



**HAL**  
open science

## Valorizzazione e mediazione del patrimonio culturale delle Tecnologie dell'Informazione: il progetto ITinHeritage

Caroline Djambian, Micaela Rossi, Giada d'Ippolito

### ► To cite this version:

Caroline Djambian, Micaela Rossi, Giada d'Ippolito. Valorizzazione e mediazione del patrimonio culturale delle Tecnologie dell'Informazione: il progetto ITinHeritage. 33° Convegno dell'Associazione Italiana per la Terminologia "Terminologia del patrimonio culturale materiale e immateriale: analisi e approcci di studio", Nov 2023, Naples, Italy. hal-04604846

**HAL Id: hal-04604846**

<https://hal.univ-grenoble-alpes.fr/hal-04604846v1>

Submitted on 12 Jun 2024

**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.

Caroline Djambian (Université Grenoble Alpes, Francia) [caroline.djambian@univ-grenoble-alpes.fr](mailto:caroline.djambian@univ-grenoble-alpes.fr)

Micaela Rossi (Università di Genova, Italia) [micaela.rossi@unige.it](mailto:micaela.rossi@unige.it)

Giada D'Ippolito (Università di Genova, Italia) [giadadippolito30@gmail.com](mailto:giadadippolito30@gmail.com)

**Valorizzazione e mediazione del patrimonio culturale delle Tecnologie  
dell'Informazione:  
il progetto ITinHeritage**

**Abstract**

Il progetto ITinHeritage riunisce ricercatori nell'ambito delle Digital Humanities intorno al desiderio comune di salvaguardare il patrimonio materiale e immateriale delle Tecnologie dell'Informazione (TI). Questo settore, dall'impatto e influenza sociale importanti ma non facilmente accessibile al pubblico, è infatti ancora sottovalutato, come molti patrimoni scientifici e tecnici contemporanei.

Attraverso un approccio patrimoniale basato sui principali attori del patrimonio museale delle TI europeo, il progetto si propone di creare una piattaforma sul web semantico per portare gli artefatti e le conoscenze che questi veicolano fuori dalle mura dei musei, e collegarli al resto del mondo. La piattaforma dà accesso a un *knowledge graph* supportato da un onto-terminologia del dominio che introdurremo qui. Il linguaggio è infatti al centro del ruolo di un museo scientifico che trascrive il discorso tecnico nell'ambiente museale e lo colloca nello spazio pubblico. Questo lavoro onto-terminologico, basato sulla creazione di un dizionario multilingue e multimediale sul modello dei primi dizionari scientifici e tecnici che hanno fondato il lavoro terminologico, seguirà una metodologia collaudata in conformità con gli standard ISO 704 e ISO 1087.

Presentiamo in queste pagine il progetto ITinHeritage nel suo complesso, con un'attenzione particolare per la fase iniziale di implementazione dell'ontologia e del glossario a partire da corpora differenziati per genere e comunità d'uso. Saranno analizzate le procedure preliminari di analisi dello stato dell'arte, delle risorse esistenti per la mappatura dei concetti e della terminologia dell'IT in italiano, inglese, francese e tedesco, così come le criticità che emergono dalle prime fasi di modellizzazione.

## 1 Introduzione

Che cos'è il patrimonio delle Tecnologie dell'Informazione (TI)? Come si può circoscrivere e definire? Quali sono le sue radici e i suoi impatti tecnologici, economici, politici e sociali, inediti e sproporzionati? L'UNESCO definisce le TI come :

‘l'insieme di strumenti e risorse tecnologiche per trasmettere, registrare, creare, condividere o scambiare informazioni...’. (UNESCO, 2024).

Oggi gli oggetti informatizzati sono innumerevoli (Lazard, Mounier-Khun, 2022) e la rapida e massiccia evoluzione delle TI non ha lasciato tempo per uno studio approfondito. Esaminare questo patrimonio è fondamentale per comprendere il mondo di oggi e l'evoluzione della società dalla seconda metà del XX secolo. Il nostro approccio, patrimoniale e interdisciplinare, ha quindi come punto di partenza gli spazi museali dell'informatica. Il Musée des Arts et Métiers (Parigi, Francia), il London Science Museum (Londra, Inghilterra), ACONIT (Grenoble, Francia), il NAM-IP (Namur, Belgio), il Museo del calcolo (Pisa, Italia), l'Homecomputermuseum (Helmond, Olanda) e l'HNF (Paderborn, Germania) sono attivamente coinvolti nella nostra rete in espansione.

Il progetto di ricerca ITinHeritage si propone di condurre una riflessione epistemologica sulla questione delle TI, ponendo al centro della questione del patrimonio la storia di una mutazione sociale e l'emergere di nuove conoscenze in termini di rappresentazioni del mondo, attraverso la definizione di questo patrimonio.

