



**HAL**  
open science

# La médiation des objets aux savoirs scientifiques et techniques, terminologie et représentation ontologique du patrimoine des Technologies de l'Information : le projet de recherche ITinHeritage

Caroline Djambian, Micaela Rossi, Giada d'Ippolito

## ► To cite this version:

Caroline Djambian, Micaela Rossi, Giada d'Ippolito. La médiation des objets aux savoirs scientifiques et techniques, terminologie et représentation ontologique du patrimoine des Technologies de l'Information : le projet de recherche ITinHeritage. Conférence Terminologie et Ontologie : Théories et applications (TOTh) 2023, Christophe ROCHE; Caroline DJAMBIAN, Jun 2023, Le Bourget du Lac, France. hal-04604971

**HAL Id: hal-04604971**

<https://hal.univ-grenoble-alpes.fr/hal-04604971v1>

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.

# **La médiation des objets aux savoirs scientifiques et techniques, terminologie et représentation ontologique du patrimoine des Technologies de l'Information : le projet de recherche *ITinHeritage***

Caroline Djambian\*, Micaela Rossi\*\*, Giada D'Ippolito\*\*\*

\*Université Grenoble Alpes, Laboratoire GRESEC (Groupe de recherche sur les enjeux de la communication)

[caroline.djambian@univ-grenoble-alpes.fr](mailto:caroline.djambian@univ-grenoble-alpes.fr)

\*\*Università di Genova, Dipartimento di Lingue e culture moderne

[micaela.rossi@unige.it](mailto:micaela.rossi@unige.it)

\*\*\*Università di Genova, Dottorato in Digital Humanities

[giadadippolito30@gmail.com](mailto:giadadippolito30@gmail.com)

**Résumé.** La sauvegarde du patrimoine des Technologies de l'Information (TI) est une démarche d'intérêt général. Leur évolution trop rapide a laissé peu de place à l'étude et à la valorisation. Le projet de recherche *ITinHeritage* se propose de comprendre comment pérenniser et médier les savoirs scientifiques et techniques contemporains, partant de ce patrimoine particulier. Notre approche part des musées de l'informatique. Avec nos partenaires : Musée des Arts et Métiers, ACONIT (France), London Science Museum (Angleterre), NAM-IP (Belgique), Museo degli Strumenti per il Calcolo (Italie), Homecomputermuseum (Hollande), HNF (Allemagne), nous développons une approche innovante en Humanités Numériques par la création d'un graphe de connaissances soutenu par une terminologie et une ontologie du domaine intégrant de grands projets européens (Europeana, CIDOC-CRM, OntoMe), accessible via une plateforme sur le web sémantique (Linked Open Data (LOD)), s'appuyant sur l'analyse amont et aval des mécanismes de transfert des savoirs scientifiques et techniques.

## **1. Introduction**

Les technosciences ne se contentent pas de nous rendre le monde plus intelligible, elles le transforment et l'impactent dans des proportions et à une vitesse sans précédent. Les Technologies de l'Information (TI) sont l'emblème de ce patrimoine scientifique et technique contemporain. L'UNESCO les définit comme « *l'ensemble*

*d'outils et de ressources technologiques permettant de transmettre, enregistrer, créer, partager ou échanger des informations, ... »* (UNESCO, 2023). Leur ancrage social et leur étendue en font ce que Price nomme une “*big science*” (Price, 1963). La sauvegarde et la valorisation de leur patrimoine sont devenues une démarche d'intérêt général. Mais leur évolution trop rapide n'a guère laissé le temps à l'étude, à la valorisation et à une réelle patrimonialisation. L'étude de ce patrimoine nous permettra de le définir, d'observer ses représentations et d'agir sur ses médiations et appropriations (Schiele, 1989). Notre projet *ITinHeritage* part des espaces muséaux des TI. Le Musée des Arts et Métiers et le Conservatoire de l'Informatique (ACONIT) (France), le London Science Museum (Royaume-Uni), le NAM-IP (Belgique), le Museo degli Strumenti per il Calcolo (Italie), le Homecomputermuseum (Pays-Bas) et le HNF (Allemagne) participent à notre réseau. Si ces musées de science n'ont cessé d'évoluer sous la pression de la relation société/science qui remet en cause les modes de préservation et de communication du patrimoine scientifique, les musées scientifiques sont à présent considérés comme des espaces devant non seulement exposer la science, mais aussi mettre en scène le savoir scientifique et technique pour les publics, dans des termes passant du formel à l'informel (Wagenberg, 2006).

Nous souhaitons par le projet de recherche *ITinHeritage*, les aider à proposer de nouvelles formes d'accès aux savoirs portés par leurs collections et particulièrement adaptées à leur médiation, via une approche innovante en Humanités Numériques. Premièrement, par l'observation des mécanismes de transfert des savoirs scientifiques et techniques experts au sein des pratiques. Puis, par la mise en place de dispositifs numériques promouvant l'accès, le transfert et l'appropriation des savoirs. Pour ce faire, nous construisons une plateforme dédiée au patrimoine des TI sur le web sémantique permettant l'ouverture et la mise en relation des artefacts avec d'autres collections à travers le monde (Linked Open Data (LOD)). Cette plateforme donne accès à un graphe de connaissances construit sur l'environnement Wikibase qui permet une standardisation (CSV et RDF) et un stockage plus aisé de gros volumes de données hétérogènes (actuellement plus de 25500 artefacts constituent la base). Le graphe est structuré par une ontologie générique du domaine patrimonial, construite sur Protégé et basée sur les modèles CIDOC CRM et Europeana Data Model (EDM). Cette ontologie de premier niveau sera complétée par une terminologie et une ontologie du domaine des TI. Mais ce ne sont pas ici les considérations techniques que nous développerons. Nous nous pencherons dans un premier temps sur les spécificités de la médiation de ce patrimoine où langage de spécialité et conceptualisation consensuelle jouent un rôle prépondérant dans le transfert des savoirs. Cela nous portera à élargir notre réflexion aux fondements des travaux termino-ontologiques. Enfin, nous présenterons les premières étapes et perspectives du travail termino-ontologique qui est au cœur de nos travaux.

