Du projet symogih.org au consortium Data for History - La modélisation collaborative de l'information au service de la production de données géo-historiques et de l'interopérabilité dans le web sémantique

From the symogih.org project to the Data for History consortium - Collaborative information modeling for geo-historical data production and interoperability in the semantic web

Francesco Beretta[1], Vincent Alamercery[2]

[1] Laboratoire de recherche historique Rhône-Alpes, CNRS, francesco.beretta@ish-lyon.cnrs.fr
[2] Laboratoire de recherche historique Rhône-Alpes, ENS de Lyon, vincent.alamercery@ens-lyon.fr

RÉSUMÉ. Le projet *symogih.org*, système modulaire de gestion de l'information historique, a mis en place dès 2008 un environnement virtuel de recherche permettant la production collaborative et cumulative de données géo-historiques issues de multiples et différents projets de recherche. Une attention particulière a été portée dans ce processus aux méthodes de mutualisation et d'interopérabilité des données, et en particulier à la question de la création d'un modèle générique et ouvert, fondement du système d'information et capable de s'adapter aux problématiques des différents projets. Cette démarche a conduit les porteurs du projet à inscrire leurs travaux dans le cadre du CIDOC CRM et à constituer le consortium *Data for History* qui vise à promouvoir l'activité d'une communauté internationale de chercheurs intéressés par un modèle partagé de données dans lequel pourront se décliner les spécificités de chaque projet. Ce processus facilitera la mise en réseau, l'interopérabilité et la pérennisation des données géo-historiques produites dans les systèmes d'information des différents projets, en accord avec les principes FAIR.

ABSTRACT. The symogih.org project ("Système modulaire de gestion de l'information historique") set up in 2008 a virtual research environment allowing the collaborative and cumulative production of geo-historical data from multiple and different research projects. Particular attention was paid in this process to the methods of pooling and interoperability of data, and in particular to the question of the creation of a generic and open model, the basis of the information system and capable of adapting to issues of the different projects. This approach led the project leaders to connect their work to the conceptual framework of CIDOC CRM and to launch an initiative for a Data for History consortium which aims to promote the activity of an international community of researchers interested in a shared data model in which decline the specifics of each project. This process will facilitate the networking, interoperability and sustainability of geo-historical data produced in the information systems of different projects, in accordance with the FAIR principles.

MOTS-CLÉS. modélisation des données géo-historiques, ontologies, interopérabilité, plateforme collaborative, système d'information distribué, principes FAIR.

KEYWORDS. modeling of geo-historical data; ontologies; interoperability; collaborative platform; distributed information system; FAIR principles.

1. Introduction

Dans le champ des systèmes d'information, on reconnaît depuis longtemps l'importance d'une modélisation robuste des données en vue d'assurer leur interopérabilité et leur réutilisation pour de nouveaux questionnements. C'est ce que soulignait Matthew West pour le monde de l'industrie, au milieu des années 1990 et donc bien avant l'introduction du web sémantique et de ses technologies, dans le document de travail *Developing high quality data models* rédigé pour le *European Process*

Industries STEP Technical Liaison Executive et publié plus tard sous forme de livre (West 2011) : quels principes adopter pour faciliter l'interopérabilité entre systèmes d'information ?

Ces mêmes interrogations sont partagées dans le domaine des humanités numériques (Courtin et Minel 2017) et notamment dans celui de l'histoire numérique (Alérini et Lamassé 2011). Elles sont présentes depuis les débuts du projet *symogih.org (Système modulaire de gestion de l'information historique)*, né en 2008 de la volonté de mutualiser et de réutiliser les données produites par les chercheurs en histoire. Elles sont devenues plus urgentes au cours des dernières années avec l'éclosion du web sémantique (Meroño-Peñuela et al. 2015). La panoplie de modèles et de technologies qui se développent dans ce domaine permettent désormais de partager très facilement les données entre différents entrepôts et de disposer ainsi d'un volume inespéré de ressources qu'un chercheur peut soumettre à analyse en fonction de sa problématique de recherche. Mais comment rendre intelligibles et, surtout, interopérables les données produites par des systèmes d'informations géo-historiques qui ont été conçus de manière indépendante les uns des autres, à partir de différents questionnements ?

Cette question est soulevée en particulier par l'article 'I', *Interoperable*, formulé parmi les principes FAIR: « make data Findable, Accessible, Interoperable, and Re-usable »[1]. Issus de la mouvance *open data* et *open science* dans le domaine des sciences naturelles, ces principes visent la réutilisation des données produites par la recherche dans le contexte de nouveaux questionnements : « There is an urgent need to improve the infrastructure supporting the reuse of scholarly data » (Wilkinson 2016, Abstract). Les chercheurs sont invités à publier non seulement leurs résultats sous forme d'articles ou de livres, mais aussi les données elles-mêmes qui ont servi à les établir[2].