## 2 Il patrimonio delle Tecnologie dell'Informazione: tra materialità e immaterialità

Il patrimonio delle TI è specifico, in costante accelerazione e caratterizzato dalla discontinuità tecnologica. Questo fenomeno globale si sta integrando in tutti gli strati della società. Eppure il patrimonio che sottende questo settore non è mai stato circoscritto e messo in discussione. Dobbiamo innanzitutto interrogare questo patrimonio labile, il cui sapere è cristallizzato in varie forme.

La prima espressione del sapere delle TI è esplicita e si concretizza nell'oggetto scientifico e tecnico, materiale e immateriale. Oggi l'oggetto museale, poco patrimonializzato perché troppo recente, non permette più di materializzare la scienza come prima. Da questo punto di vista, le scienze dure cominciano a incontrare problemi già noti da tempo alle scienze umane. La mancanza di considerazione da parte delle istituzioni e il fatto che alcune collezioni tecniche stiano sempre più diventando

collezioni virtuali senza museo conferisce a tali oggetti spessore storico. Il settore delle TI è un caso emblematico del patrimonio recente in termini di mediazione e trasmissione: la sua dimensione tecnologica è forte e i suoi oggetti materiali offrono una scarsa leggibilità visiva: la loro forma non è mediatrice di per sé, a differenza degli oggetti d'arte. Le nostre analisi sono quindi progettate per fornire anche risposte alla domanda: come possiamo mediare quello che non crea senso visivamente e scientificamente?

Interessarsi a questo recente patrimonio scientifico e tecnico significa inoltre interessarsi alla questione dei musei scientifici. Più strategici che i musei d'arte, tali musei sono sempre più orientati verso il pubblico e, conseguentemente, verso la società (Schiele, 1989). Oltre al pubblico, l'obiettivo è il cittadino, cioè un individuo in grado di partecipare al dibattito pubblico sulla scienza. A tal fine, il museo è orientato all'educazione, al trasferimento della conoscenza. Il museo scientifico si considera quindi una delle componenti dell'arena pubblica. Il suo scopo è quello di diffondere la conoscenza e, oltre a questo, di fornire competenze su tale conoscenza.

Nondimeno, in questi ultimi decenni, il patrimonio scientifico tende a diventare un patrimonio di dati piuttosto che di oggetti materiali, a causa della crescente smaterializzazione della scienza e dell'accelerazione dell'obsolescenza del patrimonio, a maggior ragione nel caso delle TI (Schiele, 1989). Il patrimonio immateriale si sta sviluppando tanto quanto quello materiale. Di conseguenza, i musei scientifici si troveranno sempre più spesso ad affrontare il compito di mostrare l'immaterialità. I manufatti digitalizzati e la loro documentazione, i software e i dati sono tutti oggetti immateriali che costituiscono le nuove collezioni della scienza moderna. Il primo passo è allora definire il patrimonio delle TI, la sua crescente immaterialità e la sua patrimonializzazione (Bortolotto, 2012).

La seconda espressione del sapere delle TI è tacita. Consiste nella conoscenza esperienziale. La 'conoscenza pura' (ragionamento) (Kant, 1997) è 'exoterica' (Jacob, 2001) quando si concretizza nella scrittura o negli artefatti. Si contrappone alla 'conoscenza empirica', 'emperia' o 'metis' di Aristotele, che è sensoriale, come il movimento della mano, e può essere espressa solo dal suo portatore riferendosi all'azione, alla esperienza pratica attorno agli oggetti scientifici e tecnici. Raccogliere e trasmettere questo sapere tacito, 'esoterico' è una vera e propria sfida, perché per essere compreso, l'emperia richiede comunque un processo di formalizzazione nei discorsi.

Questo ci porta al linguaggio specialistico, che esprime l'esperienza delle realtà del mondo specifiche di un settore. Che la trasmissione del sapere sia orizzontale (tra pari),

verso il basso (divulgazione) o verso l'alto (appropriazione da parte del pubblico), è il linguaggio ad essere al centro dei processi di formazione, appropriazione e mediazione. Con l'aiuto di esperti, che hanno una responsabilità sociolinguistica nel trasferimento del sapere perché sono gli unici a padroneggiare il discorso del settore, il suo campo concettuale e la sua consensualità (Putnam, 1984), stiamo costruendo un ontoterminologia basata sui modelli fondanti (ISO 704 e 1087), a partire dalle collezioni dei musei partner, completate da un corpus di storia dell'informatica (Mounier-Kuhn & Lazard, 2022 ; Haigh & Ceruzzi, 2021). Lo studio della diversità linguistica (Djambian et al, 2024; Rossi, 2021) ci permetterà di capire i meccanismi di concettualizzazione (Gaudin, 1996), di costruire una dialettica tra scienza e pubblico e di studiare l'evoluzione congiunta lingua - tecnologia - società (Djambian et al., 2023).