## 2. Les Technologies de l'Information : des technosciences en médiation

La médiation culturelle a des fondements théoriques larges, allant de la théorie des espaces publics (Allard-Chaniai, 1998), du « *tiers symbolisant* » (Py, Debruyne, Vandiedonck, 2002) ou de la « *traduction* » au sens de Latour (Caillet, 1995). Jean Davallon (2003) la définit comme visant à construire une interface entre public et objet culturel et leurs deux univers étrangers, pour permettre une appropriation du second par le premier. Elle est lieu symbolique de rencontre public-science. Elle est aussi un concept opératoire recouvrant, comme nous le verrons, des applications diverses, mais impliquant toujours un élément tiers et produisant toujours un effet sur son destinataire : accès, apprentissage, appropriation ... Il s'agit donc d'un transfuge d'un contexte avec modification de l'élément qui en fait l'objet, que l'action soit humaine ou objectivée par un dispositif. La médiation dépasse donc l'interaction pour induire une transformation de l'environnement où elle se déroule (par exemple social). Dans « *Les figures de l'amateur* » Hennion (2000) souligne que la médiation enclenche un processus, un événement, un passage profondément modificateur. Il s'agit donc de reconnaître ce moment « *dans ce qu'il a de spécifique et d'irréversible, de le voir comme transformation, travail productif* ». Dans la médiation, la communication intersubjective se déplace de ses actants vers un « *tiers symbolisant* » aux formes diverses. Nos observations nous permettent de dissocier quatre formes, quatre conceptions de communication, mobilisées par les espaces muséaux.

### 2.1. La médiation du patrimoine scientifique et technique

Les TI conquièrent aujourd'hui de plus en plus les sphères patrimoniales (Poli, 2012 ; Dufrêne, *et al.*, 2013). D'autres projets pionniers du web sémantique ont fédéré les collections des musées pour les croiser avec d'autres à l'échelle internationale (Dijkshoorn *et al.*, 2018 ; Juanals Minel, 2016 ; Szekely *et al.*, 2013 ; de Boer *et al.*, 2013 ; Cobb, 2015), à commencer par Europeana (Freire, Isaac, 2019). En modifiant profondément la manière dont les savoirs sont diffusés, ces nouvelles ressources ont un impact fort sur l'avenir des institutions et du patrimoine qu'elles soutiennent, mais aussi sur le plan sociétal. Les TI instaurent ainsi de nouvelles médiations aux dimensions à la fois techniques « *car l'outil utilisé structure la pratique* » et sociales « *car les mobiles, les formes d'usages et le sens accordé à la pratique se ressource dans le corps social* » (Jouët, 1993). Ces dispositifs objectivent et technicisent le processus de communication (médiation technique) et mobilisent dans le même temps une dimension subjective de la pratique (médiation sociale). Pourtant, ce n'est pas la médiation par les TI qui nous intéresse dans ces pages, mais celle de leur patrimoine, notamment de leurs savoirs, car ouvrant un champ inexploré. Elle passe par des cristallisations matérielles et immatérielles.

## La médiation des objets aux savoirs scientifiques et techniques

La première forme matérielle de médiation est l'objet de science. Peu patrimonialisé car trop récent, il est peu pris en compte par les institutions. L'infiniment petit, la dématérialisation des savoirs, l'obsolescence trop rapide qui leur donne une profondeur historique, le caractérisent et le rendent emblématique du patrimoine récent. En termes de médiation et de transmission, sa dimension technologique est forte et ses objets matériels offrent peu d'apports visuels : leur forme n'est pas médiatrice en soi contrairement aux objets d'art. Il nous faut donc rendre significatifs ces objets qui ne le sont pas de prime abord.

La deuxième forme matérielle de médiation est l'espace muséal, intrinsèquement lié à l'évolution de l'objet. Le patrimoine immatériel se développant aujourd'hui autant que le patrimoine matériel, les musées scientifiques doivent de plus en plus montrer l'immontrable, l'intangible. Dans le cas de TI, les schémas muséaux classiques sont calqués au domaine technique et l'effort de médiation est encore très axé sur la gestion des collections, ce qui ne se prête généralement pas à la médiation de ses objets. En réponse à ce point, il nous faut déterminer une médiation qui réorientera la stratégie muséale de l'objet vers sa réception et son appropriation sociale, et vers la mission de transmission de savoirs experts des musées de science.

La troisième forme de médiation est mixte, car le patrimoine immatériel va croissant et les données sont les nouvelles collections de la science moderne. Ce patrimoine immatériel est lui-même multiforme : l'objet et la documentation numérisés, les documents digitaux multimédias, les logiciels (1/3 à la totalité de certaines collections) et les données. Dans le cas des musées des TI, la documentation et les données sur les objets sont aussi importantes que ces derniers. Elles concentrent les connaissances autour des objets, démontrent leur valeur patrimoniale et les incarnent matériellement et symboliquement. Elles les rendent intelligibles, accessibles et légitiment l'institution qui les détient, car symbolisent et concentrent l'expertise sur les objets (Després-Lonnet & Rizza, 2021). La documentation et les données participent donc à la muséalisation et à la patrimonialisation des objets de science et interviennent directement dans la médiation des savoirs. Elles peuvent servir de base pour rendre visibles, lisibles et transférables les savoirs du domaine.