L'article *Interoperable* préconise ainsi, lors de la production des données, « [to] use a formal, accessible, shared, and broadly applicable langage for knowledge representation »[3]. Ce principe peut être explicité en rappelant l'une des définitions habituelles de l'ontologie, au sens informatique, définie en tant que conceptualisation formalisée et partagée d'un domaine disciplinaire : « An ontology is a formal explicit specification of a shared conceptualization of a domain of interest » (Domingue 2011, 510). Mais on peut se demander dans quelle mesure cette vision est applicable aux données produites par la recherche historique. Ces données ne sont-elles pas liées nécessairement à une problématique précise et donc non réutilisables pour d'autres questionnements ? Et aussi, les données extraites des sources historiques ne sont-elles pas, par définition, incertaines, tant dans l'identification des objets dont elles parlent, notamment les personnes, que dans les dates des événements et dans leur contenu-même ? Comment alors exprimer ces degrés d'incertitude sous forme de données ? Et ce, de surcroît, dans un langage quasi-universel, permettant l'interopérabilité au niveau de l'ensemble des données produites par les historiens ?

Sans avoir la prétention de répondre ici à ces questions complexes de manière exhaustive, nous présenterons l'expérience de mutualisation de données telle qu'elle a été développée au sein du projet *symogih.org* avec une attention particulière à la question de la modélisation générique et ouverte comme condition de l'interopérabilité. Nous présenterons également le modèle *factoïd* développé au sein des projets de prosopographie du Kings College de Londres, tout en l'articulant avec les choix de modélisation opérés dans le contexte du projet *symogih.org* afin d'en mettre en évidence les spécificités. Puis nous présenterons les raisons qui ont conduit à la création du consortium *Data for History* (http://*dataforhistory.org*) et son potentiel pour la mise en place de

[1] Cf. les instructions dans le cadre du Programme H2020 : *Guidelines on FAIR Data Management in Horizon 2020*, Version 3.0, 26 July 2016, de même que le site https://www.force11.org/group/fairgroup/fairprinciples .

[2] Voir par exemple la revue *Scientific data* publiée par le groupe Nature.

[3] https://www.force11.org/group/fairgroup/fairprinciples .

l'interopérabilité des données dans le domaine géo-historique grâce à l'utilisation du modèle conceptuel CIDOC CRM en tant que contexte sémantique permettant l'interopérabilité. En conclusion, nous exposerons quelques considérations concernant les enjeux liés à la pérennisation des données produites par les projets de recherche et aux modèles économiques permettant de créer l'infrastructure nécessaire à cette fin.

2. Le projet *symogih.org*

Le projet « Système modulaire de gestion de l'information historique » (http://symogih.org/) est né en 2008 de la volonté de quelques historiens du Laboratoire de recherche historique Rhône-Alpes à Lyon (UMR5190 LARHRA) de mutualiser les données structurées produites au cours de leur recherche afin de permettre leur réutilisation par d'autres chercheurs, tout en se référant aux standards en vigueur dans le domaine de la modélisation des bases de données (Akoka et Comyn-Wattiau 2001, Soutou 2007, Audibert 2009). Cette démarche s'inscrit dans la logique de la « curation des données » entendue au sens d'un enrichissement et d'une amélioration constantes des données afin de garantir leur qualité, leur accessibilité et leur préservation[4].

À titre d'exemple, les données produites au cours du projet SIPPAF (http://www.patronsdefrance.fr/), projet financé pendant trois ans par l'Agence nationale de la recherche et qui a abouti à la mise en place d'un système d'information prosopographique consacré au patronat français (XIX^{e}-XX^{e} siècles), continuent à être enrichies et utilisées par des chercheurs et des étudiants, notamment dans le cadre du projet SIPROJURIS (http://siprojuris.symogih.org/) consacré aux professeurs de droit en France de 1804 à 1950. Ces deux projets disposent chacun d'un site web dédié mais la collecte des données repose sur un système d'information collaboratif unique qui favorise leur échange et leur réutilisation. Les données peuvent ainsi être enrichies par de nouvelles recherches menées par d'autres utilisateurs de l'environnement de recherche collaboratif mis en place : elles continuent ainsi à vivre après la fin des projets et des financements précédents, afin d'être réutilisées pour de nouveaux questionnements. En retour, le volume des données publiées sur les sites web des projets précédents continue à augmenter et leur qualité à être améliorée, ce qui prolonge leur durée de vie et leur utilisation par le public.

Un nombre croissant de projets internes et externes au LARHRA, français et européens (actuellement plus d'une soixantaine d'utilisateurs et une quinzaine de projets)[5] utilisent cet environnement virtuel de recherche afin de produire et de mutualiser leurs données. Celles-ci sont mises à la disposition du public sous licence *Creative Commons Attribution-ShareAlike 4.0 International* dans un site web générique, le site *symogih.org*, qui représente le point d'accès central des données, ainsi que dans les sites des différents projets. La plateforme *symogih.org* comprend également un système de gestion de données spatiales, GEO-LARHRA (http://geo-larhra.ish-lyon.cnrs.fr/) (Beretta F. et Butez C. 2013) et un environnement d'encodage et d'annotation sémantique des textes au format XML/TEI (http://xml-portal.symogih.org/) (Beretta et Letricot 2017), accessibles depuis le site web générique. Bien que la dimension géographique de la recherche en histoire et la mise en relation explicite des textes et des données qui en sont extraites, grâce à l'encodage sémantique (Beretta 2016), représentent une partie importante de l'environnement virtuel de recherche mis en place, ces deux volets ne seront pas présentés ici car les développements importants qu'ils demanderaient dépassent les limites de cette contribution.