### **3 Il progetto ITinHeritage come unione tra gli approcci knowledge-based e la terminologia testuale**

La caratteristica che identifica il progetto ITinHeritage e ne costituisce al tempo stesso la sfida più difficile è senza dubbio la sua natura profondamente interdisciplinare, all'incrocio tra terminologia testuale, approcci volti alla gestione della conoscenza specialistica (*knowledge-based*) e valorizzazione del patrimonio. La natura profondamente eterogenea della disciplina delle IT, la sua costante e multiforme evoluzione in termini di linguaggi e media, la stratificazione del pubblico di riferimento e delle produzioni discorsive, la presenza di diverse tradizioni linguistiche e culturali, ma soprattutto la necessità di adottare un approccio metodologico realmente interdisciplinare nella prospettiva delle Digital Humanities, costituiscono altrettante sfide per il gruppo di ricerca.

La sfida maggiore e più urgente è proprio quella di ridefinire il peso delle componenti dei saperi e dei termini nell'economia generale di ITinHeritage. Se infatti da un lato il progetto si propone di sistematizzare e di valorizzare i saperi specialistici, le pratiche diffuse nelle comunità degli esperti nell'ambito delle IT, tra gli obiettivi fondamentali spicca a parità di importanza l'analisi terminologica. Questo movimento continuo, che oscilla tra i due poli dei *saperi*, sostanziati nella creazione di un'ontologia formale, e dei *termini*, descritti nel loro utilizzo attraverso il dizionario multilingue che accompagna l'ontologia, è un movimento che obbliga costantemente ad una riflessione più ampia, sul contesto teorico e metodologico che sottende il presupposto del progetto stesso.

La metodologia che sta alla base del progetto si inserisce pienamente in quest'ottica negli attuali sviluppi teorici dell'ambito degli studi terminologici. Se infatti per un decennio

almeno (1990-2000) gli studi terminologici si sono focalizzati soprattutto sulla componente testuale, variazionista, legata alla realizzazione discorsiva dei termini, negli ultimi anni si assiste ad un ritorno all'analisi dei concetti, alla formalizzazione e alla modellizzazione delle conoscenze specialistiche, anche in virtù del legame sempre più stretto tra terminologia e tecnologie del web semantico, nonché della crescente importanza attribuita alle discipline correlate (prima tra tutte il *knowledge management*). Il recente volume collettaneo curato da Faber e L'Homme (2022) rappresenta in questo contesto un contributo importante nel delineare le linee di sviluppo per la ricerca terminologica del XXI secolo, tra cui spicca l'interazione tra terminologia e approcci volti alla gestione della conoscenza; i saggi contenuti nel volume, tra i quali si notino in particolare il capitolo affidato a Nuopponen, che ripercorre l'evoluzione degli studi sulle relazioni concettuali, e la ripubblicazione del saggio di Meyer (1992) sulla costituzione delle basi di conoscenza terminologiche, mirano a una forte focalizzazione sugli aspetti della formalizzazione concettuale, un ritorno ai postulati alla base della teoria wüsteriana, dopo una fase fortemente incentrata sulla variazione discorsiva (cfr. Humbley e Candel, nello stesso volume), un ritorno chiaramente illustrato da Elena Montiel-Ponsoda nel suo contributo sull'interazione tra terminologia e ontologia.

Il legame indissolubile tra ontologia e terminologia non è un tema nuovo, come dimostra l'abbondante bibliografia in questo campo a partire dalla "svolta cognitiva" degli anni Duemila e dall'avvento degli approcci cognitivi (cf. il lavoro fondante di Temmerman, 2000, e la sua formulazione dell'unità di comprensione - Units of Understanding, che richiede che le terminologie siano prese in considerazione nella loro natura concettuale, come espressione della conoscenza specializzata). Se L'Homme si domanda, nello stesso volume, in quale misura approcci *knowledge-based* e linguistici siano conciliabili:

'Given the differences between the assumption of lexical-based and knowledge-based approaches and the principle on which they rely, the question is whether they can be used simultaneously in terminology work' (L'Homme, 2022: 259).

si può considerare il progetto ITinHeritage come il tentativo di fornire una risposta affermativa a questa domanda.

Il quadro teorico e metodologico privilegiato dal progetto è infatti l'approccio onto-terminologico elaborato da Christophe Roche e dal gruppo di Condillac (<http://ontoterminology.com>). Dagli anni Duemila (Roche, 2005), questo paradigma ha promosso l'analisi concettuale nei saperi specialistici, combinando i principi dell'analisi

ontologica in ambito filosofico con le sue manifestazioni più recenti, legate alla creazione di ontologie nei campi delle scienze della comunicazione e dell'intelligenza artificiale. Secondo questo approccio, è possibile correlare il livello concettuale e il livello linguistico attraverso un ambiente tecnologico *ad hoc*, che permetta al tempo stesso di normalizzare i saperi e di preservare la diversità linguistica nella quale tali saperi trovano espressione:

‘permet de normaliser la seule chose qui peut l’être, à savoir les connaissances du domaine, et de préserver ce qui doit l’être, à savoir la diversité linguistique’ (Roche e Papadopoulou, 2019).