La quatrième forme de médiation est immatérielle : le langage de spécialité. Un musée des sciences contribue à reformuler le discours scientifique dans le champ muséal et à l'inscrire dans l'espace public. Que la transmission des savoirs soit horizontale (entre pairs), descendante (vulgarisation) ou ascendante (appropriation par le public), c'est le langage qui est au cœur de leurs processus de formation et d'appropriation. L'étude des diversités langagières (Djambian, 2012), du rôle sémiotique des images et tropes (Rossi, 2021) afin d'en détecter les mécanismes, nous permettra d'observer comment et quand les technosciences peuvent devenir des représentations sociales et quel rôle jouent les langues de spécialité dans ce processus.

## 2.2. La médiation des savoirs scientifiques et techniques

Les médiations patrimoniales que nous venons de voir (objet, musée, document) recourent à des réalités matérielles qui objectivent des savoirs experts figés dans l'écrit ou l'objet et que Jacob (2001) nomme « *exotériques* » ou Popper (1979) « *connaissance représentée sans sujet connaissant* ». Le langage, qui en est la quatrième forme, permet quant à lui une réactivation de ces savoirs dans l'espace public, pour leur transmission et appropriation. Il est lieu de savoir. Pour Lamizet (1992) la médiation met en place une dialectique entre le singulier (l'objet de science ou l'expert qui y a inscrit ses savoirs), et l'espace social. La nature du « *tiers symbolisant* » tient de la réflexivité et de la représentation au sens Lacanien, où le sujet est dédoublé dans le langage et le social dédoublé dans la convention consensuelle ou *foedus*. Penser la dialectique entre ces deux sphères est tout le propos du projet *ITinHeritage*. Notre rôle est d'instaurer une médiation permettant le passage des savoirs scientifiques et techniques vers l'espace social. Ce transfert n'est possible que par le symbolique du langage, matrice, « *neutre de la communication* ». Le *foedus* quant à lui, nous porte vers le monde de référence de Caune (1995). Pour lui, la médiation culturelle est « *permutation circulaire* » d'une manifestation en tant que fait perceptible (ex : la monstration physique ou numérique de l'objet muséal), d'un individu et d'un monde de référence, soit le « *cadre culturel et social dans lequel la manifestation prend sens* ». Quéré (1982) rappelle lui, que « *l'échange social est interaction entre sujets, médiatisée par du symbolique* », et Davallon (2003) de noter que « *la médiation se construit autour d'un point de fuite – appelé extériorité, neutre, négatif, c'est selon – qui intervient dans le processus de communication sans que ceux qui y participent puissent avoir prise sur lui* ».

Nous voyons à travers ces auteurs que deux niveaux symboliques coexistent dans le lieu de médiation : celui du langage, encodage des savoirs qui appartient au singulier, et celui des références consensuelles, d'une conception commune du monde, qui appartient au politique – Gaudin parle du « *phénomène social de la référence* » (2012), structurées par une convention propre à chaque communauté et permettant le décodage des savoirs pour un retour *in fine* au singulier. L'appropriation des savoirs passe donc, de façon encore plus prégnante dans des environnements hautement techniques, par l'intégration de la conceptualisation et du langage communs au domaine. Nos travaux prendront en compte ces deux niveaux par le travail terminologique et ontologique qui favoriseront l'imprégnation des représentations consensuelles à travers l'usage de lieux et signifiants de médiation.

La notion de culture présentée par Crespi (1983) motive d'autant plus la nécessité d'orienter le public vers une représentation consensuelle du monde : « *La culture, en tant que dimension anthropologique, peut-être considérée dans l'ordre du vivant comme le résultat évolutif de la complexité croissante des modes de relation et de communication intersubjectifs et intermondains* ». Pour Bayard (2012) la culture se fonde sur la connaissance des discours au-delà des connaissances expérientielles de

*l'emperia* de Kant (1997) ou de la *métis* d'Aristote (1986). La culture scientifique introduit ainsi l'expérience des réalités par le discours qui les valide, les confirme. Cependant, les connaissances et représentations autour d'un terme sont requises, mais non nécessaires pour en déterminer le référent selon Putnam (1984). Il dissocie donc l'intégration sociale d'un terme, de sa référenciation effective dans le collectif. Le *foedus*, la convention, est ici indispensable pour que l'usage du terme participe d'une véritable appropriation des connaissances. Putnam répartit ces deux niveaux d'usages en deux catégories d'individus locuteurs (c'est la division du travail linguistique qui structure l'échange au sein des communautés) : l'individu *lambda* et l'expert, qui occupe une position et responsabilité socio-linguistique particulière, car seul capable de dire réellement ce que sont les choses. Selon Gaudin (2012), les experts portent une responsabilité « envisagée à deux niveaux, épistémique (liée au savoir lui-même) et sémantique liée à la définition des termes utilisés ». Effectivement, pour Putnam, le sens n'est pas la représentation intellectuelle d'un individu, « les significations ne sont pas dans la tête » (Gaudin). Nous en déduisons donc plusieurs axes à prendre en compte pour nos travaux. Premièrement, pour représenter les savoirs du domaine des TI, le travail terminologique ne suffira pas et ne sera que l'*a priori* du travail ontologique. Deuxièmement, les discours ne sont pas porteurs en eux-mêmes des notions dont ils parlent. Comme le rappelle Rastier (2006), on ne trouve pas de concepts dans les textes. La sémiologie de l'image et toutes les connaissances accumulées autour des objets (documents multimédia, données) sont à prendre en compte dans l'appropriation des références communes. Ensuite, l'approche sémasiologique ne sera pas suffisante et devra être corrélée avec une approche onomasiologique. Enfin, les experts (scientifiques et amateurs) occuperont une place cruciale à chaque étape de nos travaux, puisqu'ils seuls capables de donner les bons termes (au sens originel de limites) au collectif.