L'environnement virtuel de recherche mis en place par le projet *symogih.org*, fondé sur un système d'information générique multi-utilisateurs, permet de faciliter et de rationaliser le suivi de

[4] Cf. https://en.wikipedia.org/wiki/Data_curation .

[5] Cf. la liste sur le site du projet http://symogih.org .

projets et d'établir un plan de gestion des données. Un projet financé peut ainsi s'adresser au Pôle histoire numérique du LARHRA et demander l'hébergement de la partie numérique en décrivant sa problématique de recherche et ses objectifs : en application des principes du cycle de vie des données (*data life-cycle management*), un modèle de données susceptible d'être intégré à la plateforme et, en même temps, adapté aux besoins du projet est alors mis au point en collaboration avec les participants, ainsi qu'une stratégie de dépouillement des sources et de production des données. Il est aussi possible de solliciter les compétences d'autres spécialistes ou de favoriser des collaborations avec des laboratoires d'informatique pour des tâches avancées. Les chercheurs sont accompagnés, si nécessaire, au cours de l'analyse et de l'exploitation des données ; enfin, un site web dédié spécifiquement à leur projet, exposant les résultats de leur recherche, peut être mis en place. Les données elles-mêmes peuvent être publiées sous forme d'*open data* et pérennisées grâce au processus de réutilisation et d'enrichissement esquissé ci-dessus.

Le fait de gérer les données de différents projets au sein d'un seul système d'information, conçu de manière modulaire, permet d'assurer leur pérennisation après la fin du financement et leur réutilisation pour d'autres projets. La pérennité physique des données est assurée par leur sauvegarde sur les infrastructures mises à disposition par la TGIR Huma-Num. Cette architecture permet également d'intégrer de nouveaux modules à partir de technologies existantes et standardisées, ou de services mises à disposition par d'autres organismes. Ainsi, l'analyse et la publication des données de l'environnement de recherche du projet *symogih.org* peut aisément être déportée sur l'infrastructure RStudio déployée par la TGIR Huma-Num[6].

Pour les projets ouverts à la logique de l'*open data*, un point d'accès SPARQL$_7$ permet d'interroger directement la portion des données que les chercheurs ont décidé de publier au format RDF. La structure du modèle de données adopté, documentée publiquement sur le site *symogih.org*, sera décrite ci-dessous. Dans une logique de données ouvertes liées (*linked open data* ou LOD), il est essentiel, outre de publier le modèle, de relier les instances présentes dans l'ontologie avec d'autres référentiels. Dans ce domaine, une expérience pilote d'alignement avec le référentiel IdRef[8] est en cours autour des données du projet SIPROJURIS, sous la direction de François Mistral de l'Agence bibliographique de l'enseignement supérieur (ABES). L'alignement effectué entre les IdRef et les notices d'autorité propres au projet *symogih.org* permet d'afficher, dans les notices dédiées aux professeurs de droit du projet SIPROJURIS, la liste de leurs publications en les récupérant en temps réel des notices du catalogue des bibliothèques du SUDOC[9].

3. Une modélisation générique et ouverte au cœur du système d'information

Depuis le début du projet *symogih.org* en 2008, une attention particulière a été consacrée à réfléchir à un modèle de données ouvert, susceptible de s'adapter à tout type d'information historique quelles que soient les thématiques de recherche ou la période étudiée et, en même temps, en dialogue avec les standards existants. On visait ainsi à garantir l'interopérabilité des données produites avec celles issues d'autres projets ayant la même démarche, ainsi qu'avec celles des institutions patrimoniales comme la Bibliothèque nationale de France ou de certains musées. Cette approche était indispensable afin de mettre en place un système d'information générique permettant

[6] Cf. quelques exemples sur ce site https://frama.link/phn-shiny, sachant que ces analyses et visualisations de données ont une fonction heuristique et ont été mis en place pour être principalement utilisés par les chercheurs des projets concernés.

[7] Cf. http://symogih.org/?q=rdf-publication .

[8] https://www.idref.fr/ .

[9] http://siprojuris.symogih.org/siprojuris/enseignant/44315 (onglet : Bibliographie externe).

aux porteurs de projets scientifiques très divers de travailler dans un environnement virtuel de recherche collaboratif.

Deux principes fondamentaux ont guidé l'opération de modélisation au sein du projet *symogih.org*. D'une part, une séparation claire a été introduite entre la production des données et la problématique de recherche qui accompagne leur collecte. Certes, toute production de données trouve son origine dans un questionnement. Toutefois, l'information stockée dans l'environnement de recherche doit être modélisée de la manière la plus objective possible afin de permettre sa réutilisation pour d'autres recherches. La dimension historienne de la recherche est appliquée au moment d'interroger les données récoltées dans le système d'information ; elles sont agrégées en fonction de la problématique et soumises à des requêtes qui découlent du questionnement des historiens. Il s'agit par exemple de reconstituer le déroulement d'un procès, de spatialiser un ensemble d'événements ou de comparer les carrières d'un ensemble d'acteurs (Beretta 2014).