L’approccio introdotto dal gruppo di ricerca Condillac, e annualmente aggiornato nei convegni TOTh, porta alla condivisione studiosi di *knowledge management* e terminologi che hanno individuato come esigenza comune la necessità di un metodo basato su un approccio multidisciplinare, i cui risultati possano essere proposti all’insieme della comunità scientifica, nel rispetto degli standard attuali:

‘[...] une méthode outillée reposant sur une approche pluridisciplinaire, [...] puisant à la linguistique, la terminologie et la modélisation des connaissances permettant la collaboration des différents intervenants et dont les résultats peuvent être partagés grâce à l’utilisation de formats d’échange respectant les standards en vigueur dont ceux de l’ISO et du W3C (RDF/OWL, SKOS, HTML, CSV, JSON, Ontolex-Lemon, TEI-Lex0)’ (Roche, Papadopoulou, 2019).

La creazione di un'interfaccia comune, che integri aspetti concettuali e linguistici, trova la sua applicazione concreta in TEDI, uno strumento che consente di modellare la conoscenza attraverso il contributo congiunto dell'ingegneria e delle scienze umane (Roche, Papadopoulou, 2019).

Le ricerche in corso nell’ambito del progetto ITinHeritage trovano infine spazio, sempre in ottica onto-terminologica, nell’ambito scientifico legato alla valorizzazione e alla comunicazione del patrimonio culturale, ambito nel quale l’approccio onto-terminologico può rivelarsi di grande efficacia; il progetto partecipa quindi ai dibattiti che animano la comunità dei professionisti della gestione e della digitalizzazione del patrimonio, in relazione alla creazione e alla modellazione di ontologie, alla gestione dei dati nel Web semantico, all'interoperabilità dei formati e alla loro accessibilità al pubblico. In tale contesto, va citata la collaborazione in fase iniziale con Europeana (<https://europeana.eu/it>) e l’inserimento di ITinHeritage in un ampio filone di studi che hanno come obiettivo la modellizzazione dei saperi in ottica patrimoniale, al fine di

modellare, condividere e perennizzare la conoscenza in un ambiente accessibile al grande pubblico. L'approccio onto-terminologico è già stato applicato a progetti di gestione e digitalizzazione del patrimonio (Roche & Papadopoulo, 2019; Wei et al., 2022); il progetto ITinHeritage ne rappresenta l'applicazione al campo delle Tecnologie dell'Informazione.

#### **4 Metodologia per un approccio alla mediazione del patrimonio basato sul lavoro onto-terminologico**

Se il sapere delle TI è multiforme e sempre più immateriale, le TI implicano anche nuovi luoghi di conoscenza. Attraverso la costruzione di tali luoghi, mettiamo le TI al servizio della loro stessa rappresentazione. Questo processo inizia con la creazione del nostro corpus, nato dalla smaterializzazione patrimoniale.

In una prima fase un *knowledge graph*, sintesi del patrimonio delle TI, è stato sviluppato raccogliendo i metadati dei musei, che sono stati armonizzati (formati XML, PDF, ecc., nativi, vari campi), messi in standard CSV, RDF e FAIR, in modo da poter essere aperti e collegati sul web semantico (Linked Open Data (LOD)). Il *knowledge graph* è strutturato da un'ontologia di primo livello su cui si innesta l'ontologia del dominio TI. Contiene attualmente più di 25.000 artefatti ed è destinato a crescere. Per raggiungere questo obiettivo, abbiamo sviluppato un metodo per la gestione di big data eterogenei. Questo knowledge graph perpetua e apre il patrimonio delle TI tramite una piattaforma web. Altre piattaforme federano collezioni museali (Szekely et al, 2013 ; de Boer et al., 2013; Doerr, 2010; ...) o salvaguardano i softwares, come 'Software Heritage' (Di Cosmo, 2018). Tuttavia, se questi progetti valorizzano gli oggetti, noi valorizziamo i saperi. La piattaforma è intesa come un luogo per trasferire la conoscenza navigando nel grafo di conoscenze tramite un'ontologia di dominio e un dizionario multimediale.

In una seconda fase la costruzione dell'onto-terminologia inizia con un approccio semasiologico di linguistica testuale, attraverso l'estrazione lessicale dai metadati dei musei e l'analisi linguistica del nostro corpus. Lo studio della terminologia informatica è senza dubbio uno dei campi che più ha interessato i terminologi, che ne hanno approfondito gli aspetti morfologici e semantici (tra gli altri, si veda Claveau et L'Homme 2005), gli usi nel discorso specialistico e la dimensione testuale, con particolare riferimento alla linguistica dei corpora (per un primo studio di base, si veda Lefevre, Condamines, 2005), e le questioni legate al confronto interlinguistico (tra gli altri, Humbley, 2005). Particolarmente interessante nella nostra prospettiva è L'Homme, 2008, che confronta risorse terminologiche, ontologiche e generali nella descrizione della

