editions is indeed fascinating and clearly deserves further investigation, however there is already a lot that can be done to improve digital scholarly editions disseminated via the web. A second direction of improvement will have to take into consideration principles of interface design. As mentioned above, this needs to be done firstly by gathering some systematic data over the use of scholarly editions by scholars and students of various disciplines in order to have a better sense of which use cases a digital edition has to serve: the audience for scholarly editions has changed since the publication format was first established, their needs are also likely to have changed.

[151] See for instance de Elsschot (2007), and Daisen (2012); other editions are accessible form the CTB website: <http://ctb.kantl.be/>.

In their book of 2011, Ruecker *et al*. provided the first comprehensive overview of the growing field of interface design for the Humanities and Cultural Heritage in general. Interface design is a booming sector of web development, both at academic and commercial levels. However, it is quite a niche field within the Digital Humanities, and certainly a sector where editors may find themselves little at ease. On the one hand it is not hard to understand why this is the case: it is a completely different field of study which sits in the intersection of cognitive sciences, communication, marketing and computational sciences. On the other hand, however, the understanding of the way people interact with computers and the way research and reading is undertaken in a digital environment has to become a central concern when developing digital resources that are relevant to the right target. Nevertheless, it is actually arguable that an interest in interface design is really something alien to editorial concerns; Gabler (2008) produced an historical overview in order to investigate how book design has conveyed meaning to the readers and how the scholarly discourse has been spelled out at various moment of the history of the learned edition, showing the active involvement of editors in such endeavours and demonstrating how book design has influenced and shaped scholarship, and vice versa.

User-friendliness, while essential, is not enough: it is important as well to identify the audience we want to use the resource and what they are likely to do. However, to analyse users' expectations and requirements can be more difficult than expected; in fact, when developing something completely new, there might not yet be a body of users to enquire and to test it, and as a result it may be difficult to imagine the way something new may work or how it may affect the way we conceive a certain aspect of our digital behaviour (Ruecker *et al*., 2011, p. 11). This concept has been well expressed by Willard McCarty in a contribution of 1991: 'A new tool is not just a bigger lever and a more secure fulcrum, rather a new way of conceptualising the world' (§ 8).[152]

New resources build new affordances. Ruecker *et al*. (2011) devotes much space to the concept of 'affordance', which was formulated by James Gibson in the first half of the twentieth century and then further expanded in his book of 1979. In his definition, an affordance is 'an opportunity for action in the environment of a given perceiver' (Ruecker *et al*., 2011 p. 25); that is to say, what a perceiver (human) thinks a given object can do. For instance a knob affords twisting and perhaps pulling and pushing, a web link affords clicking and perhaps mouse-over. In design, affordance has been expanded to cover behaviours expected by users, for instance one expects pressing the return button of the keyboard means 'submit' or 'go' for a certain string query and not, for example, go back to the previous

[152] The passage is used for the epigraph by Ruecker *et al.* (2011), p. 1.

screen. Designing something new, then, will have to build on perceived affordances, while providing new ones. However, the provision of new affordances, may not be done knowingly; as pointed out by McCarty, new tools are sought (and developed) with the aim of simplifying and speeding up known tasks and activities but often have the consequence of changing the way we conceived the activity itself. One example of this is given, for example, by automatic collation. The scholarly community approached this activity initially only to relieve editors of a form of 'idiot work', to use Shillingsburg's colourful expression (1996, p. 139). However, the availability of accurate, comprehensive collation such as the one offered by computational methods has stimulated the development of methodologies such as phylogenetics and stemmatology, as well as opening new exciting perspectives and questioning existing heuristics, as maintained by Andrews (2013, p. 73). The availability of large datasets (the so-called 'big data') also suggests the existence of new, unthought-of affordances.

One of the tendencies that have dominated the development of user interfaces is skeuomorphism, that is, the attempt to digitally imitate analogue, real-life objects, which is a design approach that has particularly but not exclusively characterized Apple's interfaces up to 2013.[153] For example, one can think of the 'recycle bin' or 'trash' icon in most operating systems which is shaped as an office paper bin, or the fact that the home screen of a computer is called a 'desktop'; as such skeuomorphism is deployed at both visual level (the appearance of the trash bin), and at linguistic-symbolic level (the screen called a desktop). Skeuomorphism is not born with the digital, though, as clearly demonstrated by electric lights fittings shaped as candles or vinyl and laminate flooring that imitates wood. In the pre-digital history of the book the most famous example of skeuomorphism are incunabula which, in order to become an acceptable alternative to manuscripts to a highly educated and snobbish class of intellectuals, were designed to mimic as closely as possible the characteristics and features of the best handcrafted codices, and indeed it is not uncommon to find hand-illuminated early printed books.

The eBook market is flooded with examples of skeuomorphism: from the name itself (eBook), to the act of turning the pages in touchscreen devices, to virtual bookshelves, to the possibility of having yellowed paper-like backgrounds, you name it. This situation presents striking similarities with respect to the case of incunabula mentioned above: in order to make digital 'books' acceptable to the users, they have to look and behave as much as possible like printed books. Skeuomorphism characterizes also digital scholarly editions, sometimes in more subtle ways than, say, yellowing backgrounds, but nevertheless the guiding principle of

[153] On skeumorphism and why Apple has decided to abandon it, see Judah (2013).

interface design for digital editions has been to try to reproduce the printed scholarly edition as much as possible, in order to give a familiar, comforting feeling. The article of Tim McLoughlin (2010) demonstrates very clearly that this preoccupation is not without cause: the disorientation of the scholar in front of XML markup who simply wants to use a square bracket to signify that a date has been supplied by the editor hides, but not by much, the desire of reproducing the page paradigm in the new environment which has to look as much as possible like the old one. The success of word processing software (such as Microsoft Word, for instance) can be traced back to the fact that they are built on a typewriter metaphor: their relative friendliness, conveyed by the familiar action of typing on a keyboard (itself modelled – it goes without saying – on that of the typewriter), accompanied by new affordances such as, for instance, the copy-and-paste facility, determined an early and unconditional acceptance of the on-screen composition of text. It looks like a typewriter, but it is better: what is not to like about it?

Is, then, skeuomorphism a good thing? Do we need to adopt skeuomorphic principles in order to produce digital scholarly editions that are acceptable to the scholarly community? Certainly skeuomorphic interfaces present great advantages in terms of intuitiveness and user-friendliness, but, again, the commercial world of software development may have some suggestions that are worth considering here. Designers are moving away from the concept, seeking a more minimalistic and simplified look for digital objects (Judah, 2013), claiming that once you do not have to waste your time in imitating analogue devices, new ideas, and innovative and more efficient solutions could be found (Thompson, 2012). This is also the story of the evolution from incunabula to early printed books to contemporary books: once the product was fully accepted by a booming market of literates, new features were introduced (the frontispiece, the alphabetical index, for instance), while other features which came from manuscripts were gradually abandoned as not useful or worth the effort (two column layout, illuminated initials, etc.). These examples and tendencies seem to suggest that in the not too distant future skeuomorphism in digital publishing will exhaust its function (that is, introducing the unknown and the new by using a familiar design), and completely new formats will be produced in order to be consumed by born-digital users. As discussed above, when it comes to digital editions, it does indeed seem that the page paradigm is doing more harm than good, where 'what we used to do' is dangerously shaping, even conceptually 'what is possible', so this evolution may be something to aim for.

Ruecker *et al*. provide some useful ideas on how to work toward the redefinition of useful interfaces for cultural heritage material. They present tips on how to integrate interface design within the research lifecycle from its very beginning, where the interface grows with

the research, shaped by the interaction between designers and scholars, and they call for essential activities like modelling – the importance of which has been already discussed and cannot be overestimated –, prototyping and iterative design (2011, pp. 6-12). They also stress two essentials principles: aesthetics and confidence. In order to be useful an interface has also to be pleasant, or beautiful, and has to be able to inspire trustworthiness. Building on the principles developed within commercial design (in particular by Patrick Jordan),[154] they outline a short history of digital design: in the first stage, the guiding principle was functionality, a second phase was dominated by usability; a third stage complements functionalities and usability with pleasantness. They then reflect on what makes a digital object pleasant, listing concepts like the 'aesthetic experience', 'confidence', 'trust', 'willingness' and 'satisfaction': users will find a digital object pleasant if it is beautiful, if it can be trusted (if results are accurate, behaviours are reassuring), if it stimulates interaction and if the predefined goals can be achieved (Ruecker *et al*., 2011, pp. 12-19).

It is worth stopping and considering this latter aspect from a different point of view. One of the purposes of scholarly editing is to establish a stable text of a given work, 'purifying' it from the debris (the noise) that textual transmission has accumulated around the original message. Stability and reliability are essential features that complement this task. However, digital scholarly editions are mutable objects by nature and their mutability goes against the very principle of trustworthiness and confidence mentioned above.

> Editors may justly feel that electronic editions have translated them from a stable environment with difficult but well-known problems into a river of Heraclitean flux, in which everything is changing from moment to moment, and the editor and edition are expected to adapt actively to those changes from moment to moment, without being able to rely on many of the principles which used to be stable guides to editorial thinking (Sperberg-McQueen, 2009, p. 30).

Sperberg-McQueen then concludes that in order to engage with mutable editions, editors need to deploy 'different way of thinking about editions and the choices they embody' (Ibid.): trustworthiness need to be established on new bases, by providing ways to handle such mutability and by providing interfaces that respond to the needs of their intended users. The former aspect will be discussed in the next chapter, here it is important to recognize how the creation of interfaces that are pleasant, user-friendly and targeted to the right user-base is therefore crucial in order to encourage the deployment the new way of thinking invoked by Sperberg-McQueen. Michael Whitelaw (2011) speaks about 'generous interfaces', that is

[154] See his contribution of 2000, cited by Ruecker *et al*, 2011, pp. 12-19.

interfaces that offer users many access points to the content to users before they ask for it, a concept which is not far from the 'rich-prospect interfaces' proposed by Ruecker *et al*. (2011). In both cases, the scholars campaign in favour of inviting, interactive interfaces, able not only to satisfy the established needs of users but also to anticipate and build new affordances, thereby becoming fundamental research tools. Siemens *et al*. (2012b) advocate for the extensive use of prototyping, experimenting with different solutions alongside, and even prior to, developing workable interfaces. If overcoming the page paradigm is a desirable outcome, experimentation and daring new solutions represent the most promising way to achieve this.

7.6 A Standard User Interface?

Experimenting with interfaces may be a very trendy research perspective, however scholars have pointed out the dangers of the proliferation of different approaches which risk confusing the scholarly body of intended users. Roberto Rosselli Del Turco (2011) maintains that the fact that each digital edition presents a different user interface is disorienting and ultimately off-putting for scholars who are used to printed books. Printed books, in fact, present a very stable and standardized interface: they have a cover, some introductory matter, the main body of the text, normally subdivided according to some internal and meaningful structure, indices and bibliographic references at the end. Within the page, there might be some words on the top margin, such as running titles for instance, the main body of the page and a footer which may include footnotes and the *apparatus criticus*. On the other the hand, websites are all different in the organisation of data, the way the menu entries are labelled, the way users access and query the data, and so on. For instance, Rosselli Del Turco seems to have a point when we compare two websites we have mentioned before: suppose a user is interested in understanding the editorial principles that have guided the text reconstruction or to understand the meaning of the symbols deployed. In the *Jane Austen's Fictional Manuscripts Digital Edition* the criteria of the edition are to be found in a section called 'The Edition', the third from the left, and then, within that, in a page called 'Editorial Principles'.[155] In *Henry III Fine Rolls Project* the same type of information can be found within a section called 'Style Book', the fourth from the top, and then within a page called 'Editorial Conventions'.[156] One of the reasons for such diversity is given by the fact that the two websites, while being both digital scholarly editions produced in the same research centre, they have two different

[155] http://www.janeausten.ac.uk/edition/editorial.html

[156] http://www.finerollshenry3.org.uk/content/book/edit_conv.html

scholarly communities of reference (Nineteenth Century English Literary Studies for the former case, and Medieval English History for the latter), with different conventions and habits, and different printed models from which they draw. But is the expectation that these editions work in the same way a reasonable one? Probably not. In fact, if they have to respond to the needs of their intended target of readers/users, this is not only an unreasonable expectation, but also a wrong one. Digital scholarly editions are a genre of publication, but they are also grounded in specific disciplines, each with their own tradition and special characteristics, connected with their scholarly practices and specific disciplinary purposes. Scholarly editing is a practice that takes place within many disciplinary fields, and in spite of the fact that editorial practices in, say, Byzantine studies are not very dissimilar from those of French sermon studies, they are diverse enough to make two editors argue over allegedly irreconcilable differences.[157]

Nevertheless, if user interfaces are gradually to abandon their connection with printed tradition in favour of more born-digital interfaces, as claimed before, it is likely that some of these differences may fade away. In spite of the disciplinary differences, editions in print fit, more or less, the same standard format, where the order of the preliminary or back sections may change, but they are still at the front and at the back, often distinguished by different page layout and numbering. Editions from French sermon studies and Byzantines studies will still look different, but such differences may be determined by the different purposes of the editions, rather than by some *a priori* differences of discipline.

In his contribution Rosselli Del Turco lists a series of requirements for a good digital edition interface, and this includes good hypertextual functionality, special character handling, image manipulation tools, advanced search functionality and integration of ancillary tools such as glossaries and concordances. This list of specific requirements complements another list of principles according to which user interfaces have to offer consistency, readability, recognition (that is familiarity), discoverability (a concept closely related to that of generous and rich prospect interfaces seen above), user control, easy and intuitive navigation, ergonomics and attention to human interaction, and scalability (Rosselli Del Turco, 2011, §§ 16-29). While the list mixes principles with tools, it clearly manifests a growing interest in producing scholarly guidance for interface design, as well as dissatisfaction with some of the digital editorial products currently available.

[157] This is the experience reported by Tara Andrews (2013) as well as the one experienced by the author of the present contribution.

Rosselli Del Turco concludes his contribution by casting a hopeful look toward the possibilities offered by touchscreen devices, wondering what their impact will be on the current conception of digital scholarly editions once they are implemented in such an environment. Tablet computers seem to have a lot of potential: the intuitiveness of the haptic interactivity is perceived as pleasant and playful, to the point that the focus group used by Nielsen (2010) rated the experience of reading on the iPad more satisfactory that reading a printed book. However, tablets are not immune from risk, in particular with respect to their sustainability, given the rate at which new versions of the operating systems of tablets are released (more or less every six months at the time of writing), any of these updates is potentially able to make applications unusable. We will discuss the question of sustainability in more depth in the following chapter; here it simply suffices to raise the issue.

7.7 Are User Interfaces Necessary?

From what has been described before, it seems clear that interface design is becoming a central part of digital editorial work. As digital editions are all different and do not follow the same publication structure, the design of meaningful interfaces becomes a fundamental vehicle for the delivery of the scholarly discourse. The way the text is displayed, the way the searching and browsing facilities work, how easy it to access certain functionality with respect to others, all of this contributes to the rationale of the edition. In a previous contribution, an argument has been made for the inclusions of scripts such as CSS and JavaScript within the components of digital scholarly editions (Pierazzo, 2011, p. 474). There the reader was invited to disable the styling scripts (CSS) of any one page of the manuscript edition of the digital edition of *Jane Austen's Fiction Manuscripts*, in order to appreciate how fundamental the interface has become. Figure 15 presents the XML source of the last two lines of page [1] of *The Watsons*; fig. 16 presents the same text as it is readable on the web, alongside the equivalent text with styling disabled. As one can see from locating the word 'Ball', that is the first word of the first of the two main lines of writing (red circle), not only is the display different, but also the order in which the words occur. For instance the sequence 'some merit in chearfully' (blue circle) is located in the bottom half of the XML specimen, but it is over 'Ball' on a second level interlinear insertion in the web version, and finally quite a few words before 'Ball' in the version with the CSS disabled. All versions are in a sense correct: in the XML the sequence is written quite a bit after 'Ball' because Austen wrote it after, and also after the first substitution; in the web version it is placed over 'Ball' because that is where it is placed in the manuscript, and in the version without the CSS is before 'Ball' because if you consider the lines of texts, from top to bottom the sequence

‘some merit in chearfully’ is found two lines before ‘Ball’. As they are all correct in one way or another, it is the interface that as a regulative principle provides the version considered more correct by the editor.

```
            the important morning of the <lb xml:id="pmwats-del4"/>
            <w rend="capital">Ball</w>; without being able to stay &amp; share
            the pleasure herself, because <lb xml:id="pmwats-del5"/> her <w
                rend="capital">Father</w> who was an <w rend="capital"
                >Invalid</w> could not be left to spend the <w rend="capital">
                <choice>
                    <abbr>Even<hi rend="sup">g</hi>.</abbr>
                    <expan>Evening</expan>
                </choice></w>
            <lb xml:id="pmwats-del6"/> alone </del>
        <add place="superlinear" type="rewriting" n="1"> who was very recently
            returned to her family from the care of an <w rend="capital"
                >Aunt</w> who had <lb corresp="#pmwats-del3"/> brought her up
            was to make her first public appearance in the neighbourhood <g
                ref="#dash2"/> &amp; <lb rend="interlinear_multi"/> her eldest
                <w rend="capital">Sister</w>, whose delight in a <w
                rend="capital">Ball</w> was not le<g ref="#ls"/>sened by a ten
            years <au:revision>
                <del rend="overwritten" extent="wPart">
                    <unclear reason="illeg">Engage</unclear>
                </del>
                <add place="superimposed">
                    <w rend="capital">Enjoy</w>
                </add>
            </au:revision>ment<note>'Enjoy' written over previous word, possibly
                'Engage'.</note>, had <lb corresp="#pmwats-del4"/>
            <au:revision>
                <del rend="overstrike">her eldest <w rend="capital">Sister</w>
                    who had kindly</del>
                <add place="superlinear" n="2" rend="+3 gl">some merit in
                    <distinct type="spelling">chearfully</distinct></add>
            </au:revision> undert<au:revision>
                <del rend="overwritten" extent="wPart">ook</del>
                <add place="superimposed">aking </add>
            </au:revision><note>'undertook' altered to 'undertaking' by writing
                'aking' over 'ook'.</note> to drive her &amp; all her finery in
                <lb corresp="#pmwats-del5"/> the old <w rend="capital">Chair</w>
            to D. on the important <choice>
                <abbr>morn<hi rend="sup">g</hi>.</abbr>
                <expan>morning</expan>
            </choice><g ref="#dash3"/>
            <lb corresp="#pmwats-del6"/>
        </add>
    </au:revision>
```

FIGURE7.1: XML SOURCE OF THE LAST TWO LINES OF *THE WATSONS*, P. [1]

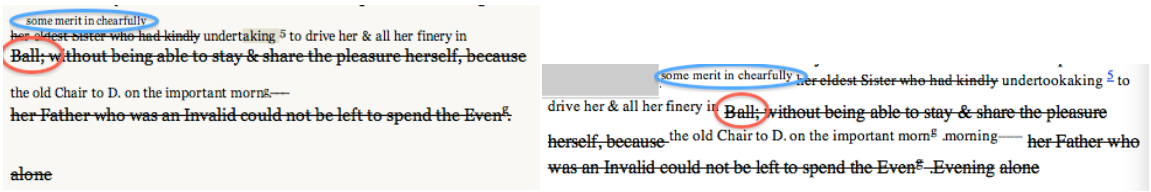


FIGURE7.2: WEB DISPLAY OF THE LAST TWO LINES OF *THE WATSON*, P. [1], WITH AND WITHOUT STYLING

This small example shows the importance of the provision of an interface: the text not only looks nice, but it is presented in the format that is considered optimal by the editor. Interfaces build meaning and include editorial decisions and principles.

However, the provision and the relevance of the interface have been questioned in many ways both at institutional levels and by the scholarly community. From 1995 to 2008 the Arts and Humanities Data Service functioned as a long-term deposit for digital artefacts produced in the United Kingdom by research projects in the Arts and Humanities. Its purpose was to guarantee the preservation and survival of valuable data, encouraging best practice in the production of such data and enabling their reuse for other research endeavours. However, interfaces and tools were not included in the deal: a project team who wanted to deposit with the AHDS would only be allowed to send their data, and not the scripts, interfaces, and structure that made them meaningful. To be fair, toward the end of its life the AHDS had started to explore ways to support some sort of interface, as it had been recognized how crucial the presentation of data had become to convey scholarship; however this desirable evolution never took place as the main funding body of the AHDS (namely the Arts and Humanities Research Council) took the not so forward-looking decision of terminating its support in 2008, preventing any further research on the matter.