### **3. L'approche termino-ontologique au cœur de la transmission des savoirs scientifiques et techniques**

Le projet qui est présenté dans ces pages se situe dans une perspective interdisciplinaire au carrefour de la terminologie, de la gestion des connaissances et du patrimoine. Du point de vue théorique et méthodologique *ITinHeritage* trouve sa place à l'intérieur du panorama des études actuelles en terminologie par : la valorisation des savoirs professionnels qui sont liés aux contextes spécialisés, l'attention attribuée à la définition et à la gestion des connaissances expertes ainsi qu'aux patrimoines matériels et immatériels qui forment notre corpus, les apports fondamentaux de l'ingénierie des connaissances à l'analyse des concepts impliqués.

La centration sur la modélisation des concepts et des connaissances est par ailleurs l'un des paradigmes dominants des études terminologiques après 2000, avec

l'importance croissante attribuée aux disciplines connexes (de l'informatique au *knowledge management*). Le volume récent dirigé par Faber et L'Homme (2022) dessine l'état de l'art de la discipline, identifiant les lignes directrices de la recherche en terminologie au XXIème siècle, parmi lesquelles l'importance des influences réciproques entre terminologie et gestion des connaissances occupe un rôle de premier plan. Le chapitre de Nuopponen notamment, qui retrace l'évolution des études sur les relations conceptuelles, mais également la nouvelle publication de l'essai de Meyer (1992) s'intéressant à la constitution de bases de connaissances terminologiques, visent à une centration puissante sur les aspects de la formalisation conceptuelle, un « *retour au concept* », une reprise des postulats à la base de la théorie wüstérienne, après une phase fortement focalisée sur la variation discursive (cf. Humbley et Candel, dans le même volume), retour clairement explicité par Elena Montiel-Ponsoda sur l'interaction entre terminologie et ontologie. Le lien indissoluble entre ontologie et terminologie n'est pas une thématique nouvelle, comme le prouve l'abondante bibliographie dans ce domaine depuis le « *cognitive turn* » des années 2000 et l'avènement des approches cognitives (cf. l'ouvrage fondateur de Temmerman, 2000, et sa formulation de l'unité « *UoU unit of understanding* » qui exigent une prise en compte des terminologies dans leur nature conceptuelle, de *savoirs* spécialisés).

D'un point de vue méthodologique, *ITinHeritage* se réclame en premier lieu de l'approche termino-ontologique théorisée par Christophe Roche et l'équipe Condillac (<http://ontoterminology.com/>). Depuis les années 2000 (Roche, 2005), ce paradigme termino-ontologique propose la valorisation de l'analyse conceptuelle dans les savoirs spécialisés<sup>1</sup>, conjuguant les principes de l'analyse ontologique dans le domaine philosophique avec ses avatars les plus récents, liés à la création d'ontologies dans les domaines des sciences de la communication et de l'intelligence artificielle. La possibilité de corréler le niveau conceptuel et le niveau linguistique par un environnement technologique *ad hoc* « *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 et Papadopoulou, 2020). Les recherches du groupe Condillac, à l'origine des rencontres annuelles TOTh<sup>2</sup>, rassemblent des spécialistes de knowledge management et terminologues identifiant un besoin commun : « *[...] l'importance, voire la nécessité, de disposer d'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, 2020). La création d'une interface commune, intégrant les aspects conceptuels et linguistiques, trouve son application concrète dans TEDI, outil qui permet la modélisation des connaissances par l'apport conjoint des ingénieurs des

<sup>1</sup> En complément à la *termontographie* de Temmerman (voir Humbley 2022 : 32).

<sup>2</sup> Disponibles à l'adresse <https://toth.fr.condillac.org/actes> (cons. le 24/09/2023).

connaissances ainsi que des humanistes, trop souvent exclus des processus de formalisation.

L'approche termino-ontologique s'avère particulièrement efficace dans les contextes de communication et digitalisation du patrimoine culturel. *ITinHeritage* s'insère également dans la tradition de valorisation et divulgation du patrimoine auprès des publics par des applications informatiques, participant aux débats qui animent la communauté des professionnels en gestion et digitalisation des patrimoines, en relation à la création et modélisation d'ontologies, à la gestion des données dans le Web sémantique, à l'interopérabilité des formats et à leur accessibilité par les publics.

La création d'ontologies dans le domaine patrimonial – notamment, muséal – est un enjeu fondamental afin de modéliser les connaissances, les partager et les pérenniser dans un environnement accessible au grand public. La création de standards d'interopérabilité devient dans ce contexte une exigence incontournable, comme l'attestent de nombreuses études et projets, parmi lesquels Du Château *et al.* (2020) citent Courage (Faraj & Micsik 2021), WarSampo (Koho, *et al.* 2019), ArchOnto (Koch, Ribeiro & Lopes 2020), MMM (Mapping Manuscript Migrations) et MémoMines (Kergosien, Eric, *et al.* 2019). L'approche termino-ontologique a déjà été appliquée à des projets de gestion et digitalisation du patrimoine (Piccini, 2016 ; Roche et Papadopoulo, 2020 ; Wei *et al.*, 2020) ; le projet *ITinHeritage* se propose maintenant de l'appliquer au domaine des Technologies de l'Information. La nature profondément hétérogène de cette discipline, son évolution constante et multiforme pour ce qui est des langages et des supports, la stratification des publics et des discours, la présence de traditions linguistiques et culturelles diverses, mais surtout la nécessité d'adopter une approche méthodologique réellement interdisciplinaire dans l'esprit des Humanités Numériques, constituent à présent autant de défis pour l'équipe de recherche.