terminologia informatica. La questione principale che ci interessa in questa fase del progetto è l'identificazione delle unità terminologiche e delle relazioni lessicali, che riflettono le denominazioni e l'organizzazione nozionistica della conoscenza del dominio. L'estrazione terminologica è stata condotta seguendo i criteri di selezione del testo ISO/DIS 5078:2023. Il nostro approccio di estrazione, applicato a tutte le lingue incluse nel progetto, può essere considerato semi-automatico. Ci siamo affidati a un approccio prevalentemente ibrido, combinando inizialmente tecniche statistiche e poi linguistiche. Partendo dal corpus francese, 'Description' e 'Use' sono gli unici due campi di metadati composti da più frasi testuali, mentre gli altri campi come 'Name' possono essere definiti come costituiti da un vocabolario più controllato. Una prima fase di analisi è rappresentata da un'estrazione terminologica automatica. Sono stati testati e confrontati diversi strumenti, tra cui Sketch Engine (Kilgariff et al., 2014), TermoStat (Drouin, 2003), e ci si è affidati anche al pacchetto NLTK del linguaggio di programmazione Python (Bird et al, 2009): per quanto riguarda i linguaggi di programmazione, sono ormai considerati alternative rilevanti agli strumenti di estrazione tradizionali quando si affrontano corpora derivati da big data (Anthony, 2021). Sketch Engine, strumento di gestione e analisi testuale per l'estrazione, si è rivelato l'opzione migliore. Questo lavoro ci ha permesso di evidenziare la terminologia del dominio, in particolare attraverso i termini complessi come 'Système mécanographique électronique Gamma 3A', 'Armoire des modules de connexion Gamma 30' o 'Tabulator BS120', e di osservarne le relazioni semantico-sintattiche come 'oggetto tipico', 'azione tipica' e 'agente tipico' (Lerat, 1988), a cui Gaudin (1996) aggiunge 'applicazione tipica', utili per trasmettere conoscenze complesse con l'obiettivo di costruire un dizionario. Esse elencano le collocazioni tipiche di un'unità lessicale per dare, attraverso questa combinazione, un'idea precisa dell'uso che si fa dell'oggetto nominato. Particolarmente adatte alla divulgazione scientifica, le relazioni semantico-sintattiche conducono il lavoro di analisi testuale verso la modellizzazione della conoscenza. Permettono di esprimere le pratiche che circondano l'oggetto. Contestualizzarne l'applicazione non può che favorire la comprensione, anche se questa descrizione non è sufficiente per l'acquisizione della conoscenza. Il percorso verso il concetto da comprendere è quindi supportato da una descrizione delle proprietà dell'oggetto, cioè un 'dare senso' e una delimitazione, con queste proprietà che rinviano alle caratteristiche del concetto. Attraverso questo lavoro di estrazione, abbiamo osservato anche le relazioni lessicali, come quelle di genere e partitive, che sono essenziali per l'acquisizione di nuove conoscenze e che la norma ISO

704 identifica come il fondamento del lavoro terminologico. Le relazioni di iperonimia, e soprattutto di meronimia, sono emerse grazie al rilievo dato al verbo nel discorso del patrimonio delle TI, verbi che Prandi e De Santis (2019) segnalano come indicatori di relazioni concettuali.

Questo approccio semasiologico si traduce infine in lessici, ma che riflettono il corpus iniziale e mettono in evidenza la varietà linguistica di un dominio piuttosto che risolverla. Non si tratta ancora di una concettualizzazione consensuale. E tuttavia, se siamo interessati al modo in cui gli esperti nominano il loro dominio, siamo altrettanto interessati al modo in cui lo concettualizzano. La nostra esperienza ci ha mostrato quanto sia difficile elaborare una rete semantica solo sulla base di questo lavoro semasiologico (Djambian, 2011): si presenta infatti il problema di identificare solo la varietà delle denominazioni dei concetti del dominio nei testi, e non i concetti stessi. Questo tipo di studio, pur essendo essenziale per gettare le basi del lavoro terminologico, si ferma al significato delle parole osservate nel discorso, definito in base ai loro usi. Per compensare le variazioni linguistiche e far emergere un senso comune è opportuno interessarsi alla parte extralinguistica del lavoro terminologico, incentrata sulla relazione del concetto con l'oggetto. Il concetto è inteso come indipendente dagli usi delle sue denominazioni ed è definito come consensualmente standardizzato all'interno di una comunità.

‘Occorre ricordare che tutto lavoro terminologico dovrebbe basarsi sui concetti e non sui termini’ (Felber, 1987).