When the website of the Archimedes Palimpsest was first launched in 2008, the decision not to provide the images and the transcribed text without a Graphic User Interface (GUI) made a sensation. The reasons behind this choice are expressed by the website: '[o]ur thinking behind this is that GUI's [sic] come and go, and our mission at this stage in the project is to provide a permanent document'. The project team realizes that the material is 'unwieldy to use' and that '[t]he digital product is therefore not so much a publication as an archive'. They finally 'invite and encourage any interested parties to apply whatever software they like to create further images based on our data set, further text based on our dataset, or graphic user interfaces through which the data that we have presented may be more easily managed.'[158] The idea that interfaces 'come and go' has probably stemmed by analogy from the typical division of source and output so common for publishing frameworks: with TeX and LaTeX, for instance, assuming one has installed a version of the required software, one can generate as many copies as wished from the same source files, and all such copies will all be identical. The use of markup language such as SGML and XML have reinforced this conception (also embraced by the TEI), as one of the biggest advantages of their adoption lies in the separation between data and presentation, where all the scholarship is allegedly embedded in the data and the presentation is just its pretty face. However, while for any text

[158] http://archimedespalimpsest.org/digital/

encoded in TeX a standard visualization can be produced, the flexibility and customizability of XML, for instance, requires the development of personalized interfaces able to showcase the markup. The example of the *Jane Austen's Fiction Manuscripts* seen above, demonstrates how crucial the function of the interface is as a way of embodying editorial statements and also the need to advocate for the inclusion and even recognize interfaces as part of the scholarly output.

But there are reasons why interfaces have even been seen as potentially dangerous. As pointed out by Peter Robinson, interfaces are very expensive, and require a lot of technical expertise and considerable infrastructural investments to be supported and maintained. By making them an essential part of the digital edition, one consequence is that digital editions can only be developed within well-funded projects and with the support of specialized research centres, excluding therefore more personal enterprises such as PhD dissertations (Robinson, 2010, p. 158). On the other hand, if editions are provided with minimal or no interface and rely on freely available web services and browsers' capability to suggest connections with similar objects, then access to the production of digital scholarly editions can be made much easier, a fact that will certainly increase their diffusion and their use. Robinson radicalized his position in a presentation delivered at the Social, Digital, Scholarly Editing conference in 2013, not only attacking the production of custom interfaces as dangerous, but also the workflow model that generates it. In his vision, editors should be provided with tools that they can control by themselves, without the need to work in research centres that develop ad-hoc standalone systems that are very expensive and will be excruciatingly difficult to maintain in the long term (Robinson, 2013a). The sustainability of digital scholarly editions will be discussed in detail in the next chapter, but here it is necessary to analyse the scholarly implications of this vision. In a successive clarification of his view, Robinson stated that what is wrong with the production of digital scholarly editions is the model of 'one scholar, one project, one digital humanist' ('1S/1P/1DH'), where each edition is produced in isolation by providing it with customized, non-sharable tools and infrastructures (Robinson, 2013b). He maintains that the '1S/1P/1DH model must die' to free editors from the fear that doing digital editions is too difficult, but he recognizes that there is still work to do in order to produce those tools that will make such disenfranchisement possible. And indeed work needs to be done, both in terms of creating a preliminary scholarly agreement (for instance of the type Andrews, 2013, discusses), and of understanding the types of needs to which digital scholarly editions aim to respond. In order to develop meaningful tools and provide editors with a strong infrastructure, more reflection on acceptable models of representation is required.

However, one has to wonder here if this model is appealing after all, and if a future where editors will be able to produce editions entirely on their own is necessarily something to aim for. Scholarly editions have always been the result of many different contributions, and in particular the product of the collaboration of editors and publishers, with the latter providing services such as quality control, copyediting, *mise en page*, and so on. The model that has prevailed in the past few years of digital scholarship is instead the one where publishers have been substituted by developers (or the digital humanists, to use Robinson's terminology). Yet, the proposal now is to go completely solo, without any external contributors apart, that is, from the tool producers, who should allegedly be at the service of the editor. According to this view the only meaningful relationship that the editors will build will be with their users, who will become central to the editorial discourse and who will have most of the responsibility of making sensible use of the given text by selecting the right tools with which to access, interrogate and understand the edition. As no custom interface will be provided, what to do with the material will be their choice. But are editors really willing to assume the whole workflow by themselves? Have they the preparation, competence and energy to accept the full weight of the workflow of producing an edition? And are users so willing to become editors and to build-up their version of the editions? And again, who are these users? As maintained before, study of the target audience of the scholarly edition and their needs is essential and urgent.

In the next chapter the discussion of sustainability of a work model, such as the one stigmatized by Robinson, is taken up, and its implications examined. In general we will investigate problems of longevity and preservation of digital resources and how these affect the confidence of the scholars who use them.

8. Trusting the Edition: Preservation and Reliability of Digital Editions

Scholars seem not to trust digital editions to any great extent. They probably use them online, but then they look to printed version to cite them. Various contributions (Robinson, 2005; Deegan and Sutherland, 2009; Vanhoutte, 2009, to name a few) have presented their explanations of why this is the case, some of which will be examined below. However, it seems most likely that one of the main reasons for this scepticism is that digital editions look much less stable and durable than printed editions.

Printed scholarly editions have long shelf lives. Some texts may be edited every hundred years, and sometimes at even longer intervals. Editing texts is a monumental task, in particular when there is a large number of surviving witnesses, and can take years, if not decades, for a scholar (or a group of scholars) to achieve it. The large amount of work required has somehow encouraged the view that editions may be forever, and that once texts have been edited well they do not need to be edited again.[159] According to Eggert, the work of the critical editor, although with differences, is comparable to the conservators of artefacts such as paintings or buildings, and its purpose is 'securing the past' (Eggert, 2009). Taking a similar line is Price: 'Editorial work is one way to engage in historical criticism and to help bring the past into the present so it may live in the future' (Price, 2009b, § 10).

Digital editions, on the other hand, are subject to rapid obsolescence and potentially to constant updates. The technology is evolving at a spinning rate, as is the taste and the expectations of the users, and this is contrary to the long shelf life view that the scholarly community holds of scholarly editions. The problem is not an isolated one: with more and more data being born digital and more and more data becoming unreachable, the durability of digital objects is a general issue that society is trying to deal with. New professions, new strategies, reliable data formats and standards, good practices, all of these are being deployed across the planet in order to try to tackle the problem of longevity and the preservation of digital assets. Nevertheless, in spite of these efforts, it seems clear that digital curation and preservation is a very costly activity and a continuous one too, to the point that one has to ask if it is indeed realistic to think of a digital product able to match the durability of print products, when the latter could potentially last for centuries, and require little effort once stored on a library shelf.

[159] For instance Deegan and Sutherland (2009, pp. 60-61) present a short collection of overconfident statements about the definitiveness of scholarly editions.

It is worth noting here, however, that the alleged durability of print products needs to be reassessed. When we say that books have lasted for thousands years, we tend to forget that only a very small portion of them have actually survived. The large majority of the codices produced in the Middle Age are lost, not to mention documents produced in the classical era, for which only fragments remains. For the early modern period, many editions of books have completely disappeared too, and many more are preserved by only a handful of exemplars, saved as they are by the large number in which they were produced. In this case we are reminded by analytic bibliographers that each copy of any given impression could be potentially different from any other, and so many important versions may have been lost even if some exemplars have survived. In the modern era, acidic paper has also lead to more losses, in addition to the 'natural' loss due to fires, other disruptive agents and disinterest. The point is that not all books have survived, but only the ones that people were interested in or were capable of preserving; some books have simply survived by chance. Every generation has exercised a selection, a choice of the items to be preserved, discarding the ones that have been judged of no interest for the future. Most textual scholars know the frustration of having to deal with glimpses of the stages of transmission of any given work and are aware of how educated guess-work must sometimes supply for deficiencies of traditions. In a sense, textual scholarship is the discipline that has been created in order to deal with the missing documents, reconstructing lost evidence and circumstances.

The other myth to be challenged is the idea that preserving printed books is not an expensive undertaking. Libraries are such established assets of our culture that we tend to take them for granted, and we remember that they are actually quite expensive services only when they are threatened. Indeed the current economic climate has lead to substantial cuts to library budgets and the closure of many of them, and this is because they are expensive. Deegan and Sutherland, quoting Gabriel Zaid, report that in '1989 the British Library estimated that, though the books it receives on legal deposit are free', receiving, cataloguing, storing and in general managing them 'cost fifty pound per copy. Plus one pound per copy a year' (2009, p. 160). Even more problematic is space: many libraries are collapsing under the volume of printed books being produced yearly. To face shortage of space, some libraries have started to store books off site (such as, for instance, the British Library), or have built larger, mostly underground, buildings (such as the Bibliotheque Nationale de France), others have digitized and then destroyed some of their belongings (a practice started as a consequence of microfilming: Tanselle, 1989b, p. 44). Preserving the past is no trivial matter, whatever the format of the artefact that needs preservation; it requires a policy and a vision, a will and a handsome amount of investment. Digital preservation is only one aspect of such an endeavour. It is not denied here that digital preservation is a problem – it is, and it is of the

most serious nature – however, it is important that we do not set our targets too high and that we are not building too high expectations over incorrect premises. It is very clear that we will not be able to preserve *everything*, no matter how hard we will try; it remains to be ascertained that preserving *everything* is actually what we want and a good idea.

8.1 Data and Metadata: Standards, Formats and Preservation

It is then unrealistic to believe that we will be able to preserve all digital objects that we produce: we are now producing every few days as much data as we have produced since the beginning of history, a phenomenon known as 'data deluge' so it is not realistic to think that we will be able to preserve every byte we generate. However, there are data and complex digital objects that are more special than others and that we might want to preserve; digital scholarly editions fit this latter category. But how can digital objects be preserved long-term? According to Deegan (2006) a set of good practices can take you a long way. The use of appropriate standards for both data and metadata, as well as separation of content from interface are suggested as key factors for the preservation of data. Deegan calls for the responsibilities of editors in the preservation for their editions: by making the right choices during the preparation of the edition phase, they will make their edition preservable.[160] The reason why the use of standards may improve the durability of one's edition lies in the fact that the more such standards get used, the more likely that there will be an interest in keeping it alive; and even if the body responsible for the maintenance of such standards will change them or cease to exist. It is assumed that, if there is a sufficient amount of data that complies to this standard, someone, somewhere will find a solution (or many solutions) as to how to migrate the data from one format to the other, as it will seem to be more economical to invest in such an activity than loosing the data or redoing everything from scratch. If a critical mass of data exists in one format, tools will be produced to migrate it. So choosing widely adopted standards is a good idea. However, it is also necessary to choose one's standard well, that is, to choose standards that are extensively documented not only from a technical point of view (how they work), but also from a semantic point of view (what they mean). The TEI *Guidelines*, for instance, represent such a standard; the same can be said for METS,

[160] Digital Preservation and Curation are quite lively disciplines which comprise a large number of publications. At the moment of writing, the most respected one are Harvey (2012) and Brown (2013). I would like to thank here Gareth Knight and Richard Gartner for pointing me, as always, in the right direction.

PREMIS, MODS, and Dublin Core.[161] While the TEI is at once a format for data and for metadata, all the other formats are for metadata only. These have a crucial role in order to ensure a longer and safer life for digital objects, as described by Lavoie and Gartner (2013):

> [Preservation metadata] facilitates the process of achieving the general goals of most digital preservation efforts: maintaining the availability, identity, persistence, renderability, understandability, and authenticity of digital objects over long periods of time […] preservation metadata establishes an informational frame of reference around a preserved digital object that remains attached to that object over time.

All the standards mentioned above are open, developed and maintained by public institutions or charitable organisations, a fact that in theory should be a further guarantee of durability. However, the recent 'shutdown' of the US government has seriously damaged this belief. On October 1, 2013, the United States federal government entered a shutdown after Congress failed to enact regular appropriations or a continuing resolution for the 2014 fiscal year. As a consequence of this, many non essential federal services were also closed, including the Library of Congress, which however, restored access to their website on the following 3rd of October. The consequence of the closure of the LoC's servers was that an undetermined but certainly large number of web services became suddenly unreachable, resulting in the inaccessibility of thousands if not tens of thousands websites worldwide. The reason why this happened is due to one of the characteristics of XML, the language of choice for most of the LoC standards. In order to declare the specific XML vocabulary one is actually adopting, XML files normally include at the very beginning a reference to an URI, a link to a site where the format is documented and maintained; this mechanism is called *namespace*. However, if the URI of the namespace is not reachable, all files that contain a link to such URI when recalled by a web user will return an error instead of the content of the file. This was what happened to all web services that made use of any of the LoC metadata standard. The disruption only lasted for three days, however this was enough to demonstrate the fragility of the allegedly safest of standards and the global interdependency of digital objects.

---

[161] The first three standards mentioned are developed and maintained by the Library of Congress: METS (Metadata Encoding and Transmission Standard): <http://www.loc.gov/standards/mets/>; PREMIS (Preservation Metadata): <http://www.loc.gov/standards/premis/>; MODS (Metadata Object Description Standard): <http://www.loc.gov/standards/mods/>. Dublin Core is otherwise an independent standard supported by the Dublin Core Metadata Initiative: <http://dublincore.org/>.

When it comes to standards that are maintained by commercial companies, their adoption represent a bigger risk, as their support and maintenance will only be guaranteed as long as the company considers them commercially sustainable. In some cases, however, a proprietary format can become an open standard. This is the case, for instance, of one of the most used formats for textual storage, the PDF (Portable Document Format) developed by Adobe.[162] The format was developed in 1993 and its specifications were made available from the very beginning, but it remained a proprietary format until 2008, when it became an open standard, certified by ISO.[163] The same route has been taken by Microsoft Office formats, that have become ISO standards in 2012.[164] It is worth noticing how both formats have become XML-based on the occasion of their opening up as formal standards. XML, in fact, as been defined as the 'acid-free paper of the digital age' (Price, 2008, p. 442), because of the fact that it is text-based, human and machine readable without the need of any specific software, and platform independent, all facts that make it suitable for long-term storage.

8.2 Preserving Data, Preserving Interfaces

We have seen the argument according to which the use of an open standard helps with the preservation of a digital object if there is enough critical mass of data to make it viable to create a converter if the standard happens to be abandoned. This argument is germane to the one that says that if an artefact will be perceived to be valuable, it will be taken good care of. These are hardly reassuring arguments for editors who would like to be told that their editions will survive for sure, but these reassurances are hard to give. And even if weak, this type of reassurance comes at a price, and not a cheap one. The price is the interface. In the previous chapter both the scholarly relevance and the problems connected with the political

---

[162] http://www.adobe.com/uk/products/acrobat/adobepdf.html

[163] ISO 32000-1:2008 - Document management — Portable document format — Part 1: PDF 1.7". Iso.org. 2008-07-01 <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=51502>

[164] Office Open XML (also informally known as OOXML or OpenXML – not to be confused with Open Office XML, the format adopted by Open Office: <http://www.openoffice.org/xml/>) is a zipped, XML-based file format developed by Microsoft for representing spreadsheets, charts, presentations and word processing documents and it is recognisable by the suffix 'x' and the end of the file extensions (such as, for instance, 'myfile.docx'). The format was standardized by ISO and IEC (as ISO/IEC 29500): *Information technology -- Document description and processing languages -- Office Open XML File Formats -- Part 1: Fundamentals and Markup Language Reference.* <http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=61750>

sustainability of interfaces was introduced, now is the moment to analyse the problem they pose for long-term preservation, namely their technical sustainability. Provided that the text of a digital edition is encoded following an international standard and equipped by a suitable set of metadata, then it should be relatively easy to preserve and, if necessary, to migrate; the same cannot be said for the interface.

Interfaces are the territory where the negotiation of meanings between the producers of the digital resource and the user takes place. However, they are complex and fragile objects. For instance, it should be remembered that the role of the 'producer' of a digital resource cannot be attributed to the editors alone, nor to their team of developers. The production of complex digital artefacts relies on several layers of software: the operating system, a set of libraries of utilities, the web server, the browser, and so on; all of these components are mutable, are controlled by different entities, are available in many versions, are competing against each other and present a set of reciprocal compatibilities and prerequisites. Each digital edition is based on a custom, idiosyncratic selection of these components, and this makes the effort of maintaining more than one edition within the same system a very demanding task. The many editions developed at the Department of Digital Humanities at King's College London can be taken as a case study. Established in 1991 with the name Centre for Computing in the Humanities, it has contributed to the creation of around ninety digital resources across its more than twenty years of existence,[165] all of them created at different stages of maturity of the digital technologies we are now accustomed to use, and using slightly different sets of software, distributed along a continuous line of development. The curation and maintenance of these resources are mammoth tasks, only made possible by modular development that allows the reduction of the number of variables of the software deployed. This effort does not come cheap, and it is made affordable almost entirely by the fact that all these resources have been developed within the same institution and therefore their architecture is more or less familiar.

But what about all the valuable digital resources that have been developed outside a specialized centre for Digital Humanities: how does one maintain them? In 2004 I was technical director of a project called OperaLiber, a digital collection of opera librettos (Pierazzo, 2006). The project was based at the University of Pisa and had been supported by the Italian research council (MIUR) on two consequent occasions, 2002 and 2005. The website has been developed by a company specializing in digital humanities developments,[166]

---

[165] See <http://www.kcl.ac.uk/artshums/depts/ddh/research/index.aspx> for list of project produced by DDH.

[166] The company that conducted the development is Netseven: <http://www.netseven.it/en>.

using a set of solutions widely used within the DH community;[167] the librettos themselves had been encoded in XML-TEI and well documented. The problems started when a web host had to be found. After several unfruitful attempts, the University of Bergamo eventually accepted the task, but the IT team there was, and is, unfamiliar with the set of software deployed. The result is that the search and the browsing of the website does not work anymore, and nobody on site seems to know how to fix the problem; the resource as a consequence is unusable. This sad story is emblematic of the difficulties that are encountered if digital resources are to be kept alive outside the organization that produced them.

The problems encountered by OperaLiber also hint at a further issue that has been only cursorily mentioned above, namely the cost of maintaining the resources. The choice of hosting the resource in a university that does not provide the necessary support is due to the fact that it was the only institution that was able to do it for free. Keeping digital resources alive after they are finished is not trivial in terms of effort and cost. However, research funds only rarely include provisions for the long-term maintenance of the resources, leaving each producer of digital resources to develop their own financial model to support them. In some cases the maintenance is guaranteed by good will and by personal and institutional interest. In other cases, the resources are subscription based, kept behind a pay-wall. Such pay-walls are despised by many who would like to be able to access all scholarly content for free, but they are considered a necessity in many cases. A few years ago the case of the *Parker on the Web* project made a sensation, as the pay-wall in this case was (and still is) perceived to be exorbitant.[168] Yet, it is hard to argue with the fact that sustaining such resources is expensive, and many scholars have recognized it: printed books are available at a price, why then must digital editions be for free? Peter Shillingsburg, for instance, admits that there is no other choice than publish digital editions 'for profit … for if it fails to maintain itself, it will fail the world of knowledge and scholarship' (2006, p. 105). The problem, however, is not whether to charge or not, but what is a reasonable amount to pay, and what is a sustainable financial model. Most publishing houses have an annual subscription model, meaning that digital publications are not technically purchased but are acquired through a lease. This model, which responds to the need of digital resources to be curated regularly and indefinitely, means that the final cost paid by the library or by private customers cannot be established, as it will grow every year. Furthermore, the day a library or a private scholar decides to stop

[167] The software architecture is made of a combination of the XML native database, eXist, MySQL database and PHP: see the technical documentation at <http://193.204.255.27/operaliber/index.php?page=/operaLiber/documentazione>

[168] See the price list published by Harrassowitz at <http://www.harrassowitz.de/documents/ParkerPricing.pdf>.

paying, the entire resource becomes inaccessible, therefore not only the cost is ever growing, but it will need to remain as such in order to see long-term return for the investment. Maintaining digital resources in good health is then hardly sustainable for their producer, nor they are sustainable for their users to keep accessing them.

Joris van Zundert and Peter Boot (2012) examine many of the issues discussed above. In their opinion the solution for the sustainability of editions is to conceive them as 'networked' and by making use of tools and services already available on the web.

> Editions should make use of services (e. g. a collation or indexing engine) that are already published on the web, for reasons of both technological and scientific efficiency and reliability. This implies that software/services will be highly distributed (p. 144).

In this way interfaces, so expensive to develop and so hard to sustain, will not be necessary, or could be just minimal, and the edition will move from a self-contained home-grown complex digital object, to a dataset that makes use of distributed services, more or less following the model of the *Archimedes Palimpsest Project*. In their vision, digital libraries will have to provide services for the maintenance and semantic enrichment of digital editions. Some of the type of tools envisaged by this model already exist. Let us consider, for instance, *Voyeur*, a tool suite developed by Stéfan Sinclair and Geoffrey Rockwell (2010). The tool allows for a text or a group of text to be loaded within the system which analyses the input and provides the users with some linguistic analysis (such as word frequency) and visualisation tools, such as word cloud, graph, and KWICK format. The interface is pleasant, simple and intuitive and it accepts both plain texts and XML texts (although the markup is almost ignored). Another tool is Juxta (desktop version) and Juxta Commons (online version).[169] Developed within the NINES framework, they provide support for automatic collation and comparison of two texts at a time. Here too the input format can be both plain text and XML, but this time some support is offered to exploit the markup.