## **4. Naissance d'un travail termino-ontologique sur le domaine des Technologies de l'Information**

### **4.1. Méthodologie de constitution de corpus et extractions lexicales**

L'un des objectifs dans la création d'une termino-ontologie est la conception d'un réseau lexical à partir des données de corpus. Nous décrivons ici cette première étape de l'élaboration de la termino-ontologie, à travers le développement d'une analyse de corpus multilingue<sup>3</sup>. Notre approche étant résolument patrimoniale, nos travaux se

---

<sup>3</sup> Il s'agit de créer un dictionnaire quadrilingue (français, anglais, italien, allemand) et multimédia sur le modèle des premiers dictionnaires scientifiques et techniques qui ont fondé

basent principalement sur les collections des espaces muséaux partenaires rassemblés dans le graphe de connaissances. Afin de compléter ce corpus conséquent de plus de 25500 items amenés à augmenter, et d'éviter les biais induits par les représentations et spécificités propres à chaque musée, nous avons choisi de travailler également sur un sous-corpus représentant la terminologie déjà normée du domaine, constitué d'ouvrages d'histoire de l'informatique<sup>4</sup>. Ce corpus complémentaire a été choisi avec les experts des différents musées partenaires par le biais d'entretiens semi-structurés. Ce premier travail lexical nous permettra également d'étudier, outre les variétés linguistiques, les mutations diachroniques des termes de ce domaine et l'évolution conjointe aux technologies.

Une première extraction lexicale de ces corpus a donc été réalisée depuis le graphe de connaissances à l'aide des logiciels *Sketch Engine* et *Termostat*, puis du langage de programmation Python. Grâce à la bibliothèque NLTK de Python, il a été possible d'extraire les premiers syntagmes nominaux, en appliquant un nettoyage préliminaire du lexique et en éliminant ponctuation et mots vides. L'utilisation des expressions régulières de *Sketch Engine* a permis d'extraire des lexiques spécifiques tels qu'acronymes ou termes commençant par une lettre majuscule suivis d'une série de chiffres. Les outils d'extraction ont tendance à écarter ces schémas, mais ils représentent pour nous une source terminologique majeure dans le domaine. Le plus souvent, il s'agit en effet de noms de machines et de leurs modèles ou fabricants (ex : Gamma 30, IBM 1130).

Sur la base de cette première extraction, quelques considérations s'imposent. Le fait est que le vocabulaire du domaine des TI est, selon les champs, assez normé, d'autant qu'en nous appuyant sur les métadonnées des collections muséales, les vocabulaires issus du graphe de connaissances sont déjà relativement contrôlés et figurent mieux les dénominations de concepts, souvent composées de termes complexes à minimum deux composants, que les outils d'extraction lexicale peinent à révéler. Toutefois, certains champs sont plus foisonnants (ex : les logiciels). Ces fluctuations indiquent des évolutions différentes des connaissances dans le domaine très récent des Technologies de l'Information. Tel quel l'évoque Philippe Seloise (2023), la stabilisation d'une terminologie témoigne de la constitution d'une discipline scientifique, passant du descriptif, au support nominal, à la caractérisation nominale, à l'abstraction, la standardisation par l'élimination des variations nominales, et l'adjectivation avec simplification de la structure syntagmatique. La standardisation qui se fait jour dans certaines branches du patrimoine des TI, comme les vocabulaires relatifs au matériel ou à l'architecture informatique, relate une conceptualisation, une évolution vers la précision terminologique libérée de ses scories, qui accompagne un raffinement des savoirs du domaine. On assiste donc dans certains sous-domaines de

---

le travail terminologique (tels que Vesalius, Schlomann, Wüster, ...) et construit selon les normes ISO 704 et 1087.

<sup>4</sup> Exemple : Mounier-Khun, P. & Lazard, E. (2022). *Histoire illustrée de l'informatique*. Paris : Edp Sciences.

ce patrimoine à une normalisation par des concepts génériques, les fluctuations de sens dues à des subjectivations semblent éliminées, ce qui semble témoigner d'une évolution cognitive. Nous devons aller plus avant dans nos études de la terminologie du domaine des TI pour confirmer ces premières hypothèses. Mais de façon pragmatique, nous allons toutefois être confrontés à l'étendue du domaine et donc du champ lexical. Des choix devront être faits et la délimitation du domaine devra donc jalonner chaque stade du travail.

Sur cette base, une première ébauche de réseau lexical est en cours de construction afin de classer et de mettre en relation les termes du domaine. À ce stade, la difficulté d'une approche purement sémasiologique se pose. Par exemple, pour inclure un abaque et un Gamma 30, nous devons au préalable appréhender les propriétés qui les distinguent d'un point de vue physique et logique, l'évolution des machines dans le temps, leur localisation dans une période historique précise, les inventeurs, etc. Il est donc question de créer au préalable une base de classes et de sous-classes dans un monde extraordinairement vaste. Dès la première étape d'étude de l'état de l'art et de l'existant, nous avons réalisé combien il était novateur d'organiser la connaissance du domaine des TI par une approche patrimoniale et onomasiologique. Comme on l'a évoqué, le choix entre un paradigme plus sémasiologique - à partir de la théorie textuelle de Bourigault et Slodzian (1999), et de la théorie communicative de Cabré (2000), dans ses dérivations socio-terminologiques (Gaudin, 1993 ; Boulanger, 1995) - et une approche résolument onomasiologique (qui maintient les postulats exprimés par Wüster) rend le projet *ITinHeritage* particulièrement actuel dans le panorama des études terminologiques. L'approche la plus efficace dans un contexte de valorisation du patrimoine peut combiner les méthodologies de la termino-ontologie avec les apports des développements récents en terminologie cognitive (Geeraerts (1993) ; Temmerman (2000)), en particulier, le concept d'unités de compréhension UoU, c'est-à-dire des catégories dont les structures prototypiques évoluent sans cesse.