D'autre part, il est impératif de procéder à l'atomisation des données, c'est-à-dire d'entreprendre un processus de décomposition de l'information en éléments correspondants à des propositions simples et autonomes, idéalement elles-mêmes non décomposables ultérieurement (Dedieu 2004 ; Beretta et Vernus 2012). Ce processus d'atomisation doit être documenté explicitement en identifiant le sens de chaque proposition ainsi que le rôle de chaque objet impliqué. On distinguera ainsi entre un événement tel qu'un congrès comme tel, et les multiples événements que sont la présence ou l'absence de différentes personnes à différents moments de cet événement, la durée et les interruptions de chaque présence pouvant être différentes et significatives pour une reconstitution historique. Aussi, dans le cas d'un meurtre il faudra expliciter le rôle de chacune des personnes impliquées, victime, assassin, complice, et éventuellement décomposer l'événement dans toutes ses phases si elles présentent des caractéristiques différentes que l'historien souhaite étudier.

Pour atteindre ce but, le projet a retenu un modèle générique de bases de données qui transforme en données le modèle lui même : le système d'information permet ainsi aux chercheurs de créer de nouveaux aspects du modèle correspondant aux besoins de leur problématique. Ces instances du modèle sont discutées avec la communauté des utilisateurs puis validées et deviennent ainsi utilisables par tous. Les définitions des instances du modèle sont publiées sur le site principal du projet *symogih.org* afin d'expliciter le sens des données publiées et d'en permettre la réutilisation, notamment dans le cadre de requêtes sur le point d'accès SPARQL[10].

Cette démarche de construction d'un modèle générique et ouvert, dont les composantes sont définies progressivement par les chercheurs, s'inscrit dans un contexte général qui a vu les modèles relationnels des systèmes d'information évoluer vers les technologies du web sémantique : les auteurs d'un ouvrage d'introduction dédié à ce domaine soulignent que les métadonnées décrivant le modèle dans un système de bases de données générique, enrichi progressivement, explicitent les relations sémantiques entre les instances produites dans le système et deviennent elles mêmes des données, permettant une grande flexibilité dans la gestion du modèle qui sera ensuite mis en valeur dans le formalisme plus souple du RDF (Segaran 2009, 16). Dans ce contexte, un processus de réécriture du modèle générique du projet *symogih.org* a été entamé en 2013 afin de le reproduire sous forme d'ontologie et de réfléchir à son alignement avec les référentiels utilisés dans le monde de la conservation des biens patrimoniaux, tels ceux du CIDOC CRM et du FRBRoo[11].

Au centre de la représentation simplifiée de l'ontologie dans la figure 1. se trouvent les deux classes principales: la classe *Object* et la classe *KnowledgeUnit*. La première regroupe tous les « objets » possédant une identité propre qui est stable dans la durée en dépit des transformations de

[10] Cf. http://symogih.org/?q=type-of-knowledge-unit-classes-tree .

[11] Cf. http://www.cidoc-crm.org/ et http://www.cidoc-crm.org/collaborations .

leurs caractéristiques ou apparences. Il s'agit d'objets concrets (tels une personne, une maison, un manuscrit) ou abstraits (tels un concept, une référence bibliographique, une profession).

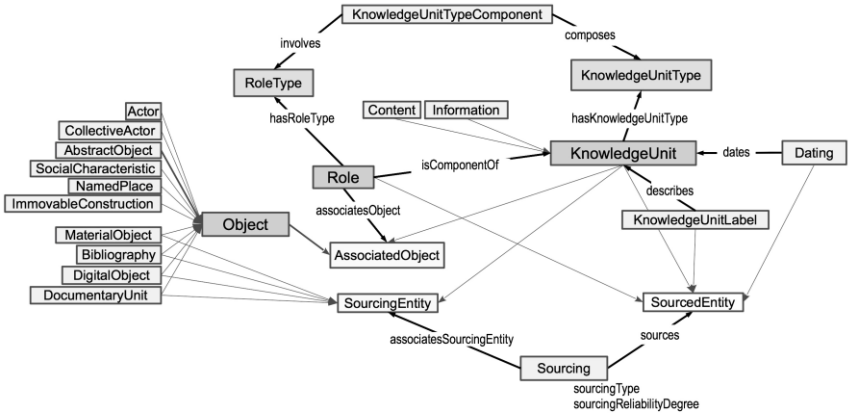

Figure 1. *L'ontologie du projet symogih.org - version 0.2.1*

À titre d'exemple, prenons en considération la proposition « En 1592, Galileo Galilei est engagé par l'Université de Padoue où il enseignera les mathématiques jusqu'en 1610 » qui exprime une information élémentaire. On reconnaît dans cette proposition les objets « Galilée », la discipline des « mathématiques », « l'Université de Padoue » ainsi que, implicitement, le lieu « ville de Padoue ». Chaque objet sera décrit par un identifiant stable, publié sous forme de *Uniform Resource Identifier* (URI) et déréférençable sur le site du projet *symogih.org*, et par une notice qui exprime succinctement ses caractéristiques essentielles afin que d'autres chercheurs comprennent aisément de quel objet il s'agit. Dans la fig. 1, les objets sont regroupés en dix sous-classes de la classe *Object* (par ex. *Actor*, *Collective Actor*, *Abstract Object*, etc.) qui sont construites de la manière la plus objective possible afin d'être adaptées à tous les contextes de recherche.