This vision is indeed tempting: freed from the concerns of developing output formats and publishing infrastructures, editors will find themselves once again confronting only the part of the editions they are familiar with and experts on: the text. In addition, editions will become much less expensive, becoming therefore affordable also to scholars without an IT infrastructure at hand, encouraging as a consequence the uptake of digital editions in general. Sustainability will also be improved: if editions only consist of the dataset, possibly expressed in a sensible format such as the ones discussed above, then their maintenance will

[169] For both software cf. <http://www.juxtasoftware.org/>

become much more feasible and cheaper too. However tempting, nevertheless this solution presents a few problems. The first problem is that it creates a separation between the creators of editions and developers of tools, in a sense recreating the same division (and problems) of editors and publishers. In a print-based framework, editors are asked to produce editions that fit publishing formats; similarly the new environment would see digital editors being asked to produce texts that comply with the tools that already exist, lacking the capability of pushing the boundaries and imagining new tools yet to come. The risk here is that, while the pioneering first generation of digital editions have tried to 'break free' from the constraints of the publishing medium (with all sort of problems and limitations, as discussed previously), reimagining functionalities, and ultimately the significance of providing editions, the second generation of digital editors will instead work with what they are given, more or less in the same way as editors working with print formats. Progress in the provision of digital editions will then be driven by the technology more than by textual research. Naturally, the reality will be much more complex than this, and there will still be room for interdisciplinary research and collaboration, but it seems rather too early to re-establish such a division now, when so much experimentation is still happening.

The second problem is that such a framework does not specify who is supposed to develop such tools or why. According to van Zundert and Boot 'services can (read: should) have multiple implementations' (2012, p. 147), but no mention is made of whose responsibility or vocation it would be to develop tools to serve the digital editors. Will it be commercial companies? If so, which business model is adopted for such production? Will it be research centres? But, arguably, tool development will hardly count as research: it does not now, as funding bodies seek advancement in scholarship more than in functionality, with tools seen at best as side-products of more complex research endeavours. In a recent intervention, Robinson (2013a) seems to identify tool development with Digital Humanities, by which he probably means Digital Humanities research centres. This seems a rather reductive vision of the discipline, but even assuming that editors will keep their affiliation within their own discipline (say, Classics or English Studies) and the development of tools to serve them will be provided by Digital Humanities centres, this does not prevent the Digital Humanities practitioners (or are they researchers?) to pursue their own idea of the needs of editors and then possibly forget about the tools they have created once they find something more interesting to develop. In fact, even if it can be argued that tool development, if not research, is certainly research facilitation, tool maintenance is a completely different matter which presents the same problems of the maintenance of interfaces, such as constant updates because of the rapid update of software and hardware. Both *Juxta* and *Voyeur* were developed within research projects located within Digital Humanities research centres, which

is the model that both van Zundert and Boot (2010) and Robinson (2013) seem to have in mind; however, what will happen when funds run out and the research team will have to devote their time to developing other tools? Furthermore, these tools provide a very simple and restrictive view of what a digital edition may offer in terms of highlighting scholarship and understanding: to be usable by the largest number of editions, they have to be very generalized, as demonstrated, for instance, by the poor support which the aforementioned tools offer for XML markup. In fact, XML markup can be very diverse, idiosyncratic and specialized, and supporting it in a meaningful way for a large number of editions is a very demanding task and therefore it is understandable that only very limited support is offered; nevertheless in many cases the markup is also the area where the scholarship of the editor is embedded, and so this scholarship is more or less ignored as a result.

In addition, as datasets (the editions) are to be produced independently from their tools, in this view either the tools have to assume a specific format for the data, therefore forcing other subsequent datasets to follow the same format over and over again with implications for editorial arguments, or the dataset has to comply with the expectations of the tools, therefore forcing other tools to adopt the same format. The argument is circular and the result is that the development and the research will be limited by the expectations that new tools will work with old editions and new editions will work with existing tools. As a consequence, the data format to be chosen as a base for this evolution toward 'editions 2.0' will be of capital importance, as it will deeply influence (almost freeze) any future development. Van Zundert and Boot seem to propose the TEI as a suitable candidate for this role (2012, p. 147), a choice that makes sense for the many good reasons mentioned above but which is not immune from problems such as, and above all, the difficulty of handling overlapping hierarchies.

The problems and the speculative nature of this model are not ignored by its proposers, though: '[t]he architecture as described above is a theoretical ideal. Realizing such a triple redundancy based network of resources and services would be a considerably ambitious venture even with major industry partners involved' (p. 148); however they maintain, 'we should nevertheless consider it seriously'. They conclude their contribution by indicating that in future the (digital) library could have responsibility for providing support not only for the maintenance of digital editions, but also for their versioning.

Robinson seems to be of a similar opinion (2013b), when, as we saw in the previous chapter, he advocates for the death of the 'one scholar, one project, one digital humanist' ('1S/1P/1DH') model, in favour of the provision of essential tools to editors who should then be left on their own to produce their editions. However, the call for the death of the model

seems at least premature, if necessary at all, coming as it does before the community of textual scholars has been able to elaborate an editorial model for which the development of tools fitting multiple editions will become possible. The assessment of the real potential of the digital medium, with its capability of re-conceptualising the world (McCarty, 1996) is not only not over yet, but has just begun. It is now that the collaboration between editors, digital humanities specialists, research centres, and digital designers is more crucial than ever: Robinson is right when he argues that a phase of digital edition development is finished, but this is the phase of the production of editions based on the assumption that it is possible to transfer scholarly editions from the printed book to the web without questioning all the basics of the editorial work, and not the phase in which we collaborate to find out what we are actually doing and why.

8.3 Preserving Data: Digital Repositories

Data, particularly if conveniently marked up according to a shared, international standard, is then easier to preserve than interfaces. Yet, even this is not a trivial matter either, as the past (and present) have taught us. In the previous chapter the AHDS (Arts and Humanities Data Service) and its closure in 2008 has been briefly mentioned; the parable of the AHDS in fact emblematically represents the difficulty of guaranteeing support for digital data. Founded in 1995, the AHDS was supported by the AHRC (the UK Arts and Humanities Research Council) and JISC (Joint Information Systems Committee) and aimed to provide a permanent deposit of digital research data. In a scheme similar to the regulation of legal deposit within libraries, Higher Education Institutions producing digital research data were encouraged (and, in the case of AHRC-funded projects, required) to deposit their data within the centre. The data service also provided guidelines to good practice, training and consultancy, in order to ensure a homogeneous format and quality of the data. However, in spite of the evident value of the institution, the funding was discontinued in 2008, undermining the very reason for its creation, namely ensuring long-term preservation. The AHRC's decision to cease its support for AHDS[170] demonstrates how the preservation of digital artefacts is not only an editorial concern, but also an institutional, national and possibly (desirably) international problem. The AHRC's press release of 2007 quite extraordinarily stated:

[170] The reasons for ceasing the support to the AHDS most probably reside in its high costs and the feeling that was not the role of a research council to support infrastructure, as suggested by Professor Sheila Anderson (the former director of AHDS) during a private conversation (16 October 2013). However, an official statement explaining such reasons was never issued.

> Council believes that Arts and Humanities researchers have developed significant IT knowledge and expertise in the past decade. The context within which the AHDS was initially supported by the AHRC has changed. Much technical knowledge is now readily available within HEIs, either from IT support services or from academics. Much that generally can be safely assumed now, for example that web sites can be put together and run effectively for the duration of a project, could not be assumed ten years ago. Council believes that long term storage of digital materials and sustainability is best dealt with by an active engagement with HEIs rather than through a centralised service.[171]

According to this statement, then, the responsibility of long-term preservation resides in the hands of the producer of the resource itself, namely the Higher Education Institution (HEI) that holds the intellectual responsibility of the resource. However, as clearly demonstrated by the OperaLiber and by the CURSUS project (cited by van Zudert and Boot, 2012) examples, the curation of digital products requires expertise and skills that are not necessarily in the hand of IT support centre of the average Higher Education Institution, nor it is clear why they should assume the curation function: HEI are supposed to produce education and research, not necessarily provide long-term preservation, a task that requires the reconfiguration of some of the core tasks of Universities and Research Centres.

Regarded at the time as best practice, the legacy of the AHDS has been continued nationally by several institutions,[172] but its influence has deeply shaped worldwide initiatives, the most relevant of them is perhaps TextGrid, supported by the German government. TextGrid offers a virtual research environment (with tools such as text editors, collation and text mining modules) and a repository where to store digital files in some convenient format, in particular encoded in XML-TEI (Zielinski *et al.*, 2009). Similar initiatives exist in other countries at national or institutional levels.[173]

So there are three possible institutional models for the preservation of digital data:

---

[171] The original enigmatic press release issued by the AHRC is currently only available from the Web Archive: <http://web.archive.org/web/20080607065438/http://www.ahrc.ac.uk/news/news_pr/2007/information_for_applicants_to_AHRC_june_deadline.asp>.

[172] The Centre for e-Research at King's College London, Archaeology Data Service, Oxford Text Archive, History Data Service, Performing Arts Data Service, Visual Arts Data Service (see http://www.ahds.ac.uk/).

[173] See, for instance the TGIR Huma-Num in France is providing a repository service (http://www.huma-num.fr/service/archivage-a-long-terme).

1. *Ad hoc* institutions: mirroring the function of libraries and archives, new specialized bodies are created to take care of digital artefacts; examples are the AHDS and TextGrid. These initiatives not only provide a home for the data once created, but they also support scholars in best practice, the adoption of which facilitates the curation of the deposited data.
2. Producers: the institutions that have developed the resources are responsible for their maintenance. Guidance toward best practice is not formally provided, but the adoption of it is implied in the funding infrastructure: no funding is given if the institutions do not provide convincing preservation strategies. This is the post-AHDS model in the UK, for instance.
3. Libraries: the institutions that historically have preserved a large part of our cultural heritage will be also entrusted with the preservation and maintenance of the virtual patrimony as well. Some libraries already provide some sort of data set deposit system,[174] but most of them tend to provide more cataloguing and delivering activity without formally stating their preservation strategies (if any)

8.4. 'Lots Of Copies Keep Stuff Safe': Redundancy and Preservation

Similar to the use of open international standards guaranteed by stable public institutions, the deposit of digital files within an institutional repository may not be enough to guarantee their preservation: the case of the AHDS is very instructive in this sense. To maximize the chances of preserving one's work, it is important to preserve it in more than one place: the reason why many print publications have survived hundreds of years is because there were many copies of them, a fact that increases exponentially the chance that at least one of them will last for long stretches of time. This is the spirit, for example, of the LOCKSS programme. Based at Stanford University, this acronym quite significantly stands for "Lots Of Copies Keep Stuff Safe".[175] Born from the consideration that 'No single library can achieve robust long-term preservation alone',[176] the service is aimed at libraries that not only will be able to access the digital content, but also will 'own' it in a peer-to-peer environment, helping in this way to keep such content alive. In the same way that the multiplication of print copies helps the global preservation of a particular published product, many copies of, say, a PDF file,

---

[174] The British Library, for instance, collects all the working files of important authors. But cf. <http://databib.org> and the < http://www.datacite.org/> for lists of libraries that offer dataset deposit service.

[175] Cf. the LOCKSS website <http://www.lockss.org/>

[176] Cf. <http://www.lockss.org/about/principles/>

will ensure its preservation. Nevertheless, this is not a problem-immune system: while it is almost impossible to alter a printed object beyond recognition (tampering is normally easily spotted as it leaves 'physical traces' which are of different nature with respect to the ones produced by the publishing houses), it is quite easy to seamlessly alter digital objects. When multiple copies are made, it is possible that some, if not all, will be altered in concurrent and unauthorized ways. Redundancy, therefore, maximizes and makes even more crucial the role of preservation metadata and of digital curation in general as only an accurate set of metadata and good practices will allow the reconstruction of the files' original format.

Not everybody agrees about this concern, though. Reflecting on how texts from antiquity have been transmitted to us in many altered ways, but, in a sense, in formats still available to their readers, Hugh Cayless advises us 'not to be overly concerned about maintaining the integrity of copies of digital resources outside our control', especially when it comes to 'the original appearance of such resources' (Cayless 2010, pp. 148-50); again, it seems that the price to pay in the hope that the digital edition will survive over time is the interface: data can be preserved, their appearance, much less so.

8.5 Open Access and Preservation

LOCKSS and similar systems are mostly intended to preserve eBooks and recognisable standalone items, rather than complex digital resources such as digital scholarly editions. The complex architecture of sources, outputs, interfaces and tools that constitute digital scholarly editions is beyond the scope of the programme. However, the principle that redundancy helps preservation is nevertheless applicable, particularly when it comes to the underlying data. This is especially true if the data, the sources, are freely available without copyright restrictions or DRM (Digital Rights Management). Cayless observes that the control over the digital object provided by DRM 'seems likely to be incompatible with long-term preservation goals, which will necessitate actions like making and distributing copies and migrating from one format to another for an indefinite period of time' (Cayless, 2010, p. 146). Making one's source files freely available for the user to download therefore not only responds to the scholarly requirement of transparency and verifiability, as stated by Bodard and Garcés (2008), but also seems likely to considerably improve the chances of the edition surviving over time and space.

Recent debates in the academic world have revolved around Open Access.[177] Under the label of open access, however, go many different phenomena, behaviours and misunderstandings, the most diffused of which seems to be related to the confusion between Intellectual Property Right (IPR) and copyright. Originally Open Access referred to the possibility of making available within an institutional repository a pre-print version of an article also published on a peer-reviewed journal; this availability could happen simultaneously to the official publication or after a delay of, typically, six months to one year ('Green Route'). Some journals are Open Access native (*DHQ* and *Digital Medievalist*, for instance) and offer all their content freely from their own website; the latter is called the 'Golden Route' of Open Access. More recently the concept of Open Access has been enlarged to cover chapters in books and monographs, the latter, however, being rarer than journals and articles. In order to publish an article in an open access peer-reviewed journal, authors are often required to pay an 'article processing charge' to the publishing houses; however, in most cases authors are at least allowed to deposit a pre-print version on their institutional repositories. In spite of many misunderstandings, Open Access has a lot to do with copyright (that is the right to copy and distribute a digital object), but has little to do with IPR (Intellectual Property Right). In fact, offering a scholarly contribution openly does not mean that the author is giving away his or her intellectual claim over the content of the contribution itself, nor his/her copyright, in fact: Open Access does not mean free licence to plagiarize, it only means that the work is accessible without restriction, in the same way that one does when reading a printed book or an article from a public library. In reality the concept of copyright itself is quite complex as it includes not only the right to copy, but to distribute, alter, and make a profit out of these operations; Open Access does not regulate these aspects, which are left to the individual repositories and publishers to decide by the provision of appropriate licencing. Many open access digital resources are nowadays distributed under a Creative Commons licence, an international licencing format which offers the possibility of choosing which rights over the content to retain and which to relinquish.[178] Founded in 2001, it is now the licencing format of choice for high profile institutions (such as the White House and Europeana), and of social networks such as Flickr. While Creative Commons mostly applies to textual content, other similar licencing systems are applicable to software and other type of digital products. For instance, the TEI Guidelines are licenced

[177] The field of Open Access is wide and ever changing. For a 'light' introduction, it may be useful to consult Suber (2012), which is, needless to say, available as Open Access monograph form at <http://mitpress.mit.edu/books/open-access> as well as in print.

[178] Creative Commons, <http.creativecomons.org>.

under a Creative Commons licence and a BSD 2-Clause licence,[179] as stated on the TEI website 'the CC+BY licence is generally appropriate for usages which treat TEI content as data or documentation. The BSD-2 licence is generally appropriate for usage of TEI content in a software environment'.[180] Open Access and Open source does not mean anarchy and legalization of piracy: quite the contrary, it provides a regulated framework to enable the sharing of scholarship.

In the United Kingdom the decision of the Research Councils UK to make Open Access more or less mandatory from 2013 for all publicly funded research has fuelled passionate discussions which have seen scholars in the humanities struggling to advocate for the necessity to consider the case of humanities scholarship differently with respect to the sciences, due to the different role of the typology of publications in the two disciplinary groups. Similar debates are happening in many countries and at international level as well.[181] While for the sciences the most common format for the publication of scholarship is the article in peer reviewed journals, for the Humanities the monograph is of a fundamental importance too; the evolution toward open access has affected journals publishers since the beginning of the century, but has only marginally affected the publishing houses of monographs. The debates around this topic are varied and include issues like funding (who is responsible for paying for articles and books), self storage (how, when, which access to provide), the role of OA for career progression, and so on, and will have deep implications for the way we will disseminate our scholarship which at the time of writing is hard to imagine in their full extent.

A detailed discussion of the topic is outside the scope of this book; however, open access has been mentioned here as there seems to be a correlation between the provision of an edition as open access (freely accessible) and open source (in all its components), and long-term preservation. Whether this correlation will be proved to be significant, only time will tell. In the meantime it seems clear that providing free access to the sources as well as to the interface is a very good idea which has more pros than cons; in fact not only is it a practice that meets scholarly criteria in making the edition more accountable, but it also improves its impact and most probably its duration. Actually, it seems hard to find any

---

[179] See The Open Source Initiative http://opensource.org/licenses/BSD-2-Clause.

[180] 'TEI: Licensing and Citation'. TEI http://www.tei-c.org/Guidelines/access.xml

[181] For instance in the US: in 2013, John Holdren, Barack Obama's director of the Office of Science and Technology Policy issued a memorandum directing United States' Federal Agencies with more than $100M in annual R&D expenditures to develop plans within six months to make the published results of federally funded research freely available to the public within one year of publication.

detrimental factors, besides what James Cummings calls the 'concerns of delicate academic egos' reluctant to expose their 'dirty laundry' to scholarly inspection in the fear of having incurred so-called 'tag-abuse' (2009, p. 316). However, as pointed out by Cummings, it is only 'through the discovery of our joint errors and misunderstandings' that scholarly progress can happen.

8.7 Trusting Editions: Heraclitean Editions

The ephemeral nature of digital editions has been held responsible for the scholarly community's distrust of digital editions. But, however important, this is not the only reason why the reception of digital editions is not as enthusiastic as the producers of these resources hope it would be. Digital editions are mutable objects: they change because the technology around them changes and therefore they are forced to adapt to it; they change because they can be changed; and they change because they are inherently mutable, interactive objects. Printed editions freeze in time the achievements of scholarship: once the text is printed, it cannot be changed easily, even if new compelling evidence is found or really disturbing errors are spotted. This offers disadvantages, like the ones just mentioned, but also advantages. Printed versions are citable, and the citation is permanent: if a particular quote is said to come from a specific page of a given edition, this fact will not change with time. New editions may be produced with a changed distribution of text per page, however, the edition that was originally referenced will remain the same.

The mutability of digital editions not only affect their citability, but also makes post publication peer review quite difficult, as maintained by Bodard (2008, § 32), if indeed one can apply the concept of publication to digital editions at all. Digital editions are open-ended by definition: one can always add, correct, modify, or update something, especially if the editions are available on the web. This allows for a growing knowledge base, and for progress in scholarship. However, it makes it hard to cite, affecting its reliability and the grounds of the editorial endeavours themselves. One of the main purposes of editorial work is to provide readers with a stable edition, but digital editions on the web are not stable, and this has deep implications for the perception of both the editorial work and the editions themselves. As stated by Sperberg-McQueen:

> Editors may justly feel that electronic editions have translated them from a stable environment with difficult but well-known problems into a river of Heraclitean flux, in which everything is changing from moment to moment, and the editor and edition are expected to adapt actively to those changes from moment to moment, without

> being able to rely on many of the principles which used to be stable guides to editorial thinking (2009, p. 30).

He then concludes that in order to successfully engage 'with the mutability of electronic edition… may require a different way of thinking'.

The need for a new theoretical framework for editorial endeavours is perhaps most evident within the so-called social or open editions, including here all those editions that acquire content on a public call, counting therefore crowd-sourcing and post-publication comments and annotation (Siemens, 2010; see chapter 1). Van Zundert and Boot rightly question the academic status and acceptability of such editions: 'do we really accept an edition will change if one of its components or data resources elsewhere is modified? […] Or do we accept possibly resulting internal inconsistencies?' (2012, p. 149).