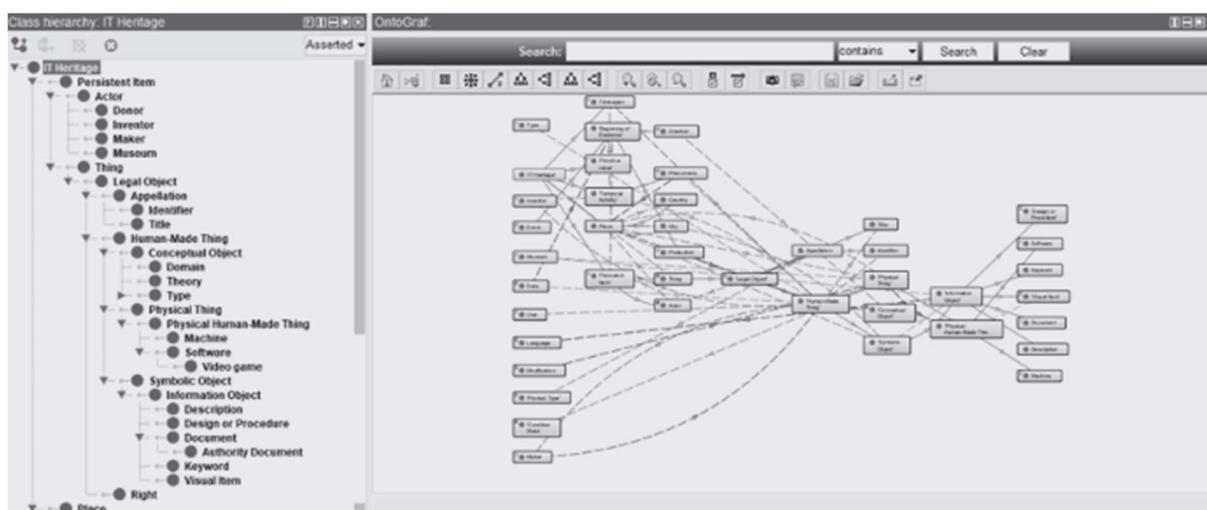
I termini sono interessanti per il terminologo perché denotano un concetto in una sorta di forma biunivoca, un ponte che riflette la dinamica continua tra la parte linguistica del mondo reale e la parte extralinguistica del mondo simbolico. Il mondo nozionale è la traduzione del modo in cui percepiamo gli oggetti nel mondo reale. La standardizzazione richiesta dal lavoro terminologico permette di fissare questo sistema nozionale consensuale all'interno di una comunità di pratica.

Per tenere conto delle specificità e delle variazioni del sistema linguistico, il sistema nozionale è più facilmente costruibile con un approccio onomasiologico. In questa formalizzazione basata sulla convenzione nel senso latino di ‘foedus’, i nomi dei concetti devono essere situati all'interno del sistema nozionale nel suo complesso, anche se ciò significa renderli artificiali per fornire una significazione che vada oltre l'uso. Ciò può essere ottenuto con l'aiuto di esperti che forniscono ‘prototipi’ concettuali raggruppando

le loro caratteristiche essenziali. Nel fare ciò, strutturare i concetti secondo relazioni di sussunzione e di differenza secondo una differenziazione specifica (logica aristotelica), ci permette di costruire un vocabolario standardizzato che sfuma le varietà e le ambiguità del linguaggio naturale. Le relazioni linguistiche citate in precedenza e osservate nel nostro corpus, soprattutto la relazione di genere e la relazione merologica, consentono una descrizione dettagliata degli oggetti in quanto forniscono l'ambiente nozionale (caratteristiche specifiche o costituenti di un oggetto).

Il nostro lavoro consiste quindi nel modellare il sapere tramite un'ontologia che rifletta la concettualizzazione del mondo delle TI. Mentre esistono modelli sulla conoscenza informatica (Salatino et al., 2020; Cassel et al., 2013), non ne esistono per il patrimonio delle TI. La costruzione di questa onto-terminologia è facilitata dall'uso sia dell'ambiente Protégé per la strutturazione del *knowledge graph*, sia dell'ambiente TEDI, più completo di Protégé per la creazione di onto-terminologie secondo gli standard ISO 704 e 1087 (Roche, Papadopoulou, 2019) e che utilizziamo per costruire il dizionario.

Nel campo del patrimonio culturale, esistono diversi formalismi del web semantico per descrivere gli oggetti culturali e le loro relazioni. Da parte nostra, abbiamo basato la costruzione dell'onto-terminologia su CIDOC-CRM (Conceptual Reference Model) (Le Boeuf et al., 2015) che ci permette di strutturare il *knowledge graph*, allineandolo al Data Model di Europeana (EDM), di cui siamo partner, e OWLTime.