La seconde classe regroupe les « unités de connaissance » (classe *KnowledgeUnit*) définies dans l'ontologie du projet *symogih.org* en tant qu'assertions de l'historien qui décrivent une information, c'est-à-dire une relation entre objets située dans le temps et dans l'espace, considérée comme ayant virtuellement existé. Comme il a été indiqué ci-dessus, ces assertions sont atomisées et conçues de la manière la plus objective possible afin de permettre leur réutilisation dans d'autres contextes. On peut ainsi extraire de la proposition mentionnée à titre d'exemple une information atomisée ou unité de connaissance qui met en relation pendant une période donnée —avec un sens bien précis : l'enseignement— une personne (Galilée), une institution (l'Université de Padoue) et une discipline (les mathématiques). Une instance de la classe *KnowledgeUnitType* est définie pour chaque type d'information qu'on souhaite stocker et publiée sur le site *symogih.org*[12]. Elle explicite le sens des données produites et permet de comprendre l'articulation des objets qui entrent en relation au sein de la connaissance, la participation de chacun d'entre eux étant définie par un rôle précis (classes *Role* et *RoleType*).

Il est à relever qu'à partir de la même proposition on aurait pu extraire d'autres informations, tel le fait que Galilée réside désormais dans la ville de Padoue, ou qu'il a été engagé par l'Université de cette ville, ou qu'il a le statut de professeur indépendamment du fait qu'il enseigne effectivement ou pas. C'est donc le questionnement appliqué à un texte qui permet d'en extraire différentes informations ou, en d'autres termes, de construire la donnée en fonction d'un modèle : ce qui compte c'est d'expliciter et de documenter ce processus, ce qui s'effectue grâce à la documentation

[12]Voir la définition du type d'information 'Enseignement', URI: http://symogih.org/resource/TyIn97 .

des instances de la classe *KnowledgeUnitType*. Le modèle n'est donc pas figé d'avance, il s'adapte aux différentes problématiques de recherche tout en visant —grâce à l'atomisation et la séparation entre problématique et production des données— le plus d'objectivité possible.

4. Comparaison avec le modèle du *factoïd* et le CIDOC CRM

Relevons aussi qu'une distinction essentielle subsiste entre les assertions qui modélisent les « faits » comme tels (*states of affairs*), par exemple le fait que Galilée a enseigné à Padoue, et celles qui reproduisent pour ainsi dire littéralement le contenu d'un document, chaque source apportant différents points de vue tant sur les dates, les contenus ou les circonstances de tel événement que sur son interprétation. C'est ce que soulignent John Bradley et Michele Pasin dans un article qui publie le modèle de données *factoïd* développé dans le contexte des projets de prosopographie du Moyen Âge portés par le Département d'humanités numériques du King's College de Londres. D'un côté, affirment-ils, il y a les « faits historiques » (« *states of affairs* »), de l'autre les assertions des sources concernant ces mêmes faits : « The factoïd approach prioritizes the sources, rather than our historians' reading of them » (Bradley 2015, 89).

En d'autres termes, les *factoïds* modélisent le contenu des sources, alors que les « informations » définies par le projet *symogih.org* modélisent la « réalité » du passé. Afin de prendre en compte cette distinction essentielle pour la recherche historique, les « contenus » ont été introduits dans le système d'information du projet *symogih.org* dès 2010. Ils sont construits de manière analogue aux « informations », c'est-à-dire en utilisant un modèle générique instancié sous forme de types de contenus, mais ils ont un niveau épistémologique substantiellement différent : les « contenus » (tout comme les *factoïds*) modélisent les assertions de la source, avec tout le lot d'incertitudes, de redondance et d'ambiguïtés qu'elle peut comporter, alors que les « informations » modélisent les assertions de l'historien après qu'il a appliqué la méthode critique. En tant qu'expression du contenu de la source, les « contenus »/*factoïds* peuvent être également annotés directement dans la transcription d'un document, par exemple en le reproduisant sous forme d'un texte structuré au format XML, selon le standard de la *Text encoding initiative*[13], et en procédant ensuite à son annotation sémantique en lien avec un référentiel partagé (Eide 2014-2015 ; Beretta 2016).