One way to deal with the mutable, Heraclitean nature of digital editions is to implement some type of versioning system, possibly patrolled and supported by the digital libraries of the future, as suggested by van Zundert and Boot (2012). The *Inscription of Aphrodisias Project*, for instance, already follows a versioning pattern; the project exists in fact in two distinct versions, the *Aphrodisias in Late Antiquity* of 2004 (ala2004) and the *Inscriptions of Aphrodisias* of 2007 (IAph2007). The project team preferred, then, rather than update the 2004 version of the website, to produce a second updated edition, in order to provide a more stable and therefore quotable set of resources (Bodard 2008, §§ 30-33; Rouché 2009, p. 163). The *Devonshire Manuscript project*,[182] which is at the base of the theoretical development of the social edition paradigm, is published within a wiki-like platform: Wikibooks.[183] This platform allows for an historical overview of all interventions on the texts, including references to those responsible for such interventions, also providing the users with the possibility of 'undoing' any undesirable intervention. In fact the software development community has elaborated elegant ways for dealing with versions and modification distributed over a number of collaborators, which are embodied by repositories such SourceForge and GitHub.[184] These platforms serve as distributions hubs for software, and while they highlight the latest version, they also allow for browsing backward all the versions of a file and of a group of file, and seeing who is responsible for each change, as well as comparing any two given versions. The same functionalities are available for private storage and backup resources such as Dropbox and, to a lesser extent, Google Drive. Digital

[182] Cf. <http://en.wikibooks.org/wiki/The_Devonshire_Manuscript>.

[183] Cf. <http://en.wikibooks.org/wiki/Main_Page>.

[184] Cf. respectively <http://sourceforge.net/> and <https://github.com/>.

products, as any type of text, come in versions, in particular when these products are the result of a social effort, and while this might cause some problems in terms of stability, it could also be made a strength. By embedding a versioning system into a digital scholarly edition, for instance, the editorial work will become even more accountable and it would be possible for the user to follow the evolution of scholarship; In this way a superseded version, which may have been quoted and peer reviewed, will not disappear, but will remain accessible over time. The fact the scholarship evolves is hardly a new concept, therefore by exposing the variation instead of hiding it under a glossy front cover will also avoid the false sense of certainty about textual facts that is considered so dangerous for the future itself of textual scholarship. The new ways of thinking invoked by Sperberg-McQueen (2009) may lead editors to consider the possibility of working in 'the open', as software developers have done for years, showing more stable alongside intermediate versions of an endeavour. Trustworthiness is not only given by stability, but it is also built around accountability and accessibility over time.

But beside the active intervention of web users and the updating of the resources by the project team, there are other reasons why digital editions change in ways that are beyond the control of their creators. Quite simply, editors and developers of digital editions cannot predict the hardware setups of their users. Will they access the edition on a small notebook, a big desktop screen or a mobile device with a touchscreen? And which browser will they use, a modern, up-to-date one that supports advanced functionalities, or an old one, which will prevent access to some more or less essential features? And what about the colours and brightness? Which level of distortion will they bring to the inspection of, say, a digital facsimile? While it is relatively easy to predict what readers will be able to see when looking at the pages of a printed editions, when it comes to digital editions the variables are simply too many, with each reader potentially looking at something that is slightly (or greatly) different with respect to someone else (Sutherland and Pierazzo, 2012, p. 200).

Long-term preservation of digital editions in all their components will remain in the list of desiderata for quite a while, but it may be worth asking here if these should be pursued at all. The discussion above has demonstrated how, in spite of good practice, the right choice of format, the creation of preservation institutions, the creation of professions, the multiplication of copies, we cannot guarantee we will be able to preserve editions more than a handful of years. The temporary or permanent closures of trustworthy institutions, the modifications of files, the endless work of migrations and the consequent rising costs make the preservation of source data a mammoth task; it may just prove impossible to preserve interfaces. Best practices such as the choice of the right data format, the provision of

lightweight and 'disposable' interfaces and the distribution of sources may make it possible to preserve the editions, but they will not necessarily make it feasible. On the positive side, though, it may be said that we are not alone in fighting for the preservation of digital artefacts: this is a problem on a global scale and massive investments have and are being made. Good results are being achieved, in particular with the enforcement of the use of standards and preservation-friendly data formats. The more we commit to the digital, the more likely it will be that feasible solutions will be found. In the meantime, though, scholars and editors may have to rethink the way we conceive scholarship. If stability is no longer a requirement for editions, what are the new requirements? And what are editions for? The mutability and the shaky ground on which digital editions are built may even have some positive effects over scholarship and textual scholarship. Editions will require the help of their readers to stay alive, both in term of storage, but also in the engagement with the content itself, annotating it and paying careful attention to valuable research output, but this is deeply undermining the belief that an edition is a 'definitive' endeavour and that texts are entities that can be established once and forever. Editions and the texts they contain are only one of the episodes of the history of a particular work. Editions are worthwhile activities not because they establish the perfect text that will enable scholarship, but because they are scholarship and represent the advancement of knowledge in a particular branch of the culture. As maintained by Sperberg-McQueen "The possibilities for mutability in electronic editions will require a great deal of editorial theory to be rebuilt" (2009, p. 30), a fact that may have many positive implications.

8.8 Trusting Digital Editions? Peer Review and Evaluation of Digital Scholarship

The mutability and instability of editions may be considered the biggest drawback in the adoption of a digital paradigm in editing, in particular if digital includes the format of the output. Connected with and derived from these are problems with citability and peer review of complex digital products, as seen above. More generally the scholarly community seems to be having difficulties in assessing the quality of digital editions, while their producers struggle to get recognition in terms of career development and arguing the scholarly value of digital outputs in general.

Digital resources are difficult to review because they require specialized competences, not only to examine the particular scholarly argument made by the resource, but also the choices made to deliver it, namely the technologies, the formats, the interfaces, and so on. Digital outputs are normally produced within a research project that is peer

reviewed in order to obtain the funds from a research body, but that leaves open the question whether the deployment of the project has achieved the scholarly potentials envisaged by the proposal. Funding bodies have different rules, but in most case at the end of a project they only require a self-evaluation from the main investigators and no external assessment. When publishing a monograph or an edition in print on the other hand, even if the book embodies results from a funded research project, the publication itself is peer reviewed before being published within the publishing house, and again afterward, in specialized journals. These mechanisms have no correspondence in the digital world: no review is necessary to publish a website and websites are only rarely reviewed in journals. For this reasons digital scholarship looks less, well, scholarly than when in printed form, as it lacks, or seems to lack, the quality assurance that characterizes print scholarship. In fact, digital resources can be assessed and peer reviewed as much as traditional formats, but this evaluation takes different shapes. A chapter of the book *Digital_Humanites* is devoted to 'How to evaluate digital scholarship' and provides guidelines of good practice to both the authors of digital resources (how to credit collaborators, for instance) and evaluators, reminding us how 'peer reviews can happen formally through letters of solicitation but can also be assessed through online forums, citations, and discussions in scholarly venues' (Burdick et al. 2012, p. 129).

Pre-publication peer review was introduced to the print world in order to assure that only the best contributions would be published in a world of scarcity such as the one regulated by academic publishers. Peer review, therefore, fulfilled two main functions: provide a reasonable estimation of the commercial viability of a given scholarly outcome, and to give a reasonable assurance about the scholarly quality of the same. For digital publication the first function is irrelevant, while the second may reasonably be moved to the post publication stage, as it does not really correspond to the new workflow, especially because, unlike print publications, digital publications can be, and are, modified after their first publication. The questions therefore are: how can we ensure a healthy stream of post-publication reviews for digital resources, and how can we ensure that post-publication modifications do not invalidate the reviews.

An answer to the first question has been attempted by initiatives such as NINES (*Networked Infrastructure for Nineteenth-Century Electronic Scholarship*). NINES is a scholarly organization which offers a peer review infrastructure and access to 'digital work in the long 19th-century (1770-1920), British and American'. The NINES peer review process is aimed as addressing the fact that digital resources 'often had difficulty establishing their credibility as true scholarship' because of the lack of an appropriate peer review

mechanism.[185] This is meant to be a type of post-publication peer review, however, and NINES also provides a set of Guidelines for new projects in the belief that some crucial choices (such as, for instance, the choice of the appropriate data format), have to be taken at the beginning of the life of project, otherwise it would be almost impossible to rectify them at the end. NINES attempts to assess the two main component of a digital resource, namely the content and the digital infrastructure: 'first, is the content important and interesting to existing scholarship; second, is the material presented in a clear, accessible, well-organized, and well-documented fashion?'.

A similar mechanism has been created for Eighteenth-century scholarly production by *18 Connect, Eighteen-century Scholarship Online*[186] and for the Middle Ages by *MESA-Medieval Scholarly Alliance*.[187] This proliferation of peer review infrastructures demonstrates how crucial the evaluation and the scholarly trust of digital scholarship is. The MLA (Modern Language Association) in 2012 issued a set of *Guidelines for Evaluating Work in Digital Humanities and Digital Media*, including sections on how to appoint and promote staff and a set of guidelines for candidates aspiring to improve their career paths. A section of the 2011 issue of the MLA peer-reviewed journal *Profession* is also devoted to evaluating digital scholarship: it contains six articles which are available as open source content, unlike the rest of the journal which is behind a paywall. Here the contribution of Kathleen Fitzpatrick is particularly relevant, as she reminds her readers that the real challenge 'is to recognize and accept that expert assessment where it does, in fact, already exist' (2011, p. 197); she adds how '[d]igitally published projects are often advised, reviewed, and commented on by experts in the field, and yet that oversight is often conducted in the open, frequently after the material has been released, and usually it does not result in "publication" by a neutral third-party organization' and that in spite of the fact that such projects are 'published under the auspices of respected university or disciplinary organizations', they still 'are tainted by association with the self-publishing that ostensibly pervades the Internet, where anyone can post anything' (p. 199). She concludes her contribution with an appeal to all to engage in full with digital scholarship and with its innovative forms of assessment such as, for instance, 'recommended links on Twitter' (p. 201).

Innovative types of scholarship require innovative forms of assessment and the flexibility of the community within which it has to be assessed. The Digital Humanities community is therefore putting a great deal of effort into providing guidance on how to

---

[185] <http://www.nines.org/about/scholarship/peer-review/>

[186] <http://www.18thconnect.org/>.

[187] <http://www.mesa-medieval.org/>.

evaluate and how to assess scholarship and scholars. The engagement of professional bodies such as the MLA demonstrates that this strategy is working and in spite of the fact that 'texts on screen look remarkably alike, despite profound differences in quality' as observed by Shillingsburg (2006, p. 87), great progress is being made in improving the digital literacy of scholars and reviewers of scholarship.

The fact that digital editions are digital is not the only reason why their evaluation is a complex matter. The other major issue with digital scholarship is that it is by nature collaborative and interdisciplinary in ways unknown to print-based work, as discussed in Chapter 5. Shillingsburg effectively notes that 'it takes a village', and that 'creating an electronic edition is not a one-person operation: it requires skills rarely if ever found in any one person' (Ibid. p. 94). Nevertheless, this is also true for print publications: nobody expects editors to print, distribute and market their own editions! What is different in the preparation of digital editions is that the type of collaborations between editors and experts of digital technologies is quantitatively bigger and qualitatively deeper (see chapter 5). In addition, most of the applied work (creation of interfaces, modelling and encoding texts, developing tools) can be qualified and is presented (including in this publication) as scholarship itself and is distributed among many people, included the editors. According to Bethany Nowviskie, in the attempt of digital scholars to get recognition for their work, there has been a strong movement to equal digital output to print objects such as articles and monographs, which in the Humanities are normally the output of one scholar. However, she objects that a 'defensive stance that asserts the uniqueness of a scholar's output by protecting an outmoded and sometimes patently incorrect vision of solitary authorship is unsupportable' (2011, p. 170). Using the words of the American Historical Association (AHA) *Statement on Standards of Professional Conduct*, she claims that digital scholarship should be analyzed 'as a process, not a product'.[188] The INKE project, for instance, offers an interesting model for the recognition of the various contributions within the project as its 'authorship convention implies a model that has more in common with the sciences rather than humanities with its focus on the sole author' (Siemens, 2009, p. 7). The result is the publication of scholarly contributions with a number of authors, which is certainly unusual in a Humanities context.[189]

---

[188] P. 171. American Historical Association. *Statement on Standards of Professional Conduct*, 8 June 2011, http://www.historians.org/about-aha-and-membership/governance/policies-and-documents/statement-on-standards-of-professional-conduct>.

[189] For instance the contribution published in the volume *Collaborative Research in the Digital Humanities*, edited by Deegan, M. and McCarty, W. (Ashgate), 2011, one can count 23 authors (implicitly divided by the use of the word 'with' in seven main authors and fifteen contributors), plus two research teams: Ray Siemens, Teresa Dobson, Stan Ruecker, Richard Cunningham, Alan Galey,

While one has to admire this brave (and ethical) challenge to the established expectations of the Humanities field, this practice also raises the question how to evaluate such a long list in terms, for instance, of career progression. While burying one's name in such a long list might not be an issue for established academics, it may prove lethal for young researchers, as observed by Nowviskie: 'junior scholars (driven by career anxiety and facing a dearth of models for expressing individual contributions to collaborative projects in the digital humanities) may be reluctant to challenge long-standing systems that locate agency in authorship' (2011, p. 170).

Innovative practices such as the ones proposed by the INKE project, as well as contributions like the ones from Nowviskie, demonstrate a raised awareness over the issue of recognition of collaborative work. Much more needs to be done, though, in particular as this change in the assessment of authorship goes together with the recognition of a new form of scholarship based on the collective effort of multidisciplinary teams.

Claire Warwick and Lynne Siemens, with Michael Best, Melanie Chernyk, Wendy Duff, Julia Flanders, David Gants, Bertrand Gervais, Karon MacLean, Steve Ramsay, Geoffrey Rockwell, Susan Schreibman, Colin Swindells, Christian Vandendorpe, Lynn Copeland, John Willinsky, Vika Zafrin, the HCI-Book Consultative Group and the INKE Research Team.

9. The Present and the Future of Digital Scholarly Editions

In an insightful contribution of 2009, Kenneth Price reflects on what to call the 'stuff' we do with texts and the other things we append to them. He examines the understandings and cultural associations behind terms such as 'edition', 'project', 'archive', and 'digital thematic research collection' to conclude that none of them really represents the work we do nor the products we craft; he finally, half ironically, propose the word 'arsenal'. Peter Shillingsburg, on the other hand, proposed the concept of 'knowledge site' (2006), while Joris van Zundert and Peter Boot speak of 'digital edition 2.0'. The title of this book sticks, perhaps unimaginatively, with 'digital scholarly editions'. Clearly this variety of names is a manifestation of the changing nature of the outcomes of editorial acts. In his quest for the right word, Price underlines two main characteristics that the name he is looking for should illuminate: firstly, that it is a website, a digital and physical artefact which is infinitely extensible; secondly its collaborative nature. In his proposal to use 'arsenal' he declares that he 'like[s] the emphasis on workshop [included in the definition of arsenal] since these projects are so often simultaneously products and in process. I also like the stress on craft and skill, a reminder that editing is not copyist work'; he then adds how the terms imply a collaborative effort (Price, 2009b, § 40). However, in spite of the very good reasons for proposing the adoption of such a term, it still does not seem adequate to the task. It belongs to totally unrelated semantic fields, the lack of recognisability of the term in such a context, and the fact that none of the meanings that, according to Price, make arsenal a good choice are immediately associated with the term itself. It is anyway fair to say that Price's proposal was a sort of tongue-in-cheek suggestion, highlighting the inadequacies of the current terminology to describe complex and evolving objects such as the ones he had in mind.

One of the terms examined by Price is project. In his argument, Price notes how an edition may be the product of a project, which includes much more than the edition even if the production of the edition itself is the main purpose of the project. In fact, it includes also all the 'back office' of a collaborative editorial endeavour, namely what Boot has defined the 'mesotext' (Boot, 2009), the scholarly relevance of which has been discussed in Chapter 5:

> the compiled email discussion list that fitfully records the building of the *Archive* and the thinking that has gone into it. The documentation of a project, in our case, includes the behind-the-scenes Works-in-Progress page, with its assortment of information, including grant proposals, minutes from Whitman planning meetings over the years, a manuscript tracking database, an image warehouse, and project-related humor (Price, 2009, § 14).

The same distinction offered by Price between edition and project is displayed by the *Jane Austen's Fiction Manuscripts Digital Edition*, as the website includes a section where the editorial principles are outlined, and another one containing details of the project itself. While the website shows a distinction between the two concepts, it also includes materials relevant to both, as they are considered equally relevant components of the same 'thing'. Keeping them separate may prove complex in practice; and in fact for Price expressions such as the 'Whitman edition' and the 'Whitman archive' are quite interchangeable. However, talking about projects over editions may not be well received: Price reports for instance that he was discouraged from using 'project' within a grant application to the NEH, the rationale being that a project sounds more ephemeral than edition.

The project is allegedly the basic unit of Digital Humanities scholarship (Burdick et al. 2012, pp. 124-125). Projects in fact possess two defining characteristics that have been presented as core features of the Digital Humanities, namely they are collaborative and they are projective, 'in the sense of futurity, as something which is not yet' (Ibid. p. 124). They are also connected to a specific funding mechanism, that is the research grant provided by private and interdisciplinary funding bodies. The result is that Digital Humanities features and promotes a 'project-based scholarship' (p. 130-131), that is to say scholarship that is encapsulated by the duration of the funding, with people aggregating around a finite endeavour as long as funding lasts.[190] Of course project-based scholarship is not born with Digital Humanities, but that in the same way as collaboration, it is quantitatively bigger and qualitatively deeper than other disciplines in the Humanities. The fact that projects are meant to build 'something which is not yet' and that involve collaborative and iterative processes fits quite well with digital scholarly editions. As well as reflecting on the reality of the production of such resources we have seen how the workflow and publication of a digital scholarly edition requires a certain number of people and infrastructures which can probably only be funded by research grants. However, even if we could push it and call digital scholarly editions project-based scholarship, such a label does not say anything about which type of scholarship it involved to build digital scholarship. As Price observes '*Project* can describe everything from fixing a broken window on the back of a house to the Human Genome Project' (2009, § 14).

The difficulty of distinguishing between different digital resources is at the heart of Price's attempt to find a new term. Palmer, facing the same dilemma, proposed 'thematic research collection' (2004), a definition that, in spite of its appropriateness ('may be the most

[190] The ethics and economic implication of this model are under scrutiny. Cf. further below for a discussion of these issues.

accurate term for what many of us are attempting'), 'is neither pithy nor memorable' (Price, 2009, § 26). However, a more substantial criticism can be raised against this term: 'Does collecting — the emphasis in Palmer's description — qualify as research, as a scholarly genre? […] Nothing in the term indicates editorial rigor and nothing points to the value added by scholarly introductions, annotations, and textual histories' (Ibid. § 27).

The other label examined and rejected by Price is 'archive', which proves to be unsuitable even if Price himself co-directs an endeavour actually called 'archive', namely the *Whitman Archive*. In fact, archive is a very widespread term used by some of the most important scholarly editions on the web: one can think of the *Rossetti Archive*,[191] the *William Blake Archive*, the *Whitman Archive* itself, and more recently the *Beckett Archive*.[192] Are these archives because they are collections of primary sources, or are they something else?[193] The fact that the *William Blake Archive* has won the MLA for 'Distinguished Scholarly Edition' in 2003[194] demonstrates that, whatever the name, these endeavours are somewhat more complex than their name seems to suggest. And this is perhaps the reason why Price

---

[191] The complete title is *The Complete Writings and Pictures of Dante Gabriel Rossetti. A hypermedia Archive,* edited by Jerome J. McGann. Institute of Advanced Technologies in the Humanities (IATH), University of Virginia, 2008.

[192] For the last of these, the label 'archive' comes form the URL: <http://www.beckettarchive.org/>, however the website, quite significantly, features another name: *Samuel Beckett Digital Manuscript Project*, Directors: Dirk Van Hulle and Mark Nixon; Technical realisation: Vincent Neyt.

[193] The OED defines archive as 'a place in which public records or other important historic documents are kept' and 'A historical record or document so preserved'; in both cases the word is mostly used in the plural (and exclusively so for the first meaning). The use of the term 'archive', in the singular, for endeavors such as the one mentioned above therefore does not exactly match these meanings as it refers to the website (and therefore a 'place', as in the first meaning offered by the OED), but also to the collection and the act of collecting digital objects.

[194] The list of all winners of this prize is found on the MLA website: http://www.mla.org/pastwinners_distsch; the William Blake Archive website features the rationale for the award of their prize: 'The William Blake Archive is a dazzling combination of hypertextually organized texts, bibliographical and historical commentaries, and beautifully reproduced visual images, including thousands of plates of Blake drawings, watercolors, and manuscripts. In the past the prize has been awarded to single volumes in a multivolume series. This year's prize, similarly, is awarded to major scholarly additions to the archive published in 2001 or 2002, including Blake's first group of twenty-one watercolors illustrating the book of Job and three copies of *The [First] Book of Urizen*. If, as has been frequently suggested, the future of editorial scholarship lies in online editions, the William Blake Archive has set a high mark for future editorial practice through its clarity, user-friendliness, beauty, and erudition.' http://www.blakearchive.org/blake/MLA.html

seems unhappy with the name, because it suggests far less scholarship than may actually be involved. In fact Shillingsburg is quite firm in stating how 'the level of critical intervention is minuscule in the electronic archive' (2006, p. 156). Again in a recent intervention at the *Scholarly, Digital and Social Edition* conference (2013), Shillingsburg strongly maintained that the difference between editions and archives is that the former make a scholarly argument, while the latter do not. In his vision, archives are simulacra, surrogates for the originals, and even when digital facsimiles are accompanied by transcriptions, they only provide a very good surrogate, without reaching the status of scholarly editions. In the same vein is Robinson, in his contribution at the same conference, he objected that there is a clear distinction between making editions, which requires 'making decisions' and should be done by scholarly editors, and making archives, which are done by digital humanists and, apparently, do not require any decision. He then points out

> the common first error made by digital humanists, – and, I fear, by many editors: that the needs of textual scholars, and indeed the interests of readers, can be perfectly served by digital archives. They cannot. An edition is an argument about a text. We need arguments; without arguments, our archives are inert bags of words and images (Robinson 2013, p.1).