Tuttavia, le ontologie top-level come CIDOC non sono abbastanza granulari per rappresentare la conoscenza dello specifico dominio delle TI. Il lavoro onomasiologico con gli esperti è quindi essenziale per la costruzione dell'ontologia. Con il loro aiuto e

grazie all'uso di OWLTime, stiamo adottando un approccio diacronico, che ci permette di seguire la storia e l'evoluzione dei manufatti e dei concetti associati nel tempo. Ad esempio, le macchine sono concepite in generazioni, partendo dai primi dispositivi meccanici di calcolo risalenti all'antichità, alle macchine da calcolo meccaniche, all'automazione, ... ai computer a programma memorizzato, dai primi ai più recenti, come i computer basati sull'intelligenza artificiale e le innovazioni attuali come il computer quantistico. È anche essenziale includere le teorie, basi dello sviluppo tecnologico. Questo permette di comprendere l'ampiezza e la varietà del dominio delle TI. Indagando la natura degli oggetti che popolano il mondo circoscritto delle TI, possiamo dedurre le caratteristiche eidetiche dei concetti che essi oggettivano e le caratteristiche differenziali da un concetto all'altro, per costruire gradualmente la rete nozionale e il dizionario del dominio.

L'ontologia è resa scalabile, un prerequisito in un campo così mutevole, utilizzando Natural Language Processing, clustering e Artificial Intelligence per individuare nuovi termini sulla base di web dataset (Wiktionary, BDPedia) e attingendo a lavori precedenti (Salatino et al., 2020). L'applicazione di metodi automatizzati per il completamento di ontologie costruite manualmente ci permette di tenere il passo con gli sviluppi tecnologici delle TI e di osservare le mutazioni della terminologia e dei concetti, quindi delle rappresentazioni all'interno della nostra società nel tempo.

## **5 Conclusione**

Interrogare e valorizzare il patrimonio formale e tacito, materiale e immateriale delle Tecnologie dell'Informazione è un compito maieutico (Bortolotto, 2011), che solleva anche questioni epistemologiche, sia per quanto riguarda il lavoro terminologico, sia per quanto riguarda la nostra società, attraverso i suoi cambiamenti, quelli del suo patrimonio culturale e scientifico, e le rappresentazioni che ne derivano.

Il patrimonio, peraltro, non è solo una questione di oggetti raccolti, in una sacralità dell'oggetto teatro della memoria, ma è un movimento, materializzato, più che in oggetti, in pratiche e nelle loro espressioni discorsive (UNESCO, 2022). Quindi, non dovrebbe essere definito dalla sua natura tangibile o intangibile, ma dal modo in cui viene utilizzato e espresso. E' un processo culturale in cui si costruiscono discorsi e senso. Il patrimonio è un modo di vedere, sapere e sentire. Tratta della costruzione di un'identità (Smith, 2011) in continua evoluzione, attraverso rappresentazioni labili del mondo, che cerchiamo di

studiare, preservare, rappresentare e trasmettere, attraverso il nostro lavoro, nel quadro fondante delle nostre società, e delle scienze e tecniche che raccontano la storia dell'umanità: le Tecnologie dell'Informazione.

## 6 Bibliografia

Ackerman W., *Etudes sur la diffusion et la représentation sociale des connaissances scientifiques*, 'Sociologie du travail', n° 4, 1963, p. 415-417.

Anthony L., *Programming for corpus linguistics*, in M. Paquot, S. T. Gries, *A practical handbook of corpus linguistics*, Cham, Springer, 2021, p. 181–207.

Bird S., E. Klein, E. Loper, *Natural language processing with Python: Analyzing text with the Natural Language Toolkit*, Sebastopol, O'Reilly Media, 2009.

Bortolotto C., *Le patrimoine immatériel et le tabou de l'authenticité: de la pérennisation à la durabilité*. 'Le Patrimoine culturel immatériel au seuil des sciences sociales', 2012. p. 218-235.

Cassel L. N., et al., *The new ACM CCS and a computing ontology*, in J. F. Downie. et al., *Atti del 13° ACM/IEEE-CS Joint Conference*, Indianapolis, USA, Association for Computing Machinery, 2013, p. 427-428.

Claveau V., L'Homme M. C., *Apprentissage par analogie pour la structuration de terminologie-utilisation comparée de ressources endogènes et exogènes*, in A. Baneyx, V. Malaisée, *Atti della TIA conference*, Rouen, France, 2005.

de Boer, V., et al., *Amsterdam museum linked open data*, 'Semantic Web', 4, 2013, p. 237-243.

Di Cosmo, R., *Software heritage: why and how we collect, preserve and share all the software source code*, in M. Chechik et al., *Atti del IEEE/ACM 40th international conference on software engineering*, Gothenburg, Sweden, IEEE, 2018.

Djambian C., et al., *New terminological approaches for new heritages and corpora: The ITinHeritage project*, *Atti del 3° MDTT international conference*, Grenada, Spain, 2024, to be published.

Djambian C., Rossi M., D'Ippolito G., *La médiation des objets aux savoirs scientifiques et techniques*, in C. Roche, *Atti del TOTh conference*, Chambéry, France, 2023, to be published.

Djambian C., *Le métier : son savoir, son parler*, in C. Roche, *TOTh 2011-Terminologie & Ontologie*, Annecy, France, Institut Porphyre, 2011.