Notre expérience montre que l'élaboration d'un réseau lexical est parfois complexifiante, les relations linguistiques entre les termes d'usage pouvant s'avérer inextricables. Il sera donc nécessaire de s'extraire de la langue et de s'orienter vers l'extralinguistique en organisant les savoirs du domaine selon la logique d'expert au sein d'un réseau conceptuel. Cette étape vise à se concentrer sur les concepts qui forment la connaissance du domaine et pas uniquement sur les façons de les nommer, sources de nombreuses ambiguïtés. Par rapport à une ontologie, le réseau conceptuel permet des représentations plus souples des savoirs par des relations diverses. Cette phase intermédiaire permet de s'affranchir des usages de la langue, d'identifier tous les concepts, de les nommer clairement et de manière consensuelle afin que les noms suffisent à situer le concept dans le réseau notionnel.

Au stade du réseau conceptuel, nous entrons dans le symbolique des références communes du domaine, structurées par le *foedus* évoqué plus haut, et impératives dans une démarche de médiation des savoirs. Ce travail puise les notions dans l'étude approfondie des textes et l'échange régulier avec les experts qui sont au centre de

notre approche, à savoir les experts des musées partenaires et historiens de l'informatique de l'équipe de recherche. À cette heure, nous débutons l'élaboration de ce réseau en sous-domaines représentant les grands concepts. La structuration est pensée de sorte à ce que les hiérarchies participent à la définition de l'objet pour anticiper la construction du dictionnaire. Ce réseau s'étend progressivement et nécessite une rectification continue. Il sera ensuite retravaillé de sorte à entamer la construction du dictionnaire sur l'outil d'Intelligence Artificielle *Tedi* (Roche, 2019) plus à même d'associer l'aspect linguistique des termes à l'extra-linguistique des concepts que Protégé.

L'intention est donc également d'analyser, au vu de notre corpus basé sur des partenaires muséaux internationaux, le multilinguisme et l'approche terminologique du domaine au fil du temps. Il sera intéressant d'analyser les variantes terminologiques, de reconnaître les relations taxonomiques, les synonymes, les co-hyponymes ou antonymes, les relations méronymiques, les termes et leurs cooccurrences. Les langues prises en considération, pour le moment, sont l'anglais, le français et l'italien, et l'ajout d'autres langues, comme l'allemand, sera évalué sur la base des données muséales dont nous disposons.

#### 4.2. Perspectives de construction de l'ontologie

La terminologie établie à ce stade nécessitera une représentation informatique de son système notionnel afin d'être exploitable par l'outil informatique. Pour modéliser et représenter formellement le système notionnel des terminologies, c'est-à-dire la couche extralinguistique, l'ontologie est actuellement l'une des approches les plus intéressantes. Le système notionnel qu'elle représente, associé à un vocabulaire de mots composé de noms de concepts (la terminologie), constitue le " pile et face " d'un même travail. Cette étape commencera par le regroupement plus spécifique des relations du réseau conceptuel établi précédemment pour faire ressortir ce qui est de nature ontologique. Leur clarification nécessitera de repartir des modèles pour restructurer tous les concepts en ensembles sémantiquement liés en ne retenant que des relations de subsomption, soit une réorganisation complète des concepts, voire la création de nouveaux concepts artificiels, mais plus généraux. Les experts valideront *in fine* l'ontologie sur l'organisation des relations et la présence de tous les concepts préalablement identifiés dans les réseaux conceptuels.

Le logiciel *Tedi* permettra de définir les concepts, leurs caractéristiques essentielles (axes d'analyse), leurs caractéristiques descriptives (attributs), et leurs relations. Les ontologies seront construites en utilisant la méthode de différenciation spécifique et traduites en RDF (Resource Description Framework) /OWL (Web Ontology Language) du W3C (World Wide Web Consortium), langages du web sémantique qui permettront une grande portabilité de l'ontologie et des requêtes en SPARQL. Le développement de technologies liées aux ontologies pour les SHS est une approche, encore peu développée, et est en effet, au cœur du web sémantique et des

encyclopédies en ligne. Elle est particulièrement appropriée pour la modélisation et la représentation des connaissances scientifiques à des fins d'appropriation et de partage des connaissances par le public (Ackermann, 1963). Si des initiatives similaires existent (Doerr *et al.*, 2010), elles visent la valorisation d'artefacts plus que celle des savoirs. Ainsi, nous nous concentrerons dans notre étude sur la manière dont les experts conceptualisent et comment terminologie et connaissance sont liées dans le domaine des TI (Roche et Papadopoulou, 2019). Nos contributions à l'état de l'art concerneront également la modélisation diachronique des connaissances et la gestion du temps dans le cadre de standards tels que CIDOC-CRM ou EDM (Europeana Data Model), un problème ouvert depuis Pan et Hobbs (2005) (NYS *et al.*, 2018). Afin de représenter diachroniquement les connaissances et d'intégrer notre ontologie du domaine des TI à la première ontologie du domaine patrimonial qui structure le graphe de connaissances sur la base des modèles CIDOC-CRM et EDM, nous utiliserons OntoMe (Ontology Management Environment) du consortium *Data4History* (Beretta, 2021). OntoMe permettra une adaptation aisée à notre ontologie initiale puisqu'utilisant les mêmes formalismes et permettra la représentation par périodes temporelles, particulièrement adaptée à notre domaine. Enfin, nous apporterons des solutions aux problèmes de prise en compte des principes terminologiques des normes internationales ISO (ISO 1087 et ISO 704) pour la construction d'ontologies au format OWL (Web Ontology Language). Si à ce jour, des taxonomies du domaine informatique existent : taxonomie de l'ACM (Association for Computing Machinery) (Coulter, 1997), la Computer Science Ontology (CSO) générée automatiquement par alignement avec plusieurs autres graphes de connaissances (DBpedia, Wikidata, YAGO, Freebase, Cyc, ...). (Salino *et al.*, 2020), classification du GETTY Museum, elles ne prennent en compte que l'aspect linguistique du domaine.