Price objects to this rather restrictive view of the function of an archive (2009, § 10, n.7), and it is also contradicted by the Directors' evaluation of what the *Samuel Beckett Digital Manuscript Project* (or 'Beckett archive') is and does, to pick up one of the aforementioned resources. The section of the Editorial Principles and Practice of the BDMP argue for a continuity rather than a dichotomy between archives and editions:

> Usually, digital 'archives' are distinguished from 'editions' because the latter offer a critical apparatus. An interoperable tool such as CollateX can enable any user - not necessarily an editor – to transform a digital archive into an electronic edition. As a consequence of these developments in scholarly editing, the strict boundary between digital archives and electronic editions is becoming increasingly permeable, resulting in a continuum rather than a dichotomy. (Van Hulle and Nixon, 2013, Editorial Principles and Practice).

Van Hulle and Dixon (2013) use the availability of tools able to manipulate the materials as an argument for the transformation of an archive into an edition; to this we can add that the extensive editorial work that organizes, typifies, transcribes, segments, compares, annotates different versions of the same work which the BDMP does to the highest

editorial standards hardly seems to fit the description of 'inert bags of words and images' (Robinson 2013, p.1). Jerome McGann (2010) is also uneasy with the distinction between archive and edition, when he declares that 'traditional text-centered editorial models could be surpassed in a digital medium', with the *Rossetti Archive* representing in his view a 'hypermedia scholarly "edition"'.

The harsh distinction championed by Shillingsburg and Robinson seems to suggest some sort of defensiveness over the role and peculiarities of the editorial work. If the borders between archives (and, as we will see below, libraries) and editions becomes blurred, so do actors such as editors, digital humanists, developers, librarians, interface designers, and so on, as seen in previous chapters. This merging of actors and redistribution of roles and competencies has determined the need for advocacy of the speciality of the editorial work. However, whatever distinction editors will try to enforce to defend their speciality, it seems unlikely that these will resist the transformative impact of the digital medium and of digital scholarship; it is worth here recalling, once again, the enlightening words of McCarty (1991, § 8): 'a new tool is not just a bigger lever and a more secure fulcrum, rather a new way of conceptualising the world'. In the same way that the introduction of new technologies such as, for instance, print, electricity, and the combustion engine, have deeply changed the way we live and our environment, computers and digital representations will do (and already are doing) just the same. Certainly there will still be distinctions between editors, editorial work and editions and other roles, however these might not reside where we now think they will. Continuity and transformation, inertia and dynamicity, which are the forces at stake when a new technology is introduced, will contribute to the shaping of future conceptualisations. At the present time, it seems that placing boundaries around the types of resources that can be produced might not be a productive way to look at the transformations introduced by the digital medium.

On a related but different perspective, the view that attempts to separate editions from archives is probably also connected to the understanding of documentary editions as a lower form of scholarship with respect to critical editions, an understanding that is forcefully rejected here. What these archives demonstrate is that looking for a clear-cut distinction between different types of resources may not be the wisest thing to do; the same applies to the distinction between critical and documentary editing: if it is enough to add a tool to one's archive to transform a documentary edition into a critical edition (Van Hulle and Nixon, 2013, Editorial Principles and Practice), then the distinction between the two forms of scholarship seems rather unsubstantial. It is not the tool that makes the difference, obviously, but this example underlines how the two endeavours are based on very similar scholarship.

Digital resources, and in particular those that have long lives, tend to evolve from an initial idea into something else. The case of the *Whitman Archive* as described by Price is particularly relevant in this sense. Born as a way to complement the missing parts of the printed edition of Whitman, it has grown to become a live laboratory (a 'workshop') of scholarship, with wide-ranging goals and achievements. The same can be said for the *Beckett Digital Manuscript Project*, the experimental and engaging solutions found to present the various versions of Beckett's drafts represents perhaps the most accomplished attempt yet published to account for the genesis of the works of a modern author. In the same way that distinctions such as diplomatic, semi-diplomatic, and critical are getting blurred in the digital environment, archives and editions are converging too.

A similar argument can be made around the concept of the 'digital library'. According to John Lavagnino, a digital library is everything digital editors do not want their work to be confused with, since it does not involve any analysis and 'can be created by workers who have no special knowledge of the material' (Lavagnino 2009, pp. 63-64). In his assessment, he probably has in mind endeavours such as large scale digitization, such as Google Books; however, when examining the *Perseus Digital Library*,[195] the 'editor in chief' of which is Gregory Crane, it is easy to realize that, once again, labels and terminologies can be deceiving. The *Perseus Digital Library* has produced an ever-growing collection of Classical Greek and Latin works with editorial rigour. While the project does not produce new editions based on primary sources, the editorial work required for the digitization and re-mediation of authoritative editions can hardly be achieved without analysis, knowledge and expertise in the subject area. In fact, not only are most of the current and past staff respected Classics scholars, but the library itself has become a place of critical and theoretical reflection (Crane, 2008; Crane 2010). Another example is offered by *Biblioteca Italiana* ('Italian Library'),[196] a complex project the aim and purpose of which are somewhere in between a digital library, a publishing house and a research hub. The texts available are considered new editions with respect to the printed version from which they derive, as declared by their digital frontispieces, and this is because they have been manually corrected from the OCR first, then analyzed and encoded in TEI, alongside reliable texts are offered accurate sets of metadata, which are also available for download. The main catalogue offers 2,800 texts (as of November 2013), which have been produced in more or less ten years. It is clear here that qualitative considerations precede quantitative ones; the project is also complemented by special collections and born-digital editions. A similar model is followed

[195] Cf. <http://www.perseus.tufts.edu/hopper/>.

[196] Cf. <http://www.bibliotecaitaliana.it/>.

by the *Bibliothèque Virtuelle des Humanistes*, based at the University of Tours.[197] Yet, as demonstrated by Lavagnino, the term 'digital library' has some strong connotations, some of which are quite undesirable.

The confusion in terminology reflects a problem of definition too: before a name can be found, it is first necessary to understand what that name will designate. Scholars have tried in various ways to define what a digital scholarly edition is, but the question is still open. As seen in Chapter 1, attempts to define these resources go hand in hand with theoretical assumptions, and these attempted definitions are meant to be as exclusive as they are inclusive. According to Patrick Sahle (2008) a digital scholarly edition is a 'critical representation of historical documents' which cannot be printed without loss of information. He then defines the components of his definition:

1. 'historical documents': editing is concerned with documents which are already there. In this wide sense of 'historical' the definition includes documents relevant for all subjects, history as well as literature or philosophy. Scholarly editing goes back to and starts from existing documents. To edit (to publish) a new document (which does not refer to something preexisting) is not scholarly editing.
2. 'representation': covers (abstract) representation as well as presentation (reproduction). As I use to say: transmedialization (representation by data) and medialization (presentation by media). Publishing descriptive data (e.g. metadata) without reproduction is not critical editing. A catalogue, a database, a calendar is not an edition.
3. 'critical / scholarly' (erschlieend): reproduction of documents without critical examination is not scholarly editing. A facsimile is not a scholarly edition.

From this definition we then can deduce that the definition of digital scholarly editions has to deal with the object to be considered ('historical documents'), a scholarly evaluation of this object, and the representation of the object in a way that differs from the original document. As for the medium, the fact that it cannot be printed without loss of information references the interactivity and to the dynamicity of digital editions. In another contribution Sahle identifies the defining character of digital scholarly editions in what he calls 'transmedialisation', namely the possibility of being separated from its own medium: 'the shift from the edition as a media product to the edition as a modeled information resource that can be presented in media but is rather the abstract representation of

[197] Cf. <http://www.bvh.univ-tours.fr/>.

knowledge' (2012). To summarize, digital scholarly editions are 'critical, abstract statements about historical documents which may be embodied by a digital, dynamic representation'. This definition, which is liberally derived from Sahle's statements, has the advantage of being synthetic and effective, but leave out two of the characteristics considered essential by Price: the open-endedness and the collaborative nature of these resources.

Siemens et al. 2012b, rather than defining what digital editions are, classifies them as being 'hypertextual editions' or 'dynamic editions' or both. In the case of hypertextual editions the emphasis is on the way readers can jump from one place to another within the edition itself, while 'dynamic' underlines the interactivity, the processing, and the exploitation of text as data, requiring that the edited texts are offered alongside tools to handle them. The scholarly intentionality of the editors is taken for granted. The definition that one can deduce from this classification is based on functionalities rather than on the nature of the object itself; that is to say a digital scholarly edition is defined by what you can do with it, more than by what it is. A scholarly digital edition contains a critically edited text which can be navigated in a non-linear way and can be explored by the means of sophisticated digital tools.

Shillingsburg's 'knowledge site' does not have a definition easy to summarize, but it is proposed as being 'constructed modularly and contributed to by a "village" of scholars' (2006, p. 97), and as an open-ended list of content and functions (pp. 101-102). Again, the scholarly discourse at its foundation is taken for granted (it is actually the 'point' of it), and the definition is focused on the process required to create the edition.

So far the derived definitions of what we are for the moment calling a digital scholarly edition have been focused on the object of editing (historical documents), the editorial statements that represent them as different from the original, the functionalities they have, and the process necessary to produce them. The combination of these 'definitions' is starting to delineate the kind of definition we are looking for.

Price discarded the use of the word 'edition' to define such complex resources because he felt that the term was too heavily connotated with the relics of printed editions. Indeed, the expectations in methodology and functionalities used to create printed editions are very different from those associated with digital editions. In particular, as seen in previous chapters, the expectation of stability and completeness is a defining feature of printed editions; on the other hand digital editions are characterized by open-endedness and changeability. Formats are stable in print, and mutable and unexpected in digital. However, in spite of the many differences noted in the way the edition is produced, conceived and

consumed, the main purpose remains the same: to make a scholarly argument about a historical document or a group of historical documents, in relation to their authorship, readership, circulation and reception. 'Edition', therefore, seems still a good way to define such resources.

Independently from the name we can attribute to it, a digital scholarly edition can then be defined as an interpretative representation of historical documents which encompass a combination of primary sources surrogates, edited text(s), and tools to exploit them. Such tools may have been developed specifically for the edition itself or pulled from elsewhere, or both. It may be open, encouraging the contribution of the users and readers in many ways, or closed, representing an editorial statement of a given scholar or group of scholars.

What tells digital editions apart from digital archives and digital libraries? Not much, as previously discussed. In fact, it is very clear how both archives and libraries make arguments (and many of them scholarly arguments too): they are all interpretative, they offer tools for the manipulation of the texts they contain, and they may be closed or open-ended. The fact is that, in the moment that historical documents are remediated and represented digitally a certain level of interpretation and a distance from the 'real thing' is inevitable. Furthermore, the selection of which texts to include, which editions to choose and how to organize them are all intellectual decisions that respond to an interpretation of a particular segment of the reality. The digital medium blurs traditional distinctions because all materialities become eventually the same: a stream of bytes variously organized by human beings via software and hardware. And as this operation is done by humans who at very least have to choose what to remediate and what not, operating discretionally and interpretatively, it becomes increasingly difficult to find the one thing, the one characteristic that tells different endeavours apart.

Perhaps the only safe statement one can make is that different representational resources (editions, archives, libraries), independently from the labels under which they are presented, offer different levels of scholarly engagement and surplus where 'surplus' here is used to translate Patrick Sahle's *Erschließung*. He defines it as that which 'encompasses any activity that increases the amount of information concerning a specific object and thus makes it more accessible'. The shift between archives, libraries and editions is subtle and happens along a continuous line where minimum and maximum scholarly engagements are distributed. The difference between these resources is therefore quantitative: the more scholarly argument the more likely there is to be an edition. However, the quantitative definition becomes also qualitative in the moment that scholarly engagement corresponds to an increased level of reliability, accuracy and, generally, quality of the editorial product.

While the distinctions are blurred, they still hold from a scholarly point of view. Nevertheless, no reliable criteria but the judgement of the scholar-user are offered as a means for classifying digital scholarly editions as distinguished from other text vessels.

9.1 A Glance Towards the Future

The difficulty in classifying and distinguishing digital scholarly editions as a genre is another factor in the increasing uneasiness of the scholarly community toward these new products. Furthermore, such difficulties and indistinctions are only destined to increase. Deegan and Sutherland (2009, pp. 59-88) undertake a very thorough and deeply critical analysis of digital editions, demystifying the bold optimism over the new medium and warning of the difficulties that pave the transition to the digital medium. They also maintain that the end result cannot be taken for granted as 'we are in a transitional stage as far as electronic technology is concerned. In consequence, both print and electronic editorial method are in flux' (p.76). They also argue that 'in the future we will also see an intensive merging of different media – print, audio, visual, movie clips – within the digital space, and as a consequence a blurring of editing into other kinds of remarking or adaptation as we discover further combination for text objects and text processes' (pp. 78-79). In other words, if you think it is hard now, wait and see the future. What is the future of scholarly editions then? Is it digital? Is it in video? Is it in print? The answer is bound to be 'all of the above', for at least some time.

Making predictions about the future is always difficult, and in particular in a world where technological changes happen at an increasing (and perhaps alarming) rate; nevertheless some current emerging trends seem to suggest a few lines of development.

1. **Digital as inevitable**. There is no going back: a future where the Humanities give up using computers seems highly unlikely. The past few years have shown an increasing integration of the digital into people's lives. The impact of technology on culture, education, politics, shopping, dating, travelling, eating, for instance, is hard to overestimate and is growing on a daily basis. In particular, the impact of computers on research, management and storage of knowledge, delivery and dissemination of scholarly outputs is truly transformational. Therefore imagining a future where editors decide to give up digital editing and produce only print material seems highly improbable. Perhaps the adoption of digital editions may be slower that some editors

would have expected (Robinson, 2003 and 2005), but the point here is not if we will be using digital editions or printed editions, but which type of digital, because digital it will be. But will it be in the shape of highly sophisticated editions such the one described in this book, or PDF versions of printed editions? In other words, will the inevitable switch to digital translate into a full remediation of printed scholarship?

2. **Digital as e-print**: today most digital textual objects look like printed objects. In an attempt to improve their acceptability, eBooks and eBook readers try to mimic the shape and functionalities offered by printed books. A survey conducted between 2007 and 2009 by JISC has revealed that the adoption of digital resources in Higher Education mostly means the use of PDFs of texts books and articles, with almost no concession to multimediality, interactivity and dynamic content. The same risks becoming true also for scholarly editions: the Oxford Scholarly Editions Online project, for instance, offers access to many scholarly editions published by Oxford University Press in the form of paginated PDFs. A similar route has been undertaken by Cambridge University Press and other publishers. Again, the page paradigm seems too strong to be fought. This tendency is quite a risky one, as it ignores most of the digital research deployed in the past twenty years in favour of a more reassuring, quotable by the page, print-alike skeuomorphism.[198] If this line of digitisation prevails, the humanities will find themselves cut off from the real potentials offered by the new medium. However, it is also possible that, as happened in the past with the introduction of print, for instance, once the new medium has established itself more stably within the academic community and new forms of scholarship have found the recognition that is still missing today, the reliance on the page model will likely cease to be so compelling. But for this to happen, new easy tools that can fit better within the editorial workflow have to be envisaged.

3. **Digital as hybrid**: print editions are accompanied by a digital version. This is already offered by most non-academic publishers who distribute their products as print and eBook, often at the same price, to suit most needs. The same has accompanied the evolution of music supports from vinyl, to tape, to CD, to Mp3 to online, with previous formats available for collectors and people with special professional needs and new formats gradually taking over in terms of the number of copies produced and sold. For scholarly editions we may postulate a situation where of any given published edition a

[198] Skeumorphism is the characteristic of a derivative object to imitate the real-like object it derives from. See Chapter 7, pp. 00.

decreasing number of printed copies will be sold, until the paper support will be either abandoned or maintained as luxury or storage format, to the point that it will not the digital format that will complement the print, but the print that will be offered as niche alternative to digital publications.

These possible lines of development seem more or less to suggest that the editorial work and workflow undertaken by digital editors in the past twenty years, as described by this contribution, will not be followed in the same way by the next generation of scholars for which digital editing will mean something different. This is indeed the risk and the challenge that digital editing is facing today. As Robinson (2013a) has reminded us, scholarly editions are expensive, only developable within specialized centres of research by healthy grant holders; they are difficult to support and sustain in the medium term, let alone the long term. However, before declaring the battle lost, perhaps some reflection on the significance of research in the humanities and digital humanities may be helpful here. Generally speaking, research in textual scholarship has always been interpreted as the production of a new version (or text) of a given work. The way it was delivered was only an external and instrumental preoccupation. Another way of putting it is to say that research means knowing in advance the type of research outcome that will be found at the end of the research itself. In the sciences, however, this is not always the case. The result of a research could be something unexpected that has no direct evident application.

It could be helpful here to introduce a distinction between digital editing as a research activity and digital editing as a means to produce digital editions. While the former can be highly speculative, the latter aims to present a product to the public in the most appealing way. Digital editing as research may still produce digital editions, but such editions may be considered just one of the outcomes of the research, the embodiment of a new editorial theory or model, and not its most important outcome. As Nowviskie maintains, digital editing as research may be evaluated 'as a process, not a product' (2011, p. 170), to which we may add the specification 'not *only* as a product'. Of course there is also a research component in digital editing as a means to producing a digital edition, but this lies mostly on the textual side, rather than in the digital innovation. Therefore it could be useful to consider the two activities as different, more or less in the same way that blue-sky research differs from applied research. According to McCarty, 'digital humanities have brought experiment

into the humanities';[199] the same could be said about digital editing: it has brought experiment into textual scholarship. The digital editions produced so far are both products aimed at a more or less specific public, and experiments, a way to test what is possible within the new medium and to establish new ways for scholarship. Jerome McGann significantly describe his Rossetti Archive as an hybrid enterprise, a laboratory and an edition:

> I undertook the project partly as a laboratory experiment to explore the capabilities of digital technology and partly to create a scholarly edition of Rossetti's work. As a laboratory experiment the project was a remarkable experience—a clear educational success, I should say. […] On the other hand, if the archive is judged strictly as a scholarly edition, the jury is still out' (2010, p. 11).

So while in the future the production of digital editions may take the easy route of publishing products which look remarkably like printed editions, it is essential that we maintain and reinforce a parallel strand of more daring digital editions which have to work as laboratories of textual scholarship. It is from this second strand of research that real, sustainable innovation is to be expected. If conceived in this way, it makes sense for such activity to take place within specialized centres and with the support of research grants. The quest for sustainable solutions will also be part of the deal.

Yet, in order to make this vision a reality a few essential changes will have to take place. First, a change of mentality of both editors and funding bodies is required. As pointed out by McCarty, 'failure is our most important product',[200] but until the community of textual scholars is ready to embrace the possibility of failure, it will be hard to find the courage to develop truly innovative products: the fear of failure will be too great. But the change has to happen also at the level of funding bodies, where the expectation of the delivery of finished and polished products still dominates. Discovering that one approach does not work or does not deliver what was hoped at the beginning is as an important research outcome, as important as finding a method that does work. As Prescott comments: 'The risk now is that, as digital technologies become commonplace in the academy, we will assume that there is only one way of doing things, a series of methods and standards which have to be shared and

---

[199] [Humanist] 27.444 between STEM and the human sciences? http://www.dhhumanist.org/cgi-bin/archive/archive_msg.cgi?file=/Humanist.vol27.txt&msgnum=443&start=59937&end=60030

[200] As cited by Prescott (2013) in his account of McCarty's Busa Award lecture delivered at the Digital Humanites conference in Lincoln, Nebraska, 2013 (cf. <http://digitalriffs.blogspot.co.uk/2013/07/riffs-on-mccarty.html>).

disseminated. The result will be an evisceration of the possibilities of the digital humanities'; he then calls for 'more experiments, more tinkering, more just trying out' (2013). Digital editing, or at least a part of it, should do just this: experiment with the new medium until its limits and possibilities are uncovered and explored to the end – if there is an end. Experimenting could take many forms, including the creations of prototypes and models, reflecting on or shaping new theories, advancing knowledge and understanding. The European network DiXiT promise to do exactly this (Digital Scholarly Editing Initial Training Network):[201] the network aims at blue-sky research on an international scale for 5 years, the impact of which could be truly transformational. This is also the belief of Tara Andrews (2013) who calls for the conception and the practice of a new digital philology, heavily computational and without pre-conceptions over heuristics and over what can and cannot be done. Her vision calls for an engagement without compromises of the editor with computational methodologies which may be too challenging, as the technical threshold is perhaps too high. However if we locate such a call within the framework discussed above where two strands of digital editing can be envisaged, then her proposal becomes much more achievable. If we maintain the development of a line of experimental textual scholarship, able to engage with sophisticated computational models searching for new solutions, then such a technical threshold becomes not only adequate, but inevitable.