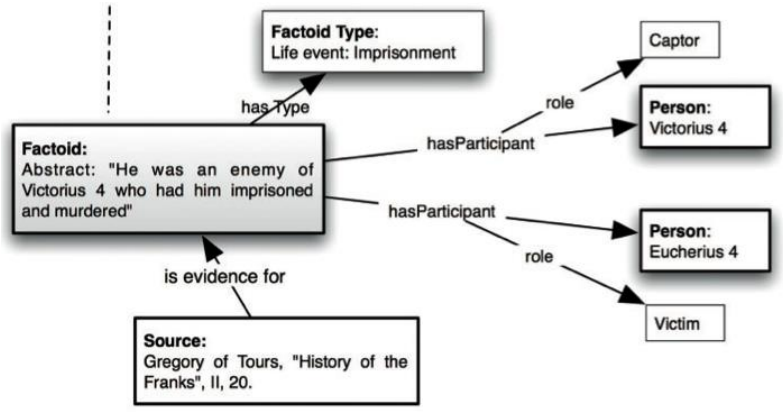

Figure 2. *Exemple d'une instance du modèle* factoïd *(Bradley 2015, 88)*

Pour revenir aux *factoïds*, la figure 2 montre que leur structure ontologique est analogue à celle de l'information/unité de connaissance du projet *symogih.org* : une assertion est qualifiée par un type, on indique quelle est sa source et quels sont les objets qu'elle associe par le biais de rôles dont la sémantique est explicitée par un type défini par les chercheurs.

[13] http://www.tei-c.org/ .

C'est donc non la structure mais le niveau épistémologique de la donnée qui fait la différence entre les deux modèles : afin de passer d'un niveau à l'autre il faut appliquer l'outillage de la critique historique, c'est-à-dire vérifier la fiabilité et le degré de véracité de chaque assertion de la source, puis agréger les contenus des différentes sources en une seule information censée reproduire avec plus ou moins de certitude les « faits » comme tels. Ce processus d'agrégation et ce changement de niveau épistémologique est indispensable pour répondre à bon nombre de questionnements de la recherche historique. En effet, en régle générale il est nécessaire de disposer, au moment de soumettre les données aux traitements et aux analyses, d'informations uniques et non redondantes concernant le même état de fait afin d'éviter que les répétitions comportent des distorsions et des biais dans les résultats. Par exemple, on ne peut pas comparer les carrières d'une population d'enseignants universitaires si les données disponibles ne contiennent pas une information unique pour chaque segment de carrière mais plusieurs mentions de chacun des segments issues de différentes sources. Dans ce cas, une agrégation préalable des données est indispensable à leur analyse en transformant les mentions en états de fait et en indiquant le degré de probabilité de chacun de ces derniers.

Si le principe est clair, son application dans un système d'information qui implémente les deux niveaux épistémologiques est plus délicate. Nous nous limiterons ici a évoquer quelques aspects de ce problème tout en renvoyant un traitement plus développé à un autre contexte. D'abord la question de la gestion de l'identité des objets dont font état les *factoïds* : s'agit-il bien de la même personne et a-t-on le droit d'agréger différentes mentions de segment de carrière dans une même information se rapportant à un seul individu ? Si la question paraît simple à régler pour le cas de Galileo Galilei — bien qu'il existe deux personnages historiques connus sous ce nom, l'un né au XIV^{e} siècle, l'autre au XVI^{e}—, elle est bien plus délicate pour une source mentionnant, par exemple, un individu appelé « Johannes Teutonicus » dans la première moitié du XIII^{e} siècle.

Puis la manière de documenter et de modéliser le processus complexe d'extraction et d'agrégation de l'information présente dans les sources en appliquant les différentes méthodes de la critique historique, telles la conjecture, l'inférence, la contextualisation, etc. L'ontologie du projet *symogih.org* propose une approche de cette question qui permet d'expliciter, au niveau du sourçage de l'information (classe *Sourcing* de la fig. 1), la méthode adoptée au moment de produire l'unité de connaissance, ainsi que le degré de fiabilité de la source utilisée. Ce procédé permet aux autres historiens de mesurer la valeur de la donnée produite, sa fiabilité. Une ou plusieurs sources, voire des contenus préalablement crées dans le système d'information, peuvent être associées à une même information. On peut renseigner également la certitude de l'identification de chaque objet associé à l'information ainsi que les libellés des objets tels qu'ils apparaissent dans la source.

Cette méthode n'est toutefois pas vraiment satisfaisante car elle ne permet pas d'expliciter de manière distincte et précise la fiabilité et le degré de véracité du contenu de chaque source ou *factoïd* mobilisé dans le processus d'intégration qui produit l'information. Cette limite importante dépend entre autres d'une sorte de « raccourci » ontologique qui a été introduit inconsciemment dans la modélisation des classes abstraites du modèle *symogih.org* et qui apparaît avec évidence en le comparant avec des ontologies ayant une structure similaire.

Le modèle présente la même structure cognitive que la *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE), une ontologie de haut-niveau conçue comme moyen d'étudier les structures essentielles du langage naturel en tant que compréhension de la réalité (Masolo *et al.*, 2003). En particulier, la catégorie des *endurants*, entités qui subsistent avec la même essence dans le temps, tels les objets physiques, les concepts ou les êtres humains, équivaut à la classe *Object* de l'ontologie *symogih.org*, tandis que les *perdurants* de DOLCE, entités qui se développent dans le temps tout en se modifiant d'un instant à l'autre, tels les événements ou les processus, correspondent aux unités de connaissance : ils expriment une relation subsistant entre objets à un moment donné du temps, qu'il soit ponctuel ou étendu (Beretta 2017).