## Conclusion

*«... museum collections often have a rich source of metadata that could potentially be used to provide more intelligent and scaffolded means of search and exploration» (Mulholland *et al.*, 2008).*

Qu'est-ce que le patrimoine des Technologies de l'Information (TI) ? Comment le circonscrire, le définir, et en tirer des visions prospectives ? Quels sont ses ancrages et impacts technologiques, économiques, politiques et sociétaux, sans précédent et hors proportions ? Se pencher sur ce patrimoine est un enjeu crucial de compréhension du monde actuel. Aux côtés de ses musées partenaires, le projet de recherche *ITinHeritage* vise à mener une réflexion épistémologique sur la question des TI en mettant au cœur de la question du patrimoine l'histoire d'une mutation sociétale et l'émergence de nouveaux savoirs en termes de représentations du monde, à travers la définition de ce patrimoine et des nouveaux lieux de savoirs et pratiques de métiers et privés qu'il engendre. Qu'il s'agisse de l'expression de ses savoirs explicites ou expérimentiels, notamment à travers le langage, dont nous concrétisons l'étude par la

création d'un dictionnaire terminologique multilingue et multimédia à l'image de ce patrimoine, de la conceptualisation du monde qu'il relate, que nous modélisons grâce à l'ontologie et au knowledge graph, et que nous œuvrons à rendre accessibles par une plateforme de web sémantique et des agents conversationnels soutenus par l'IA, le patrimoine des TI nous ouvre un champ infini d'étude. Mais les besoins de ses musées et publics ne se cristallisent pas dans les objets. Bien au-delà, c'est l'histoire que tout un chacun a vécu avec eux et qui témoigne de tout un temps, qui les passionne. Ainsi, les TI se traduisent par une multitude de domaines en jeu (technologiques, économiques, politiques, sociaux), d'usages, de pratiques, allant de l'industriel, professionnel, à l'intime, qui impliquent pour le projet *ITinHeritage* de discuter les questions et les enjeux du patrimoine des TI sur la période contemporaine et sa mutation numérique et sociétale, depuis la deuxième moitié du XXe à nos jours. L'évolution des conceptions et pratiques à travers l'histoire des TI est la perspective qui s'ouvre à notre projet pour les prochaines années à venir.

## Références

- Ackermann, W. 1963. "Etudes sur la diffusion et la représentation sociale des connaissances scientifiques ", *Sociologie du travail*, n° 4, pp. 415-417.
- Allard-Chanai, L. 1998. « Médiation esthétique et réseaux de communication: L'exemple de la cinéphilie assistée par ordinateur ». In *Médiations sociales systèmes d'information et réseaux de communication*. 3-5 déc., Metz, France.
- Aristote. 1986. *Métaphysique*, trad. J. Tricot. Paris : Vrin.
- Beretta, F. 2021. «A challenge for historical research: making data FAIR using a collaborative ontology management environment (OntoME)». *Semantic Web*, 12(2), 279-294.
- Caune, J. 1995. *Culture et communication: Convergences théoriques et lieux de médiation*. Grenoble : PUG.
- Chaudiron, S., Jacquemin, B., Kergosien, E. 2018. « L'apport du web sémantique à la sauvegarde du patrimoine immatériel. Les cas du textile, de la musique et de la mine ». *HumaNum. Information, communication et humanités numériques. Enjeux et défis pour un enrichissement épistémologique. Actes du 23e colloque bilatéral franco-roumain en Sciences de l'information et de la communication*, (pp.311-328). Université Babeş-Bolyai, Cluj-Napoca, Roumanie.
- Cobb, J. 2015. "The journey to linked open data: The Getty vocabularies". *Journal of Library Metadata*, 15, 3-4 : 142-156.
- Crespi, F. 1983. *Médiation symbolique et société*. Paris : Librairie des méridiens.
- Davallon, J. 2003. « La médiation: la communication en procès ». *Médiation et information*, 19 : 37-59.

## La médiation des objets aux savoirs scientifiques et techniques

- de Boer, V., *et al.* 2013. « Amsterdam museum linked open data ». *Semantic Web*, 4,3 : 237-243.
- Després-Lonnet, M., Rizza, M. 2021. « Quand l'archive incarne l'institution: le rôle des dossiers d'œuvre dans la fabrique documentaire de la muséalité d'un musée des Beaux-arts ». *Signata. Annales des sémiotiques/Annals of Semiotics*, 12.
- Djambian, C., Lainé-Cruzel, S. 2012. « Appropriation des savoirs et variété des langues d'usages des communautés de métier ». In *Terminologies : textes, discours et accès aux savoirs spécialisés : actes du V<sup>e</sup> colloque international du Groupe de linguistique appliquée des télécommunications (GLAT)*, (pp. 343-356). Gênes, Italie.
- Doerr, M., *et al.* (2010). "The Europeana Data Model (EDM) ", In *Actes du Congrès mondial des bibliothèques et de l'information : IFLA 76*, Gottenburgh, Suède.
- Dufrêne, B., Ijadjene, M. Bruckmann, D. 2013. *Numérisation du patrimoine. Quelles médiations ? Quels accès ? Quelles cultures ?* Paris : Hermann Editeurs.
- Faber, P., L'Homme, M-C. (eds). 2022. *Theoretical Perspectives on Terminology: Explaining Terms, Concepts and Specialized Knowledge*. Amsterdam: John Benjamins Publishing Company.
- Freire, N., Isaac, A. 2019. « Wikidata's Linked Data for Cultural Heritage Digital Resources: An Evaluation Based on the Europeana Data Model", In *DCMI International Conference on Dublin Core and Metadata Applications*, Seoul, Corea.
- Gaudin, F., Nicolae, C., Dysola, E. A. 2012. « Le référent, une construction sociale. Exemples d'astronomie ». In *Actes de la conférence TOTh 2012*, juin, Le Bourget du Lac, France.
- Hennion, A., Maisonneuve, S., Gomart E. 2000. *Figures de l'amateur: formes, objets, pratiques de l'amour de la musique aujourd'hui*. Paris: La documentation française.
- ISO (Organisation internationale de normalisation). 2022. 704:2022. Travaux de terminologie - Principes et méthodes. Genève : ISO.
- ISO (Organisation internationale de normalisation). 2019. 1087-1:2019. Travaux de terminologie - vocabulaire - partie 1 : théorie et application. Genève : ISO.
- Jacob, C. 2001. « Rassembler la mémoire ». *Diogène*, 4 : 053-076.
- Januals, M., Minel, J.-L., 2016. « La construction d'un espace patrimonial partagé dans le Web de données ouvert », *Communication*, 34, 1.
- Jouët, J. 1993. « Usages et pratiques de nouveaux outils de la communication », *Dictionnaire critique de la communication*. Paris : Presses Universitaires de France.
- Kant, E. 1997. *Critique de la raison pure*. trad. A. Renaut. Paris : Aubier.
- Lamizet, B. 1992. *Les lieux de la communication*. Bruxelles : Pierre Mardaga.
- Lamizet, B. 1999. *La médiation Culturelle*. Paris : L'Harmattan.
- Piccini, S., Vezzani, F., Bellandi, A. 2022. « Entre TBX et Ontolex-Lemon: quelles nouvelles perspectives en terminologie ». In *Actes de l'International Conference on Multilingual*