On the other hand, though, a call for ready-to-use, user-friendly tools and standards that will enable digital editing as a means for producing editions in a simplified environment is compelling too, if not even more so. And this is not just in order to convince our colleagues of the advantages offered by computers as assumed by Prescott (2013), or to demonstrate the relevance of our work (a relevance that should not be demonstrated by the means of tools, but by scholarship); it is mainly because we need to find ways of easing young scholars into digital editing. It should be possible for a Masters student, or a PhD student to edit text(s) digitally and to find enough support to make these edited texts available somewhere. At the moment obstacles such as a lack of tools for manipulating the texts and a lack of publications infrastructures to showcase texts once they have been edited are the biggest barriers for the uptake of the digital methodology in editing; and without being able to support young scholars, the existence itself of digital editing as an innovative, transformational discipline is at risk.

9.2. Ethics of Digital Editions

[201] Cf. <http://dixit.uni-koeln.de>.

This last point regarding the sustainability of digital editing in terms of support to scholars in production is a crucial component of what could be called the ethics of digital editing. The way digital scholarly editions in particular, and digital humanities resources in general, have been developed so far have been defined as project-based and team-based. And indeed funding has coalesced around research centres which then have the capability of hiring developers and research stuff for the (short) duration of the project. The lack of disciplinary infrastructure and in particular established teaching curricula in digital humanities able to ease project staff into an established academic career have aggravated a phenomenon sadly well known within higher education, that is the proliferation of short-term positions. As projects are funded in most cases for one to three years, it is quite normal that groups of scholars and developers, mostly early careers people, are hired for the duration of the project only, leaving them with uncertainty for their future, forcing them to migrate from project to project, and often leaving halfway through the lifespan of the project, if a more permanent position becomes available.

In addition to the precariousness of the job market, the recognition of the essential contribution given by project teams is also problematic. Questions connected with the attribution of authorship have already been discussed in the previous chapter, but, as discussed by Nowviskie (2012), this is just one of the issues connected to the employment of highly skilled researchers whose opportunities for career development are quite thin. The realisation of how widespread such employment difficulties are and its impact on the ethics and development of the discipline has indeed prompted Bethany Nowviskie to start a movement called *alt-ac* (Alternative Academic careers) advocating for the recognition and improvement of the non standard career patterns so common within digital humanities projects.

> "#Alt-ac" is the neologism […] about "alternative academic" careers for humanities scholars. Here, "alternative" typically denotes neither adjunct teaching positions nor wholly non-academic jobs — about which, in comparison, advice is easy to find. Instead, we are examining positions within or around the academy and requiring deep understanding of humanities scholarship, but outside of the ranks of the tenured or tenure-track teaching faculty. Such roles are taken up by capable scholars who maintain a research and publication profile, or who bring their (often doctoral-level) methodological and theoretical training to bear on problem sets in the humanities (Nowviskie, 2011a).

The project is showcased by a MediaCommons[202] website, which hosts contributions by both established and young scholars, the latter often in alt-ac careers.[203]

In her introduction to the project's collection of essays, Nowviskie points out not only the ethical implications of such a situation, but also the economic impact of such a practice: 'if the academy cannot foster more appealing and fruitful options along the #alt-ac track,[204] we will have trained a generation of humanities experts, only to lose them'. The impact on projects of losing highly trained staff who understandably leave in the quest for more stable jobs and career development has not yet been quantified, but the suspicion is that this will result in a quite a hefty bill to pay. However, tenured positions are not always possible nor are necessarily the solution. In his contribution to the same initiative, Willard McCarty (2011) points out how tenure on the one hand 'shields a nascent department and its tenured members from the distracting and enervating demands for proof of usefulness and service', but on the other had creates rigid structures not suitable for a developing discipline such a Digital Humanities.

Career path and career development are indeed the field in which the battle for the recognition of digital scholarship is and will be fought in the next few years. During a presentation at King's College London (2006), Charlotte Rouché declared that Digital Humanities 'is a business for old people', meaning by this that the constraints of an academic career path are likely to penalize digital scholarship, a fact particularly problematic for early career researchers. In a special thematic section of the MLA journal *Profession* in 2011, made available in open access as an exception, a group of scholars advocate for the recognition of digital scholarship at many levels, but in particular for career progression. In addition MLA has distinguished itself for its early support of digital scholarship and for the provision of tools for the evaluation of digital scholarship. In 2012 it produced a set of *Guidelines for Evaluating Work in Digital Humanities and Digital Media*, including sections on how to appoint and promote staff and guidelines for candidates aspiring to improve their career paths by making the most of digital scholarship. Yet the understanding of the scholarship involved in producing a database or a set of XSLT is still not common knowledge. Schreibman *et al*. (2011) in introducing the special section of *Profession*,

---

[202] Funded in 2007 'MediaCommons is a community network for scholars, students, and practitioners in media studies, promoting exploration of new forms of publishing within the field. MediaCommons was founded with the support of the Institute for the Future of the Book, and with assistance from the National Endowment for the Humanities': <http://mediacommons.futureofthebook.org/>.

[203] Cf. <http://mediacommons.futureofthebook.org/alt-ac/>.

[204] Since 2012, the new Twitter hash tag for alt-ac is #altac.

remarked how at a workshop on the evaluation of digital scholarship 'evaluators who were not themselves creators of digital scholarship found it extremely challenging to see how technical decisions are theoretically informed and constitute research-supported argumentation' (p. 128). Then they ask '[d]oes someone need to write and publish arguments in peer-reviewed journals about their databases or archives for promotion and tenure, or do those databases and archives themselves, with accompanying descriptions, count as research?' (p. 129). The question is fair: while as a community we can complain about prejudiced evaluation of our research outcomes, it is also our responsibility to find ways in which such scholarship can become more accessible to the non initiated. Schreibman *et al*. (2012) also report on a special difficulty in career development for digital editors, namely the fact that textual editing is also commonly labelled as a service, instrumental to scholarship more than scholarship itself (p. 127-28).

All of this is certainly a concern, yet some tentative signs of a fairer evaluation of digital scholarship are emerging from different sides. The MLA's statements and resources are undoubtedly one of these. The possibility of submitting digital resources for the Research Excellence Framework, the mechanism for the valuation of research in the United Kingdom, is another one. In the future the inevitable uptake of digital scholarship will probably radically change this situation, and hopefully for the better. However, it is in advocating and making our work as traceable and assessable as possible that this will be achieved. The establishment of good practices, and transparent ways of evaluating and manifesting scholarship will help to show how, in spite of the deep changes determined by the digital medium, the rigour and the hermeneutics of editing remains intact, if not enhanced by the new medium.

To conclude, having discussed the many facets of digital scholarly editing it is legitimate to ask two questions: is digital scholarly editing something truly new or just the same activity in a new medium? And: is it worth it? To respond to the first question, it is certainly evident that in digital scholarly editing there are at least as many elements of continuity as there are of innovation. However the innovative elements are truly transformative, involving as they do formats, methods, roles, heuristics and hermeneutics of editing. Yet the purpose of editing remains the same, that is, the presentation of historical documents in ways that are meaningful for a group of users according to a documented and sharable methodology. For this, we may conclude – though reserving the right to return to this point in a few years time – that digital scholarly editing is a radical evolution (but not

revolution) of print-based editing, as if in a Darwinian pattern of evolution a few steps have been jumped all at once.

As for the second question (are they worth it?), the answer cannot but be: absolutely! We have just made the jump and we do not really know how to cope with the changed circumstances, yet there are so many research paths that can be pursued that the worthiness of the endeavour cannot be doubted. Surely there is a rich agenda for the digital editor while new balances and procedures able to meet scholarly expectations have to be established, yet the challenge ahead is exciting and invigorating.

Bibliography

Aarseth, Espen, *Cybertext. Perspectives on Ergodic Literature Cybertext. Perspectives on Ergodic Literature* (Baltimore and London: John Hopkins University Press, 1997).

Amazon, *Amazon.com now Selling More Kindle Books than Print Books (Press release)*, (2011), available at <http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-newsArticle&ID=1565581&highlight>.

André, Julie and Pierazzo, Elena, 'Le Codage en TEI des Brouillons de Proust: Vers l'Édition Numérique', *Genesis*, 36/13 (2013): 155-161.

Andrews, Tara L., *Excerpts from the "Chronicle" of Matthew of Edessa* (2012), available at <http://byzantini.st/ChronicleME/>.

Andrews, Tara L., 'The Third Way. Philology and Critical Edition in the Digital Age', *Variants*, 10 (2013): 61-76.

Andrews, Tara L. and Macé, Caroline, 'Beyond the Tree of Texts: Building an Empirical Model of Scribal Variation through Graph Analysis of Texts and Stemmata', *Literary and Linguistic Computin*g, 28/4 (2013): 504-521.

Arbuckle, Alyssa, 'Considering The Waste Land for iPad and Weird Fiction as models for the public digital edition', *Digital Studies/ Le champ numérique* (2014), available at <http://www.digitalstudies.org/ojs/index.php/digital_studies/article/view/264>.

Austen, Jane, *Minor works now First Collected and Ed. from the Manuscripts by R. W. Chapman, with Illus. from Contemporary Sources*, in Robert William Chapman (ed.) (London: Oxford University Press, 1954).

Austen, Jane, *Sir Charles Grandison*, in Brian Southam (ed.), *Sir Charles Grandison* (Oxford: Clarendon Press, 1980).

Bański, Piotr, 'Why TEI Stand-off Annotation Doesn't Quite Work: and Why you Might Want to Use it Nevertheless', *Proceedings of Balisage: The Markup Conference 2010*, 5 (2010), available at <doi:10.4242/BalisageVol5.Banski01>.

Barthes, Roland, 'The Death of the Author', in *Image-Music-Text* (New York: Hill and Wang, 1977): 142-148.

Beckett, Samuel, *Stirrings Still / Soubresauts and Comment dire / What is the Word: an*

Eectronic Genetic Edition, Dirk Van Hulle and Vincent Neyt (eds), (Brussels: University Press Antwerp (ASP/UPA), 2011), available at <http://www.beckettarchive.org>.

Bédier, Joseph, 'La Tradition M anuscrite du "Lai de l'Ombre": Réflexions sur l'Art d'Éditer les Anciens Textes', *Romania*, 54 (1928): 161--96, 321--56.

Benjamin, Walter B.S., 'L'Œuvre d'Art à l'Époque de sa Reproduction Méchanisée', *Zeitschrift für Sozialforschung*, 5 (1936): 40-68.

Berrie, Phill, Eggert, Paul, Tiffin, Chris, and Barwell, Graham, 'Authenticating Electronic Editions', in Lou Burnard, Katherine O'Brien O'Keeffe, and John Unsworth (eds), *Electronic Textual Editing* (New York: Modern Language Association of America, 2006): 269-276.

Bischoff, Bernhard, *Latin Palaeography: Antiquity and the Middle Ages*, translated by Dáibhí Ó Cróinín and David Ganz (eds), (Cambridge: Cambridge University Press, 1990).

Bhaskar, Michael, *The Content Machine. Towards a Theory of Publishing from the Printing Press to the Digital Network* (New York: Anthem Press, 2013).

Bodard, Gabriel, 'The Inscriptions of Aphrodisias as Electronic Publication: A User's Perspective and a Proposed Paradigm', *Digital Medievalist*, 4 (2008), available at <http://www.digitalmedievalist.org/journal/4/bodard/>.

Bodard, Gabriel and Garcés, Juan, 'Open Source Critical Editions: A Rationale', in M. Deegan and K. Sutherland (eds), *Text Editing, Print and the Digital World* (Aldershot: Ashgate, 2009): 83-98.

Bolgar, Robert Ralph, *The Classical Heritage and its Beneficiaries* (Cambridge: Cambridge University Press, 1954).

Boot, Peter, 'Mesotext. Framing and Exploring Annotations'. In Stronks E. and Boot P. (eds), *Learned Love. Proceedings of the Emblem Project Utrecht Conference on Dutch Love Emblems and the Internet (November 2006)* (The Hague: DANS - Data Archiving and Networked Services, 2007): 211-225.

Boot, Peter, *Mesotext: Digitised Emblems, Modelled Annotations and Humanities Scholarship* (Amsterdam: Pallas Publications, 2009).

Bordalejo, Barbara, 'The Encoding System', in Prue Shaw (ed.), *The Commedia of Dante*

Alighieri: A Digital Edition. (Saskatoon: Scholarly Digital Editions, 2010): http://www.sd-editions.com/AnaAdditional/CommediaEx/CommediaExhome.html.

Bordalejo, Barbara, 'What is meant by editing in the phrase social editing?', *Presentation at Social, Digital Scholarly Editing Conference*. University of Saskatchewan, Saskatoon. (2013), available at < http://www.academia.edu/4125893/What_is_Meant_by_Editing_in_the_Phrase_Social_Editing>.

Bowers, Fredson T., *Principles of Bibliographical Description* (Princeton: Princeton University Press, 1949).

Bowers, Fredson T., 'Authorial Intention and the Editorial Problems', *Text*, 5 (1991): 49-61.

Bozzi, Andrea, 'Electronic Publishing and Computational Philology', in Macé Caroline, Baret Philippe, Andrea Bozzi, and Cignoni Laura (eds), *The Evolution of Texts: Confronting Stemmatological and Genetical Methods*, XXIV-XXV (Pisa, Roma: Istituti Editorali e Poligrafici Internazionali, 2006): 3-24.

Brown, Adrian, *Practical Digital Preservation: A How-to Guide for Organizations of Any Size* (London: Facet Publishing, 2013).

Bree, Linda and McLaverty, James, 'The Cambridge Edition of the Works of Jonathan Swift and the Future of the Scholarly Edition', in Marilyn Deegan and Kathryn Sutherland (eds), *Text Editing, Print and the Digital World* (Aldershot: Ashgate, 2009): 127--136.

Broude, Ronald, 'Musical Works, Musical Texts, and Musical Editions. A Brief Overview', *Scholarly Editing*, 33 (2012), available at <http://scholarlyediting.org/2012/essays/essay.broude.html>.

Brüning, Gerrit, Henzel, Katrin, and Pravida, Dietmar, 'Multiple Encoding in Genetic Editions: The Case of "Faust"', *Journal of the Text Encoding Initiative*/4 (2013), available at <http://jtei.revues.org/697>.

Bryant, John, *The Fluid Text: a Theory of Revision and Editing for Book and Screen* (Ann Arbor: University of Michigan Press, 2002).

Burdick, Anna, Drucker, Johanna, Lunenfeld, Peter, Presner, Todd, and Schnapp, Jeffrey, *Digital_Humanities* (Cambridge (MA): The MIT Press, 2012).

Burghart, Marjorie and Rehbein, Malte, 'The Present and Future of the TEI Community for

Manuscript Encoding', *Journal of the Text Encoding Initiative*/2 (2012), available at <http://jtei.revues.org/372>.

Buzzetti, Dino, 'Digital Representation and the Text Model', *New Literary History*, 33/1 (2002): 61-88.

Buzzetti, Dino, 'Digital Editions and Text Processing' (Aldershot: Ashgate, 2009): 45-61.

Buzzetti, Dino and McGann, Jerome, 'Critical Editing in a Digital Horizon', in Lou Burnard, Katherine O'Brien O'Keeffe, and John Unsworth (eds), *Electronic Textual Editing* (New York: Modern Language Association of America, 2006): 53-73.

Carpenter, David, Wilkinson, Louise, Crook, David, and Short, Harold, *Henry III Fine Rolls Project* (London: King's College London, 2009-2013), available at <http://www.finerollshenry3.org.uk/home.html>.

Cartlidge, Neil, 'The Canterbury Tales and Cladistics', *Neuphilologische Mitteilungen*, 102 (2001): 135-50.

Caton, Paul, 'Lost in Transcription: Types, Tokens, and Modality in Document Representation', in *Digital Humanites 2009. Conference Abstracts. University of Maryland, June 22-25* (2009): 80-82.

Caton, Paul, 'On the Term "Text" in Digital Humanities', *Literary and Linguistic Computing*, 28/2 (2013): 209-220.

Caton, Paul, 'Pure Transcriptional Markup', in *Digital Humanities 2013. Conference Abstracts. University of Nebraska-Lincoln, July 16-19* (2013): 140-141.

Causer, Tim, Tonra, Justin, and Wallace, Valeria, 'Transcription Maximized; Expense Minimized? Crowdsourcing and Editing The Collected Works of Jeremy Bentham', *Literary and Linguistic Computing*, 27/2 (2010): 119-137.

Cayless, Hugh A., 'Ktêma es aiei: Digital Permanence from an Ancient Perspective', in Gabriel Bodard and Simon Mahony (eds), *Digital Research in the Study of Classical Antiquity* (Aldershot: Ashgate, 2010): 139-150.

Cerquiglini, Bernard, *In the Praise of the Variant*, trans. by Betsy Wing (Baltimore and London: John Hopkins University Press, 1999).

Chaucer, Geoffrey, *The Wife of Bath's Prologue on CD-ROM*, Peter M.W. Robinson (ed.), (Cambridge: Cambridge University Press, 1996).

Chaucer, Geoffrey, *Chaucer: The General Prologue on CD-ROM*, Elizabeth Solopova (ed.), (Cambridge: Cambridge University Press, 2000).

CIDOC CRM, *FRBRoo Introduction* (2009), available at <http://www.cidoc-crm.org/frbr_inro.html>.

Ciula, Arianna and Lopez, Tamara, 'Reflecting on a Dual Publication: "Henry III Fine Rolls" Print and Web', *Literary and Linguistic Computing*, 24/2 (2009): 129-141.

Contini, Gianfranco, *Esercizî di lettura. Nuova edizione aumentata di "Un anno di letteratura"* (Torino, Einaudi), 1974.

Contini, Gianfranco, *Breviario di Ecdotica* (Torino: Einaudi, 1986).

Coombs, James H., Renear, Allen H., and DeRose, Stephen J., 'Markup Systems and the Future of Scholarly Text Processing', *Communications of the ACM*, 30/11 (1987): 933-47.

Coseriu, Eugenio, 'Los Conceptos de 'Dialecto', 'Nivel' y 'Estilo de Lengua' y el Sentido Propio de la Dialectologia', *Lingüística española actual*, 3/1 (1981): 1-32.

Crane, Gregory, 'Give us Editors! Re-inventing the Edition and Re-thinking the Humanities', in Jerome McGann (ed.), *Online Humanities Scholarship: The Shape of Things to Come* (Houston: Rice University Press, 2010): 81-97.

Crane, Gregory, Bamman, David, and Jones, Alison, 'ePhilology: When the Books Talk to their Readers', in Ray Siemens and Susan Schreibman (eds), *A Companion to Digital Literary Studies* (Oxford: Blackwell, 2008): 29-64.

Cummings, James, 'The Text Encoding Initiative and the Study of Literature', in Ray Siemens and Susan Schreibman (eds), *A Companion to Digital Literary Studies* (Oxford: Blackwell, 2008): 451-476.

Cummings, James, 'Converting Saint Paul: A new TEI P5 edition of "The Conversion of Saint Paul" Using Stand-off Methodology', *Literary and Linguistic Computing*, 24/3 (2009): 307-317.

Czmiel, Alexander, 'Editio ex Machina: Digital Scholarly Editions Out of the Box, *Digital Humanities 2008. Conference Abstracts. University of Oulu, July 16-22* (Oulu: University of Oulu, 2008): 101-102.

Daisne, Johan, *De Trein der Traagheid*, Xavier Roelens and Edward Vanhoutte (eds.) and

with Ron Van der Branden (Ghent: Centrum voor Teksteditie en Bronnenstudie, 2012).

Daston, Lorraine and Galison, Peter, *Objectivity* (New York: Zone Books, 2010).

De Angelis, Alessandtro, 'Le "Dita Separate": un'Ipotesi Lessicale e una Sintattica per l'Indovinello Veronese', *Zeitschrift für romanische Philologie*, 119/1 (2003): 107-133.

Deegan, Marilyn, 'Collection and Preservation of Electronic Edtions', in Lou Burnard, Katherine O'Brien O'Keeffe, and John Unsworth (eds), *Electronic Textual Editing* (New York: Modern Language Association of America, 2006): 358-370.

Deegan, Marilyn and Sutherland, Kathryn, *Transferred Illusions. Digital Technology and the Forms of Print* (Aldershot: Ashgate, 2009).