La même perspective cognitive se retrouve dans le modèle conceptuel du CIDOC CRM (Doerr 2003) créé pour permettre l'interopérabilité des données produites dans le domaine de la conservation des biens culturels (standardisée par l'ISO en 2006) : les classes *Persistent Item* et *Temporal Entity* correspondent respectivement aux objets et aux informations de *symogih.org* mais avec une nuance importante au point de vue épistémologique. En tant qu'unité de connaissance, une « information » du modèle *symogih.org* encapsule à la fois le *perdurant* comme tel, l'événement en tant qu'état de fait, et l'assertion de l'historien le concernant, fondée sur les sources. Certes, grâce à la classe *Sourcing,* les sources associées à une information peuvent être multiples, et pour chacune d'entre elles les paramètres de fiabilité et la méthode adoptée pour produire la connaissance peuvent être renseignés individuellement mais ce processus est unique pour l'ensemble de l'information car celle-ci est conçue, avec tous ses rôles, comme une unité de connaissance ou assertion.

En revanche, dans le CIDOC CRM (version 6.2.1, octobre 2015) le sourçage se fait pour chaque propriété (ou rôle pour employer le terme équivalent de l'ontologie *symogih.org*) en produisant des instances de la classe *E13 Attribute Assignement* qui, en tant que sous-classe de *E7 Activity*, dispose de tout un ensemble de propriétés permettant d'expliciter, grâce à un typage adéquat, la manière de produire la connaissance. Le sourçage sera, pour ainsi dire, accroché aux rôles/propriétés, ce qui permettra d'apporter, pour chaque composante de l'information, c'est-à-dire pour chaque propriété, un sourçage spécifique et une description détaillée de sa fiabilité, de son degré de véracité et de la méthode adoptée en termes de critique historique. Une extension du CIDOC CRM en cours de développement, CRMinf, entendue comme ontologie formelle dont la finalité est d'intégrer les « metadata about argumentation and inference making in descriptive and empirical sciences »[14], permet de détailler encore plus ce processus de production de connaissance, mais sa présentation sort du cadre de cette contribution.

Quant au modèle du CIDOC CRM, il convient de souligner que la classe *Temporal entity* représente un « fait », un *state of affairs* comme tel, par exemple une naissance ou un enseignement, et non sa description ou une assertion le concernant. Une entité temporelle n'est donc jamais sourcée directement, elle n'est, par principe, pas sourçable : c'est au niveau de l'association des objets qui participent à cet événement opérée par les propriétés, ainsi que grâce à la définition de sa classe d'appartenance, qu'on définit son identité. On retrouve l'assertion de l'historien concernant tel « fait », telle unité de connaissance, dans l'ensemble des propriétés d'une entité temporelle, réuni dans un graphe. Plus précisément, l'assertion de véracité qui résulte de l'application de la méthode historique est déplacée de l'entité temporelle (*symogih.org*) vers ses propriétés (CIDOC CRM), ce qui permet un sourçage de l'information beaucoup plus fin et pertinent.

Soulignons également que l'entité temporelle équivaut à une « information » de l'ontologie *symogih.org* (une fois enlevée la dimension d'assertion) et non à un « contenu » car il ne s'agit pas de modéliser ce que dit la source d'un événement mais l'événement comme tel. Il ne paraît donc pas légitime de modéliser le *Factoïd* en tant que sous-classe de *Temporal Entity* comme le proposent les auteurs de l'article mentionné précédemment (Bradley 2015, 92-93) car un *factoïd* exprime l'assertion d'une source, non le « fait » lui-même. Dans le contexte du CIDOC CRM, on pourra modéliser un *factoïd* (de même qu'un « contenu » de *symogih.org*) en tant que classe équivalente à *E89 Propositional Object* ou, si on souhaite aussi encapsuler sa formulation précise sous forme de texte, à la classe *E73 Information Object*[15].

L'avantage de cette méthode de modélisation réside dans le fait de permettre d'enregistrer les différents points de vue des sources par rapport à une même propriété : « The CRM has been designed to accommodate alternative opinions and incomplete information, and therefore all

[14] http://www.cidoc-crm.org/crminf/ .

[15] http://www.cidoc-crm.org/Version/version-6.2.1 .

properties should be implemented as optional and repeatable for their domain and range »[16]. Plusieurs instances de la même propriété peuvent être associées à une même *Temporal Entity* dans le système d'information afin d'associer plusieurs candidats différents pour le même rôle. L'utilisation de plusieurs instances de la classe *E13 Attribute Assignement* permettra ensuite de sourcer chacun d'entre eux tout en fournissant les paramètres de véracité de cette alternative. Avec cette méthode de modélisation on pourra aisément traiter le cas d'un homicide dans lequel l'identité de l'assassin n'est pas claire, en associant avec le même rôle « être l'assassin » plusieurs personnes susceptibles d'être coupables tout en articulant la discussion au niveau du sourçage, c'est-à-dire de l'assertion concernant chaque instance du rôle/propriété.