- digital terminology today. Design, representation formats and management systems*, CEUR-WS : 19-30.
- Poli, M.-S. 2012. « Que signifie réellement engager des médiations innovantes ? », *Musées en mutations*, Paris.
- Popper, K. 1979. *Objective knowledge: an evolutionary approach*. Oxford: Oxford University Press.
- Price (de la Solla), D. 1963. *Little science, big science*. New York et Londres : Colombia University Press.
- Puren, M., Vernus, P.. 2022. « Vers une ontologie de domaine pour l'analyse des tissus anciens. Le projet Silknow et le cas du patrimoine soyeux européen », *Humanités numériques*. <http://journals.openedition.org/revuehn/3179>
- Putnam, H. 1984. *Raison, vérité et histoire*. Paris : Editions de Minuit.
- Py, É. D. L., Debruyne, F., & Vandiedonck, D. 2002. "La recherche du sens". In *Les recherches en information et communication et leurs perspectives*, SFIC, 7-9 oct., Paris, France.
- Quéré, L. 1982. *Des mémoires équivoques. Aux origines de la communication moderne*. Paris: Aubier.
- Rastier, F. 2006. "Formes sémantiques et textualité". *Langages*, 3: 99-114.
- Roche, C., Papadopoulou, M. 2020. "Rencontre entre une philologue et un terminologue au pays des ontologies". *Revue Ouverte d'Intelligence Artificielle*, 1 : 43-70.
- Roche, C., Papadopoulou, M. 2020. « Terminologie et ontologie pour les humanités numériques : le cas des vêtements de la Grèce antique », *Humanités numériques*. Accès 20 <http://journals.openedition.org/revuehn/462>
- Roche, C., Papadopoulou, M. 2019. " Mind the Gap: Ontology Authoring for Humanists ", In *JOWO : The Joint Ontology Workshops*, Graz, Autriche.
- Roche C. 2005. « Terminologie et ontologie ». *Langages*, 157 : 48-62.
- Rossi, M. 2021. "Termes et métaphores, entre diffusion et orientation des savoirs ". *La linguistique*, 57 : 153-173.
- Schiele, B. 1989. *Faire voir, faire savoir, la muséologie scientifique au présent*. Paris : Ed. Musée de la Civilisation.
- Selosse, P. 2023. « Naissance et renaissance de la terminologie botanique au XVIe siècle », In *Terminologie et ontologie : théories et applications*, TOTh 2023, 1-2 juin, Le bourget du Lac, France.
- Szekely, P., *et al.* 2013. « Connecting the Smithsonian American Art Museum to the Linked Data Cloud », In *The Semantic Web: Semantics and Big Data, ESWC 2013*, Berlin, Germany.

- UNESCO. 2023. « Technologies de l'Information et de la Communication (TIC) », *Glossaire*. Accès 23 <https://uis.unesco.org/fr/glossary-term/technologies-de-linformation-et-de-la-communication-tic>
- Vezzani, F., Di Nunzio, G. M. 2023. « Multilingual digital terminology: Introduction to the special issue », *Digital Scholarship in the Humanities*, 38.
- Vezzani, F. 2022. *Terminologie numérique : conception, représentation et gestion*. Berne : Peter Lang.
- Wagenberg, J. 2006. "Toward a total museology through conversation between audience, museologist, architects, and builders", In R. Terradas *et al.*, *The Total Museum*, Barcelona: Sacyr.
- Wei, T., *et al.* 2022. "Using ISO and Semantic Web standard for building a multilingual terminology e-Dictionary : A use case of Chinese ceramic vases", *Applied Ontology*, 17 (3): 423-441.

## Abstract

Safeguarding our Information Technology (IT) heritage is in the public interest. The rapid evolution of IT has left little room for its study and enhancement. The *ITinHeritage* research project aims to understand how contemporary scientific and technical knowledge can be perpetuated and mediated, based on this particular heritage. Our approach starts with IT museums. With our partners: Musée des Arts et Métiers, ACONIT (France), London Science Museum (England), NAM-IP (Belgium), Museo del calcolo (Italy), Homecomputermuseum (Holland), HNF (Germany), we are developing an innovative approach of Digital Humanities by creating a knowledge graph supported by terminology and ontology of the field integrating major European projects (Europeana, CIDOC-CRM, OntoMe), accessible via a platform on the semantic web (Linked Open Data (LOD)), based on upstream and downstream analysis of scientific and technical knowledge transfer mechanisms.

"ANR-19-P3IA-0003 ": "This work has been partially supported by MIAI @ Grenoble Alpes, (ANR-19-P3IA-0003)"