Dekker, Ronald H. and Middell, Gregor, 'Computer-supported Collation with CollateX: Managing Textual Variance in an Environment with Varying Requirements', in *Supporting Digital Humanities: Conference Proceedings* (Copenhagen 2011), available at <http://crdo.up.univ-aix.fr/SLDRdata/doc/show/copenhagen/SDH-2011/submissions/sdh2011_submission_54.pdf>.

DeRose, Stephen J., Durand, David G., Mylonas, Elli, and Renear, Allen H., 'What is Text, Really', *Journal of Computing in Higher Education*, I/2 (1990): 3-26.

Dionisotti, Carlo, *Geografia e Storia della Letteratura Italiana* (Torino: Einaudi, 1967).

Doni, Anton Francesco, *Inferni. Libro Secondo de' Mondi*, in Francesco Sberlati (ed.), (Bologna: Commissione per i testi di lingua, 1998).

Dow, Sterling, *Conventions in Editing: A Suggested Reformulation of the Leiden System* (Durham: Duke University Press, 1969).

Driscoll, Matthew J., 'The Words on the Page: Thoughts on Philology, Old and New'. *Creating the Medieval Saga: Versions, Variability, and Editorial Interpretations of Old Norse Saga Literature*, ed. Judy Quinn & Emily Lethbridge (Odense: University Press of Southern Denmark, 2010), pp. 85-102, available at http://www.staff.hum.ku.dk/mjd/words.html.

Driscoll, Mattew J., 'Levels of Transcription', in Lou Burnard, Katherine O'Brien O'Keeffe, and John Unsworth (eds), *Electronic Textual Editing* (New York: Modern Language Association of America, 2006).

Dunn, Stuart and Hedges, Mark, *Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research*, (London: King's College London, 2012), available at <http://crowds.cerch.kcl.ac.uk/wp-uploads/2012/12/Crowdsourcing-connected-communities.pdf>.

Eaves, Morris, Essick, Robert, Viscomi, Joseph, and with Shaw, William (eds), *The William Blake Archive* (Chapel Hill: University of North Carolina at Chapel Hill, 1996), available at <http://www.blakearchive.org/>.

Eggert, Paul, 'Text-encoding, Theories of the Text, and the "Work-Site"', *Literary and Linguistic Computing*, 20/4 (2005): 425-435.

Eggert, Paul, 'Document and Text: The "Life" of the Literary Work and the Capacities of Editing', *TEXT*, 7 (1994): 1-24.

Eggert Paul, *Securing the Past. Conservation in Art, Architecture and Literature* (Cambridge: Cambridge University Press, 2009a).

Eggert, Paul, 'The Reader-Oriented Scholarly Edition and the Work'. *Presentation at Social, Digital Scholarly Editing Conference*. University of Saskatchewan, Saskatoon. (2013).

Elliot, Tom, Bodard, Gabriel, Mylonas, Elli, Stoyanova, Simona, Tupman, Charlotte, and Vanderbilt, Scott, *EpiDoc Guidelines: Ancient documents in TEI XML* (2011-2013), available at <http://www.stoa.org/epidoc/gl/latest/index.html>.

Elsschot, Willem, *Achter de Schermen*, in Vincent Neyt Peter de Bruijn and Dirk Van Hulle (eds), *Achter de Schermen* (Antwerp and The Hague: Center for Manuscript Genetics (Universiteit Antwerpen), Huygens Instituut, 2007).

Farley, Abraham and Ellis, Henry (eds), *Domesday Book* ([London],1783-1816).

Faulhaber, Charles B., 'Textual Criticism in the 21st Century', *Romance Philology*, 45 (1991): 123-48.

Ferrer, Daniel, 'The Open Space of the Draft Page: James Joyce and Modern Manuscripts', in Georges Bornstein and Theresa Tinkle (eds), *The Iconic Page in Manuscript, Print and Digital Culture* (Ann Arbour: University of Michigan Press, 1998): 249-267.

Finneran, Richard J. (ed.), *The Literary Text in the Digital Age* (Ann Arbor: University of Michigan Press, 1996).

Fischer Franz, 'The Pluralistic Approach: The First Scholarly Edition of William of Auxerre's Treatise on Liturgy', *Computerphilologie*, 10 (2010): 151-168, available at <http://computerphilologie.digital-humanities.de/jg08/fischer.html>.

Fischer Franz, 'All Texts are Equal, but... Textual Plurality and the Critical Text in Digital Scholarly Editions', *Variants*, 10 (2013): 77-92.

Fitzpatrick, Kathleen, 'Peer Review, Judgment and Reading', *Profession* (2011): 196-201, available at <doi:10.1632/prof.2011.2011.1.196>.

Folsom, Ed and Price, Kenneth M. (eds), *The Walt Whitman Archive* (University of Nebraska--Lincoln: Center for Digital Research in the Humanities, 1995-), available at <http://www.whitmanarchive.org/>.

Flaubert, Gustave, *Hérodias*, Giovanni Bonaccorso *et al*. (eds), *Hérodias* (Paris: Nizet, 1991).

Neil Fraistat and Julia Flanders (eds), *The Cambridge Companion to Textual Scholarship* (Cambridge: Cambridge University Press, 2013).

Gibson, James J., *The Ecological Approach to Visual Perception* (Boston: Houghton-Miffin, 1979).

Greg, W. W., 'The Rationale of Copy-Text', *Studies in Bibliography*, 3 (1950--51): 19-36.

Grier, James, *The Critical Editing of Music: History, Method, and Practice* (Cambridge: Cambridge University Press, 1996).

Gabler, Hans Walter, 'The Primacy of the Document in Editing', *Ecdotica*, 4 (2007): 197-207.

Gabler, Hans Walter, *Argument into Design. Editions as a Sub-Species of the Printed Book* (London: School of Advanced Studies, 2008), available at <http://www.ies.sas.ac.uk/sites/default/files/files/Publications/Coffin%20lectures/HWG%20Coffin%202008.pdf>.

Gabler, Hans Walter, 'Beyond Author-Centricity in Scholarly Editing', *Journal of Early Modern Studies*, 1 (2012): 15-35.

Gilbert, Penny, 'Automatic Collation: A Technique for Medieval Texts', *Computers and the Humanities*, 7/3 (1973): 139-147.

Goytisolo, Juan, *Paisajes Después de la Batalla. Preliminares y Estudio de Crítica Genética*,

in Bénédicte Vauthier (ed.), *Paisajes Después de la Batalla* (Salamanca: Ediciones Universidad Salamanca, 2012).

Garcés, Juan and Dogan, Zeki Mustafa (eds), *Codex Sinaiticus Project* (London and Leipzig: British Library and Leipzig University Library, 2009), available at <http://www.codexsinaiticus.org>.

Greetham David C., *Textual Scholarship: An Introduction* (New York: Garland, 1994).

Greetham, David C., 'A History of Textual Scholarship', in Neil Fraistat and Julia Flanders (eds), *The Cambridge Companion to Textual Scholarship* (Cambridge: Cambridge University Press, 2013): 16-41.

Grésillon, Almuth, *Eléments de Critique Génétique: Lire les Manuscrits Modernes*. (Paris: Presses universitaires de France, 1994).

Hanna, Ralph, 'The Application of Thought to Textual Criticism in All Modes—with Apologies to A. E. Housman', *Studies in Bibliography*, 53 (2000): 163-72.

Hay, Louis, 'Critiques de la Critique Génétique', *Romanic Review* 3 (1995): 403-417.

Harvey, Ross, *Preserving Digital Materials* (Berlin: De Gruyter Saur, 2012).

Hillesund, Terje, 'Will E-books Change the World?', *First Monday Online*, 6/10 (2001), available at <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/891/800>.

Housman, A. E., 'The Application of Thought to Textual Criticism', *Proceedings of the Classical Association*, 18 (1921): 67-84.

Howe, Jeff, *Crowdsourcing: Why the Powers of the Crowd is Driving the Future of Business* (New York: Three Rivers Press, 2009).

Huitfeldt, Claus, 'Multi-dimensional Texts in a One-dimensional Medium', *Computers and the Humanities*, 28/4-5 (1994): 235-41.

Huitfeldt, Claus, 'Philosopy Case Study', in Lou Burnard, Katherine O'Brien O'Keeffe, and John Unsworth (eds), *Electronic Textual Editing* (New York: Modern Language Association of America, 2006): 181-196.

Huitfeldt, Claus and Sperberg-McQueen, C. Michael, 'What is Transcription?', *Literary and Linguistic Computing*, 23/3 (2008): 295-310.

Huitfeldt, Claus, Marcoux, Yves, and Sperberg-McQueen, C. Michael, 'Extension of the Type/Token Distinction to Document Structure', *Proceedings of Balisage: The Markup Conference 2010*, Balisage Series on Markup Technologies 5 (2010), available at doi:10.4242/BalisageVol5.Huitfeldt01

Hunter, Michael, *Editing Early Modern Texts: an Introduction to Principles and Practice* (New York: Palgrave Macmillan, 2007).

International Federation of Library Associations and Institutions (IFLA), *Functional Requirement for Bibliographic Records* (1998), available at <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>.

Jackson, William A., 'Some Limitations of Microfilm', *Papers of the Bibliographical Society of America*, 35 (1941): 281-288.

Jakobson, Roman, 'Closing Statement: Linguistics and Poetics', in Thomas Sebeok (ed.), *Style in Language* (Cambridge (MA): MIT Press, 1960): 350-77.

Jannidis, Fotis, 'TEI in a Crystal Ball', *Literary and Linguistic Computing*, 24/3 (2009): 253-265.

Jessop, Martyn, 'The Inhibition of Geographical Information in Digital Humanities Scholarship', *Literary and Linguistic Computing*, 23/1 (2008a): 39-50.

Jessop, Martyn, 'Digital Visualization as a Scholarly Activity', *Literary and Linguistic Computing*, 23/3 (2008b): 281-293.

JISC National eBook Observatory Project, *Scholarly E-books Usage and Information Seeking Behaviour: a Deep Log Analysis of MyiLibrary*, CIBER, Final Report (November 2009), available at <http://observatory.jiscebooks.org/reports/scholarly-e-books-usage-and-information-seeking-behaviour-a-deep-log-analysis-of-myilibrary/>.

Joyce, James, *Ulysses: A Critical and Synoptic Edition*, in Hans Walter Gabler (ed.), *Ulysses: A Critical and Synoptic Edition* (New York: Garland, 1984).

Jordan, Patrick W., *Designing Pleasurable Products: an Introduction to the New Human Factors* (London: Taylor & Francis e-Library, 2000).

Judah, Sam, 'What is Skeuomorphism?', *BBC News Magazine* (2013), available at <http://www.bbc.co.uk/news/magazine-22840833>.

Kamp, Georg, 'Digital Publishing in the Faculties of Arts and Humanities: Prospects, Chances, and Risks', *Poiesis & Praxis: International Journal of Technology Assessment and Ethics of Science*, 5/2 (2008): 77-80.

Kant, Immanuel, *Critique of Pure Reason*, translated by Norman Kemp, *Critique of Pure Reason* (New York: Saint Martin's Press, 1965).

Kichuk, Diana, 'Metamorphosis: Remediation on "Early English Books Online (EEBO)"', *Literary and Linguistic Computing*, 22/3 (2007): 291-303.

Kiernan, Kevin (ed.), *Electronic Beowulf* (London, Chicago: British Library and University of Chicago Press, 2011).

Kiernan, Kevin, 'Digital Facsimile in Editing', in Lou Burnard, Katherine O'Brien O'Keeffe, and John Unsworth (eds), *Electronic Textual Editing* (New York: Modern Language Association of America, 2006): 262-268.

Kirschenbaum, Matthew and Reside, Dough, 'Tracking the Changes. Textual Schoalrship and the Challange of the Born Digital', in Neil Fraistat and Julia Flanders (eds), *The Cambridge Companion to Textual Scholarship* (Cambridge: Cambridge University Press, 2013): 257-273.

Kline, Mary-Jo and Perdue, Susan H., *A Guide to Documentary Editing* (Charlottesville: University of Virginia Press, 2008).

Kralemann, Björn and Lattmann, Claas, 'Models as Icons: Modeling Models in the Semiotic Framework of Peirce's Theory of Signs', *Synthese*, 190 (2013): 3397-3420.

Lachmann, Karl, *Caroli Lachmanni in T. Lucretii Cari De rerum natura libros commentarius iterum editus* (New York : Garland, 1979).

Lancashire, Ian, 'Working with Texts', *Paper delivered at the IBM Academic Computing Conference* (Anaheim, CA, 1989).

Lavagnino, John, 'Access', *Literary and Linguistic Computing*, 24/1 (2009): 63-76.

Lavoie, Brian and Gartner, Richard, 'Preservation Metadata', *DPC Technology Watch Report 13-03 May 2013* (Digital Preservation Coalition, 2013), 2nd edition, available at <http://dx.doi.org/10.7207/twr13-03>.

Lawrence, Steve, 'Free Online Availability Substantially Increases a Paper's Impact', *Nature*, 411/521 (2001), available at

<http://www.nature.com/nature/journal/v411/n6837/full/411521a0.html>.

Lebrave, Jean-Louis, 'De l'Édition Informatisé à l'Édition Électronique', in Aurèle Crasson (ed.), *L'Édition du Manuscrit. De l'Archive de Création au Scriptorium Électronique* (Louvain-La-Neuve: Academia Bruylant, 2008): 169-188.

Lowry, Martin J.C., *The World of Aldus Manutius: Business and Scholarship in Renaissance Venice* (Oxford: Blackwell, 1979).

Lucius, Wulf D. v., 'Publishers as Elements of the Scientific Communication System', *Poiesis & Praxis: International Journal of Technology Assessment and Ethics of Science Praxis*, 5 (2008): 125-137, available at <http://www.springerlink.com/content/4x375728t3006811/fulltext.pdf>.

Macé, Caroline and Baret, Philippe V., 'Why Phylogenetic Method Work: the Theory of Evolution and Textual Criticism', in Macé Caroline, Baret Philippe, Andrea Bozzi, and Cignoni Laura (eds), *The Evolution of Texts: Confronting Stemmatological and Genetical Methods*, XXIV-XXV (Pisa, Roma: Istituti Editorali e Poligrafici Internazionali, 2006).

Mahoney, Anne, 'Tachypaedia Byzantina: The Suda On Line as Collaborative Encyclopedia', *Digital Humanities Quarterly*, 3/1 (2009), available at <http://www.digitalhumanities.org/dhq/vol/003/1/000025/000025.html>.

McCarty, Willard, 'Finding Implicit Patterns in Ovid's Metamorphoseos with TACT', *CCH Working Papers*, 1 (1991), available at <http://www.digitalstudies.org/ojs/index.php/digital_studies/article/view/105/170>.

McCarty, Willard, 'Modeling: A Study in Words and Meanings', in Susan Schreibman, Ray Siemens, and John Unsworth (eds), *A Companion to Digital Humanities* (Oxford: Blackwwell, 2004): 254-270.

McCarty Willard, *Humanities Computing* (Basingstoke: Palgrave Macmillan, 2005).

McCarty, Willard, 'Working Digitally', *#alt-academy* (2011), available at <http://mediacommons.futureofthebook.org/alt-ac/pieces/working-digitally>.

McCarty, Willard, 'Collaborative Research in the Digital Humanities', in Marilyn Deegan and Willard McCarty (eds), *Collaborative Research in the Digital Humanities* (Aldershot: Ashgate, 2012): 1-10.

McGann, Jerome, *The Textual Condition* (Princeton: Princeton University Press, 1991).

McGann, Jerome, 'Marking Texts of Many Dimensions', in Susan Schreibman, Ray Siemens, and John Unsworth (eds), *A Companion to Digital Humanities* (Oxford: Blackwwell, 2004): 198-217.

McGann, Jerome, 'Electronic Archives and Critical Editing', *Literature Compass*, 7/2 (2010): 37-42, available at doi: 10.1111/j.1741-4113.2009.00674.x.

McKenzie, Donald F., *Bibliography and the Sociology of Texts* (London: British Library, 1986).

McLaverty, James, 'The Concept of Authorial Intention', *The Library*, 6/9 (1984): 121-38.

McLouglin, Tim, 'Bridging the Gap', in Malte Rehbein and Sean Ryder (eds), *Computerphilologie*, 10 (2010): 37-54, available at <http://computerphilologie.tu-darmstadt.de/jg08/mclough.pdf>.

McLuhan, Marsahll, *Understanding Media: The Extensions of Man Understanding Media: The Extensions of Man* (New York: McGraw-Hill, 1964).

Monella, Paolo, 'A Sausserean Approach to Graphemes Declaration in charDecl for Manuscripts Encoding', in *2013 TEI Conference: The Linked TEI* (2013), available at <http://www.unipa.it/paolo.monella/tei2013/saussurean.pdf>.

Murray, Elizabeth K.M., *Caught in the Web of Words: James Murray and the Oxford English Dictionary Murray, KM Elizabeth* (New Haven: Yale University Press, 2001).

Nell Smith, Martha, 'Electronic Scholarly Editing', in Susan Schreibman, Ray Siemens, and John Unsworth (eds), *A Companion to Digital Humanities* (Oxford: Blackwwell, 2004).

Nielsen, Jakob, 'How Users Read on the Web', *Alertbox* (2009), available at <http://www.nngroup.com/articles/how-users-read-on-the-web/>.

Nielsen, Jakob, 'iPad and Kindle Reading Speeds', *Alertbox* (2010), available at <http://www.nngroup.com/articles/ipad-and-kindle-reading-speeds/>.

Nowviskie, Bethanie, 'Where Credit Is Due: Preconditions for the Evaluation of Collaborative Digital Scholarship', *Profession* (2011): 169-181, available at <http://www.mlajournals.org/doi/pdf/10.1632/prof.2011.2011.1.169>.

Nowviskie, Bethanie, 'Introduction: Two Tramps in Mud Time', *#alt-academy* (2011a), available at <http://mediacommons.futureofthebook.org/alt-ac/pieces/introduction-

two-tramps-mud-time>.

O'Donnell, Daniel Paul, 'Resisting the Tyranny of the Screen, or, Must a Digital Edition be Electronic', *The Heroic Age*, 11 (2008), available at http://www.mun.ca/mst/heroicage/issues/11/em.php.

O'Donnell, Daniel Paul, 'Disciplinary Impact and Technological Obsolescence in Digital Medieval Studies', in Ray Siemens and Susan Schreibman (eds), *A Companion to Digital Literary Studies* (Oxford: Blackwell, 2008b).

Ong, Walter J., 'The Writer's Audience is Always a Fiction', *PMLA*, 90/1 (1975): 9-21.

Ore Espen, '... They Hid their Books Underground'. *Transferred Illusions. Digital Technology and the Forms of Print* (Aldershot: Ashgate, 2009): 113-125.

Panciera, Silvio, 'Struttura dei Supplementi e Segni Diacritici Dieci Anni Dopo', *Supplementa Italica*, 8 (1991): 9-21.

Parker, D.C., 'The Text of the New Testament and Computers: the International Greek New Testament Project', *Literary and Linguistic Computing*, 15/1 (2000): 27-41.

Parkes, Malcom B., *Pause and Effect: An Introduction to the History of Punctuation in the West* (Aldershot: Ashgate, 1992).

Petrucci, Armando and Romeo, Carlo, 'L'orazionale Visigotico di Verona: Aggiunte Avventizie, Indovinello Grafico, Tagli Maffeiani', *Scrittura e Civiltà*, 22 (1998): 13-30.

Pierazzo, Elena, 'Editorial Teamwork in a Digital Environment: The Edition of the Correspondence of Giacomo Puccini', *Jahrbuch für Computerphilologie*, 10 (2008), available at <http://computerphilologie.tu-darmstadt.de/jg08/pierazzo.html>.

Pierazzo, Elena, 'A Rationale of Digital Documentary Editions', *Literary and Linguistic Computing*, 26/4 (2011): 463-477.

Pierazzo, Elena, 'Of Time and Space: a New Framework for Digital Editions of Draft Manuscripts', *Variants*, 11 (2014a): forthcoming.

Pierazzo, Elena, 'Digital Documentary Editions and the Others', *Scholarly Editing*, 35 (2014b), available at <http://www.scholarlyediting.org/2014/essays/essay.pierazzo.html>.

Pierazzo, Elena and Stokes, Peter A., 'Putting the Text Back into Context: a Codicological

Approach to Manuscript Transcription', in Franz Fischer, Christiane Fritze, and Georg Voelgler (eds), *Kodikologie und Paläographie im Digitalen Zeitalter 2 - Codicology and Palaeography in the Digital Age 2* (Norderstedt: Books on Demand, 2010): 397-430.

Porter, Dot, 'Medievalists and the Scholarly Digital Edition', *Scholarly Editing*, 34 (2013): 1-26, available at <http://www.scholarlyediting.org/2013/pdf/essay.porter.pdf>.

Price, Kenneth M., 'Electronic Scholarly Editions', in Ray Siemens and Susan Schreibman (eds), *A Companion to Digital Literary Studies* (Oxford: Blackwell, 2008): 434-450.