- Doerr M., et al., *The Europeana Data Model (EDM)*, in *Atti del World Library and Information Congress: IFLA 76*, Gothenburg, Sweden, IFLA, 2010, p. 1-12.
- Drouin P., *Term extraction using non-technical corpora as a point of leverage*, *Terminology*, 'International Journal of Theoretical and Applied Issues in Specialized Communication', 9, 2003, p. 99–115.
- Faber P., L'Homme M. C., *Theoretical Perspectives on Terminology. Explaining terms, concepts and specialized languages*, Amsterdam/Philadelphia, John Benjamins, 2022, p. 237–259.
- Felber H., *Manuel de terminologie*, UNESCO, 1987.
- Gaudin F., *Terminologie: l'ombre du concept*, 'Meta', 41, 1996, 604-621.
- Haigh T., Ceruzzi E., *A New History of Modern Computing*, Cambridge, The MIT Press, 2021.
- Humbley J., *La traduction des métaphores dans les langues de spécialité: le cas des virus informatiques*, 'Linx. Revue des linguistes de l'université Paris X Nanterre', 2005, p. 49-62.
- Jacob C., *Rassembler la mémoire*, 'Diogène', 4, 2001, p. 53-76.
- Kant E., *Critique de la raison pure*, Paris, Aubier, 1997.
- Kilgarriff A., et al., *The Sketch Engine : Ten years on*, 'Lexicography', 1, 2014, p.7–36.
- Le Boeuf P., et al., *Definition of the CIDOC Conceptual Reference Model (Version 6.2.1 éd.)*, 'ICOM/CIDOC Documentation Standards Group, CIDOC CRM Special Interest Group', 6, 2015.
- Lefevre L., Condamines A., *MAR-REL : une base de marqueurs de relations conceptuelles pour la détection de Contextes Riches en Connaissances*, in I. Eshkol-Taravella, J. Y. Antoine, *Atti della 24ème Conférence sur le TALN, Orléans, France, ATALA*, 2017, p. 183–191.
- Lerat P., *Terminologie et sémantique descriptive*. 'La banque des mots', 36, 1988.
- L'Homme M. C., *Ressources lexicales, terminologiques et ontologiques: une analyse comparative dans le domaine de l'informatique*, 'Revue française de linguistique appliquée', 1, 2008, p.97–118.
- Maret P. et al., *QAnswer KG: designing a portable question answering system over RDF data*, in A. Harth et al., *Atti del Semantic Web: 17th International Conference, ESWC*, Heraklion, Crete, Greece, Springer, 2020. p. 429-445.
- Meyer I., *Concept management for terminology. A knowledge engineering approach*, in P. Faber, M. C. L'Homme, *Theoretical Perspectives on Terminology. Explaining*

- terms, concepts and specialised languages*, Amsterdam/Philadelphia, John Benjamins, 2022, p. 110–126
- Mounier-Kuhn, P., Lazard, E., *Histoire illustrée de l'informatique*, 3<sup>e</sup> ed., Paris, EDP Sciences, 2022.
- Prandi M., De Santis C., *Manuale di linguistica e grammatica italiana*, Torino, UTET, 2019.
- Putnam H., *Raison, vérité et histoire*, Paris, Editions de minuit, 1984.
- Roche C., Papadopoulou M., *Mind the Gap: Ontology Authoring for Humanists*, in A. Barton, *Atti del JOWO: The Joint Ontology Workshops*, Graz, Autriche, JOWO, 2019, p.1-12.
- Rossi M., *Termes et métaphores, entre diffusion et orientation des savoirs*, 'La linguistique', 57, 2021, p. 153-173.
- Salatino A., et al., *The computer science ontology: A comprehensive automatically-generated taxonomy of research area*, 'Data Intelligence', 2, 2020, p. 379-416.
- Schiele B., *Ethique et patrimoine scientifique : Regards sur l'évolution des musées de sciences et technologies*, 'Ethique publique', 19, 2017, URL: <https://journals.openedition.org/ethiquepublique/3036> .
- Smith L., *Heritage and its intangibility*, in A. Skounti, O. Tebba, *De l'immatérialité du patrimoine culturel*, Rabat, UNESCO, 2011.
- Szekely P., et al., *Connecting the Smithsonian American Art Museum to the Linked Data Cloud*, in P. Simiano, *Atti del Semantic Web: Semantics and Big Data, ESWC*, Berlin, Germany, Springer, 2013, p. 593-607.
- Temmerman R., *Towards new ways of terminology description*, Amsterdam/Philadelphia, John Benjamins, 2000.
- UNESCO, *Technologies de l'Information et de la Communication (TIC), Glossaire*, 2024, URL: <https://uis.unesco.org/fr/glossary-term/technologies-de-linformation-et-de-la-communication-tic>
- UNESCO. *Textes fondamentaux de la Convention pour la sauvegarde du patrimoine culturel immatériel*, Paris, UNESCO, 2022.
- Wei T., et al., *Using ISO and Semantic Web standard for building a multilingual terminology e-Dictionary: A use case of Chinese ceramic vases*, 'Applied Ontology', 17, 2022, p. 423-441.