Price Kenneth M., 'Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?', *Digital Humanities Quarterly*, 3/3 (2009), available at <http://digitalhumanities.org:8080/dhq/vol/3/3/000053/000053.html>.

Price, Kenneth M., 'Why the Whitman Archive Isn't (Really) Using Social Editing', *Presentation at 'Social, Digital Scholarly Editing' Conference*. University of Saskatchewan, Saskatoon. (2013). (2013).

Pritchard, David, 'Working Papers, Open Access, and Cyber-infrastructure in Classical Studies', *Literary and Linguistic Computing*, 23/2 (2008): 149-162.

Rehbein, Malte, 'Reconstructing the Textual Evolution of a Medieval Manuscript', *Literary and Linguistic Computing*, 24/3 (2009): 319-27.

Rehbein, Malte, 'The Transition from Classical to Digital Thinking. Reflections on Tim McLoughlin, James Barry and Collaborative Work', in Malte Rehbein and Sean Ryder (eds), *Computerphilologie*, 10 (2010): 55-67, available at <http://computerphilologie.tu-darmstadt.de/jg08/rehbein.pdf>.

Rehbein, Malte and Fritze, Christiane, 'Hands-On Teaching Digital Humanities: A Didactic Analysis of a Summer School Course on Digital Editing', in Brett D. Hirsch (ed.), *Digital Humanities Pedagogy: Practices, Principles and Politics* (Cambridge: Open Book Publisher, 2012): 47-78.

Renear, Allen H., Mylonas, Elli, and Durand, David, 'Refining our Notion of What Text Really Is: The Problem of Overlapping Hierarchies', in Nancy Ide and Susan Hockey (eds), *Research in Humanities Computing* (Oxford: Oxford University Press, 1996), available at <http://www.stg.brown.edu/resources/stg/monographs/ohco.html>.

Renear, Allen H., 'Text Encoding', in Susan Schreibman, Ray Siemens, and John Unsworth

(eds), *A Companion to Digital Humanities* (Oxford: Blackwwell, 2004): 218-239.

Robinson, Peter M. W., 'The Collation and Textual Criticism of Icelanding Manuscripts (1): Collation', *Literary and Linguistic Computing*, 4/2 (1989): 99-105.

Robinson, Peter M. W. and Solopova, Elisabeth, 'Guidelines for Transcription of the Manuscripts of the Wife of Bath's Prologue', in Norman Francis Blake and Peter M. W. Robinson (eds), *The Canterbury Tales Project: Occasional Papers*, 1 (: Oxford University Computing Services, Office for Humanities Communication, 1993): 19-52, available at <http://www.canterburytalesproject.org/pubs/op1-transguide.pdf>.

Robinson, Peter M. W., 'Where We are with Electronic Scholarly Editions, and Where We Want To Be', *Computerphilologie*, 5 (2003), available at <http://computerphilologie.tu-darmstadt.de/jg03/robinson.html>.

Robinson, Peter M.W., 'Making Electronic Editions and the Fascination of What is Difficult', in Andrea Bozzi, Laura Cignoni, and Jean-Louis Lebrave (eds), *Digital Technologies and Philological Disciplines*, XX-XXI (Pisa: Istituti Editorali e Poligrafici Internazionali, 2004): 415-437.

Robinson, Peter M. W., 'Current Issues in Making Digital Editions of Medieval Texts – or, Do Electronic Scholarly Editions Have a Future?', *Digital Medievalist*, 1/1 (2005), available at <http://www.digitalmedievalist.org/journal/1.1/robinson/>.

Robinson, Peter M.W., 'What Text Really is Not, and Why Editors Have to Learn to Swim', *Literary and Linguistic Computing*, 24/1 (2009): 41-52.

Robinson, Peter M.W., 'Toward a Theory of Digital Editions', *Variants*, 10 (2013): 105-132.

Robinson, Peter M.W., 'Why Digital Humanists Should Get Out of Textual Scholarship', *Presentation at 'Social, Digital Scholarly Edition' Conference,* Saskatoon, University of Saskatchewan (2013a), available at <http://www.academia.edu/4124828/SDSE_2013_why_digital_humanists_should_get_out_of_textual_scholarship>.

Robinson, Peter M.W., 'Why Digital Humanists Should Get Out of Textual Scholarship. And if They Don't, Why We Textual Scholars Should Throw Them Out', Academic blog (2013b), available at <http://scholarlydigitaleditions.blogspot.co.uk/2013/07/why-digital-humanists-should-get-out-of.html>.

Rockwell, Geoffrey, 'Crowdsourcing in the Humanities: Social Research and Collaboration',

in Marilyn Deegan and Willard McCarty (eds), *Collaborative Research in the Digital Humanities* (Aldershot: Ashgate, 2012): 135-154.

Rosenblueth, Arturo and Wiener, Robert, 'The Role of Models in Science', *Philosophy of Science*, 12 (1945): 316-322.

Rosselli Del Turco, Roberto, 'After the Editing is Done: Designing a Graphic User Interface for Digital Editions', *Digital Medievalist*, 7 (2011), available at <http://www.digitalmedievalist.org/journal/7/rosselliDelTurco/>.

Rouché, Charlotte, 'Why Do We Mark Up Texts?', in Marilyn Deegan and Willard McCarty (eds), *Collaborative Research in the Digital Humanities* (Aldershot: Ashgate, 2012): 155-162.

Ruecker, Stan, Radzikowska, Milena, and Sinclair, Stéfan, *Visual Interface Design for Digital Cultural Heritage: A Guide to Rich-Prospect Browsing* (Aldershot: Ashgate, 2011).

Sahle, Patrick, *A Catalogue of Digital Scholarly Editions* (2008), available at <http://www.digitale-edition.de/>.

Sahle Patrick, *Digitale Editionsformen, Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels* (Norderstedt: Books on Demand, 2013), 3 vol.

Salemans, Benedictus Johannes Paulus, *Building Stemmas with the Computer in a Cladistic, Neo-Lachmanian, Way. The case of Fourteen Text Versions of Lanseloet van Denemerken* (Nijmegen: Katholieke Universiteit Nijmegen, 2001), PhD diss. available at <http://www.neder-l.nl/salemans/diss/salemans-diss-2000.pdf>.

Sardou, Victorien, Giacosa, Giuseppe, Illica, Luigi, and Puccini, Giacomo, *Tosca. Facsimile della Copia di Lavoro del Libretto*, Gabriella Biagi Ravenni (ed.), *Tosca*, 2 (Firenze: Centro Studi Giacomo Puccini - Leo Olschki Editore, 2009).

de Saussure, Ferdinand, *Cours de Linguistique Générale*, Charles Bally, Albert Sechehaye (eds), with Albert Riedlinger, (Paris: Payot, 1916).

Schmidt, Desmond, 'The Inadequacy of Embedded Markup for Cultural Heritage Texts', *Literary and Linguistic Computing*, 25/3 (2010): 337-356.

Sculley, D. and Pasanek, Bradley M., 'Meaning and Mining: the Impact of Implicit Assumptions in Data Mining for the Humanities', *Literary and Linguistic Computing*, 23/4 (2008): 409-424.

Schmidt, Desmond, 'The Role of Markup in the Digital Humanities', *Historical Social Research/Historische Sozialforschung*, 37/3 (2012): 125-146.

Schreibman, Susan, Mandell, Laura, and Olsen, Stepjen, 'Introduction', *Profession* (2011): 123-135, available at <http://www.mlajournals.org/doi/pdf/10.1632/prof.2011.2011.1.123>.

Siemens, Lynne, ''It's a Team if You Use "Reply All" ': An Exploration of Research Teams in Digital Humanities Environments', *Literary and Linguistic Computing*, 24/2 (2009): 225-233.

Siemens, Lynne and the INKE Research Group, 'From Writing the Grant to Working the Grant: An Exploration of Processes and Procedures in Transition', *New Knowledge Environments*, 1/1 (2009): 1-8, available at <http://journals.uvic.ca/index.php/INKE/article/view/169>.

Siemens, Ray, Elkink, Mike, McColl, Alastair, Armstrong, Karin, Dixon, James, Saby, Angelsea, Hirsch, Brett D., Cara Leitch, with Martin Holmes, Haswell, Eric, Gaudet, Chris, Girn, Paul, Joyce, Michael, Gold, Rachel, Watson, Gerry, members of the PKP, Iter, TAPoR, and teams, INKE, 'Underpinnings of the Social Edition? A Narrative, 2004-9, for the Renaissance English Knowledgebase (REKn) and Professional Reading Environment (PReE) Projects', in Jerome McGann, Andrew Stauffer, Dana Wheeles, and Michael Pickard (eds), *Online Humanities Scholarship: The Shape of Things to Come* (Houston: Rice University Press, 2010), available at <http://cnx.org/content/m34335/latest/>.

Siemens, Ray, Timney, Meagan, Leitch, Cara, Koolen, Corina, with the ETCL, Alex Garnett, INKE, and Group, PKP Research, 'Toward Modelling the "Social" Edition: An Approach to Understanding the Electronic Scholarly Edition in the Context of New and Emerging Social Media', *Literary and Linguistic Computing*, 27/4 (2012a): 445-461, available at <doi:10.1093/llc/fqs013>.

Siemens, Ray, Dobson, Teresa, Ruecker, Stan, Cunningham, Richard, Galey, Alan, Warwick, Claire, Siemens, Lynne, with Michael Best, Chernyk, Melanie, Duff, Wendy, Flanders, Julia, Gants, David, Gervais, Bertrand, MacLean, Karon, Ramsay, Steve, Rockwell, Geoffrey, Schreibman, Susan, Swindells, Colin, Vandendorpe, Christian, Copeland, Lynn, Willinsky, John, Zafrin, Vika, the HCI-Book Consultative Group, and the INKE Research Team, 'Human-Computer Interface/Interaction and the Book: a Consultation-derived Perspective on Foundational E-Book Research', in

Marilyn Deegan and Willard McCarty (eds), *Collaborative Research in the Digital Humanities* (Aldershot: Ashgate, 2012b): 163-189.

Siemens, Ray, Karin Armstrong, Barbara Bond, Constance Crompton, Alyssa Arbuckle, Terra Dickson, Johanne Paquette, Jonathan Podracky, Ingrid Weber, Cara Leitch, Melanie Chernyk, Bret D. Hirtsch, Daniel Powell, Chris Gaudet, Eric Haswell, Arianna Ciula, Daniel Starza-Smith, James Cummings, with Martin Holmes, Greg Newton, Jonathan Gibson, Paul Remley, Erik Kwakkel, and Aimie Shirkie, *A Social Edition of the Devonshire MS (BL Add 17,492)*, Wikibooks (2013), available at <http://en.wikibooks.org/wiki/The_Devonshire_Manuscript>.

Shillingsburg, Peter L., 'Text as Matter, Concept, and Action', *Studies in Bibliography*, 44 (1991): 31-83.

Shillingsburg Peter L., *Scholarly Editing in the Computer Age. Theory and Practice Scholarly Editing in the Computer Age. Theory and Practice* (Ann Arbor: The University of Michigan Press, 1996).

Shillingsburg Peter L., *From Gutenberg to Google* (Cambridge: Cambridge University Press, 2006).

Sinclair, Stéfan and Rockwell, Geoffrey, *Voyeur Tools* (2009), available at <http://hermeneuti.ca/voyeur>.

Spencer, Matthew, Davidson, Elizabeth A., Barbrook, Adrian C., and Howe, Christopher J., 'Phylogenetics of Artificial Manuscripts', *Journal of Theoretical Biology*, 227/4 (2004): 503-511.

Sperberg-McQueen, C. Michael, 'Text in Electronic Age: Textual Study and Text Encoding, with Examples from Medieval Texts', *Literary and Linguistic Computing*, 6/1 (1991): 34-46.

Sperberg-McQueen, C. Michael, 'How to Teach your Edition How to Swim', *Literary and Linguistic Computing*, 24/1 (2009): 27-52.

Sperberg McQueen, C. Michael, Huitfeldt, Claus, and Marcoux, Yves, 'What is Transcription? Part 2', *Digital Humanites 2009. Conference Abstracts. University of Maryland, June 22-25* (2009): 257-260.

Stokes, Peter A., 'Modeling Medieval Handwriting: A New Approach to Digital Palaeography', in Jan Christoph Meister (ed.), *Digital Humanites 2012. Conference*

Abstracts. University of Hamburg, July 16-22 (Hamburg: Hamburg University Press, 2012): 382-384.

Stokes, Peter A. and Pierazzo, Elena, 'Encoding the Language of Landscape: XML and Databases at the Service of Anglo-Saxon Lexicography', in Silvia Bruti, Roberta Cella, and Marina Foschi Albert (eds), *Perspectives on Lexicography in Italy and Europe* (Newcastle: Cambridge Scholars Publishing, 2009): 203-38.

Stokes, Peter A. *et al*. (eds), *DigiPal: Digital Resource and Database of Palaeography, Manuscripts and Diplomatic* (London: King's College London, 2011-12), available at <http://digipal.eu/>.

Stewart, William A., 'A Sociolinguistic Typology for Describing National Multilingualism', in Joshua A. Fishman (ed.), *Readings in the Sociology of Language* (Paris: Mouton, 1968): 529-545.

Suber, Peter, *Open Access* (Cambridge (MA): MIT Press, 2012), available at <http://mitpress.mit.edu/books/open-access>.

Sutherland, Kathryn, 'Being Critical: Paper-based Editing and the Digital Environment' (Aldershot: Ashgate, 2009): 13-25.

Sutherland, Kathryn, *Jane Austen's Fiction Manuscripts: A Digital Edition*, (London and Oxford: King's Collegs London and Oxford University, 2010), available at <http://www.janeausten.ac.uk/>.

Sutherland, Kathryn, 'Anglo-American Editorial Theory', in Neil Fraistat and Julia Flanders (eds), *The Cambridge Companion to Textual Scholarship* (Cambridge: Cambridge University Press, 2013): 42-60.

Sutherland, Kathryn and Pierazzo, Elena, 'The Author's Hand: from Page to Screen', in Marilyn Deegan and Willard McCarty (eds), *Collaborative Research in the Digital Humanities* (Aldershot: Ashgate, 2012): 191-212.

Tanselle, G. Thomas, 'The Editing of Historical Documents', *Studies in Bibliography*, 31 (1978): 1-56.

Tanselle, G. Thomas, 'The Varieties of Scholarly Editing', in David C. Greetham (ed.), *Scholarly Editing. A Guide to Research* (New York: Modern Language Association of America, 1995a): 9-32.

Tanselle, G. Thomas, 'Critical Editions, Hypertexts, and Genetic Criticism', *The Romanic*

Review, 86/3 (1995b), available at <http://www.questia.com/library/journal/1G1-18137111/critical-editions-hypertexts-and-genetic-criticism>.

Tanselle, G. Thomas, *A Rationale of Textual Criticism* (Philadelphia: Pennsylvania University Press, 1989a).

Tanselle, G. Thomas, 'Reproductions and Scholarship', *Studies in Bibliography*, 42 (1989b): 25-54.

Tarte, Ségolène M., 'Digital Visual Representations in Papyrology: Implications on the Nature of Digital Artefacts' (2011), available at <http://www.academia.edu/776645/Digital_Visual_Representations_in_Papyrology_Implications_on_the_Nature_of_Digital_Artefacts>.

TEI Consortium (eds), *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. 2.6.0, 20th January 2014, available at http://www.tei-c.org/Guidelines/P5/ (Accessed 6 August 2014).

Terras, Melissa, *Digital Images for the Information Professional* (Aldershot: Ashgate, 2008).

Thaller, Manfred, 'Digital Manuscripts as Base Line for Dynamic Editions', in Jean-Louis Lebrave Andrea Bozzi, Laura Cignoni (ed.), *Digital Technology and Philological Disciplines*, XX-XXI (: Istituti Editorali e Poligrafici Internazionali, 2004): 489-511.

Thompson, Clive, 'Retro design is crippling innovation', *Wired* (2012), available at <http://www.wired.co.uk/magazine/archive/2012/03/ideas-bank/clive-thompson>.

Tillet, Barbara B., 'FRBR and Cataloguing for the Future', *Cataloguing and Classification Quarterly*, 39/3-4 (2005): 197-205.

Treharne, Elaine, 'Fleshing out the TEXT: the Transcendent Manuscript in the Digital Age', *Postmedieval*, 4/4 (2013): 1-16.

Tumariello, Giovanni, Morbidoni, Cristian, and Pierazzo, Elena, 'Toward Textual Encoding Based on RDF', in *ELPUB 2005. Proceedings of the International Conference* (Leuven: KU Leuven, 2005): 57-64, available at <http://elpub.scix.net/data/works/att/206elpub2005.content.pdf>.

Twycross, Meg, 'Virtual Restoration and Manuscritp Archeology', in Mark Greengrass and Lorna Hughes (eds), *The Virtual Representation of the Past* (Aldershot: Ashgate, 2008): 23-47.

Unsworth, John, 'Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might our Tools Reflect This?', *Paper delivered at the symposium 'Humanities Computing: formal methods, experimental practice'*, King's College, London, May 13, 2000 (2000), available at <http://people.lis.illinois.edu/~unsworth//Kings.5-00/primitives.html>.

Vandendorpe, Christian, 'Reading on Screen: The New Media Sphere', in Ray Siemens and Susan Schreibman (eds), *A Companion to Digital Literary Studies* (Oxford: Blackwell, 2008): 203-215.

Vander Meulen, David L. and Tanselle, G. Thomas, 'A System of Manuscript Transcription', *Studies in Bibliography*, 52 (1999): 201-213.

Vanhoutte, Edward, 'Traditional Editorial Standards and the Digital Edition', in Els Stronks and Peter Boot (eds), *Learned Love. Proceedings of the Emblem Project Utrecht Conference on Dutch Love Emblems and the Internet (November 2006)* (The Hague: DANS - Data Archiving and Networked Services, 2007): 157-174.

Vanhoutte Edward, 'Every Reader his own Bibliographer - An Absurdity?', in Marilyn Deegan and Katrhyn Sutherland (eds), *Text Editing, Print and the Digital World* (Aldershot:Ashgate, 2009), pp. 99-110.

Vanhoutte Edward, 'Defining Electronic Editions: A Historical and Functional Perspective', in Willard McCarty (ed.), *Text and Genre in Reconstruction. Effects of Digitalization on Ideas, Behaviours, Products and Institutions* (Cambridge: Open Book Publisher, 2010): 119-144.

Vanhoutte Edward, 'The Gates of Hell: History and Definition of Digital | Humanities | Computing', in Melissa Terras, Julianne Nyhan, and Edward Vanhoutte (eds), *Defining Digital Humanities. A Reader* (Aldershot: Ashgate, 2013): 119-156.

Van den Branden, Ron, Terras, Melissa, and Vanhoutte, Edward, *TEI by Example* (2010), available at <http://www.teibyexample.org>.

Van Hulle, Dirk, Nixon, Mark, and with Neyt, Vincent (eds), *Samuel Beckett Digital Manuscript Project* (Antwerp: University Press Anwerp, 2013), available at <http://www.beckettarchive.org/>.

Van Zundert, Joris and Boot, Peter, 'The Digital Edition 2.0 and the Digital Library: Services, not Resources', *Bibliothek and Wissenschaft*, 44 (2012): 141-152.

Vauthier, Bénédicte, 'Coups de Sonde dans un Dossier Génétique Hispanique (Essentiellement) Numérique et Essai de Théorisation : El Dorado (2008) de Robert Juan-Cantavella' in *Conference Abstracts: Variance in Textual Scholarship and Genetic Criticism/La Variance en Philologie et dans la Critique Génétique* (2013), abstract available at <http://www.textualscholarship.eu/conference-2013-abstracts.html>.

Weaver, Warren and Shannon, Claude E., *The Mathematical Theory of Communication* (Chicago: University of Illinois Press, 1963).

Whitelaw, Mitchell, 'Generous Interfaces', *Keynote Presentation at the National Digital Forum 2011 in Wellington* (2011), available at <http://visiblearchive.blogspot.co.uk/2011/12/generous-interfaces-ndf-2011-keynote.html>.

Windram, Heather F., Shaw, Prue, Robinson, Peter M.W., and Howe, Christopher J., 'Dante's 'Monarchia' as a Test Case for the Use of Phylogenetic Methods in Stemmatic Analysis', *Literary and Linguistic Computing*, 23/4 (2008): 443-463.

Zeller, Hans, 'Record and Interpretation: Analysis and Documentation as Goal and Method of Editing', in Hans Walter Gabler, George Bornstein, and Gillian Borland Pierce (eds), *Contemporary German editorial theory* (Ann Arbor: University of Michigan Press, 1995): 17-58.

Zichermann, Gabe and Cunningham, Christopher, *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. (Sebastopol: O'Reiley, 2011).

Zielinski, Andrea, Pempe, Wolfgang, Gietz, Peter, Haase, Martin, Funk, Stefan, and Simon, Christian, 'TEI Documents in the Grid', *Literary and Linguistic Computing*, 24/3 (2009): 267-279.