Concernant l'incertitude des dates, autre question délicate en modélisation des données historiques, le projet *symogih.org* présentait aussi des limites que permet de dépasser le CIDOC CRM. Certes, l'ontologie du projet insistait sur la notion de datation en tant que processus d'établissement d'une date plus ou moins probable, ou d'une fourchette de dates possibles, en adéquation avec les méthodes habituelles de la recherche historique. Ce processus, et non une simple date, était exprimé avec la classe *Dating* possédant ses propriétés spécifiques (figure 1). Toutefois le même problème de sourçage évoqué précédemment s'appliquait aussi à ce domaine, alors qu'en adoptant la conception du CIDOC CRM chaque propriété associant une date à une étendue plus ou moins précise du temps, exprimée avec des règles précises, peut être soumise à une discussion « argumentée » à partir des sources, en utilisant des instances de la classe *E13 Attribute Assignement.* Aussi le CIDOC CRM introduit une série de propriétés inspirées de la logique temporelle de Allen (Allen 1983) (*P114 is equal in time to* et propriétés suivantes) qui permet de décrire des relations d'événements et leurs positions relatives dans le temps, et ce, même si on ne dispose pas d'éléments suffisants pour proposer une datation même approximative.

Reste une question : ce processus d'extraction de contenus des sources, puis d'agrégation en informations non redondantes est-il automatisable ? Il apparaît de ce qui précède que pour effectuer une telle opération il faut disposer, d'une part, d'une modélisation robuste du processus d'extraction des informations des sources et d'évaluation de la fiabilité de chacune d'elles, et, d'autre part, concevoir les algorithmes adaptés aux principes de la méthode critique permettant de pondérer le degré de véracité des contenus disponibles afin de les agréger dans un résultat final. La discussion de ce problème, qui a donné lieu à des tentatives expérimentales au sein du projet *symogih.org*, dépasse le cadre de cette contribution et sera l'objet d'une publication ultérieure.

5. L'intégration avec le CIDOC CRM et le consortium *Data for History*

Cette réflexion au sujet de l'alignement de l'ontologie du projet *symogih.org* avec le CIDOC CRM, menée dès 2014, a permis de mettre en évidence les spécificités et les originalités de la modélisation collaborative réalisée au sein du projet (Beretta 2017) et en même temps d'en comprendre les limites, notamment celles évoquées dans les pages précédentes. Aussi, le fait de publier en RDF des données modélisées dans une ontologie spécifique, non alignée explicitement avec un standard reconnu, ne permettait pas de garantir leur interopérabilité, en tout cas pas au sens de l'article *Interoperability* des principes FAIR.

Afin de dépasser ces limites, les auteurs de cette contribution ont entamé en 2016 une collaboration active avec le *Special interest group* (SIG) qui maintient le CIDOC CRM et participent depuis activement au développement de ce dernier. Ces deux années ont été un temps intense d'apprentissage de ce modèle conceptuel, devenu une spécification ISO en 2006, et de compréhension de ses arcanes qui restent parfois peu accessibles tant à cause d'une définition fort

[16] *Definition of the CIDOC Conceptual Reference Model*, Version 6.2.1, October 2015, p. xiii.

abstraite de la norme que de la nécessité de l'interpréter à l'aune d'une tradition orale qu'on ne découvre qu'en participant aux réunions du SIG.

Deux éléments en particulier méritaient une définition plus précise. D'une part, l'articulation décrite ci-dessus de la notion d'unités de connaissance propre au projet *symogih.org* avec celle des *temporal entities* du CIDOC CRM, comprenant les notions de sourçage et de gestion de la temporalité. D'autre part, la démarche collaborative adoptée au sein du projet *symogih.org* dans la production progressive du modèle, couplée avec la volonté de permettre une certaine souplesse dans la production de nouveaux types d'information, avait amené à produire un ensemble de classes d'informations conçues à plat, hors de toute considération taxonomique. En revanche, la notion de sous-classe appliquée dans le CIDOC CRM permet, d'un côté, de construire des arbres de spécialisation avec un 'héritage des propriétés et, d'un autre côté, de garantir la cohérence de l'ensemble de l'ontologie, ce qui est essentiel afin de rendre interopérables les données issues de différents systèmes d'informations conçus avec différents niveaux de spécialisation. Un processus d'alignement des types d'information de *symogih.org* avec les classes de *temporal entities* du CIDOC CRM est actuellement en cours.

Aussi, si on constate que plusieurs projets ont adopté le CIDOC CRM comme cadre conceptuel de référence et que son usage est de plus en plus répandu[17], cette adoption se fait souvent sous la forme de développement d'extensions locales non suffisamment évaluées dans leur approche méthodologique, voire carrément non publiées, ce qui représente un obstacle majeur pour l'interopérabilité de données visée pourtant par la plupart de ces projets. Il paraissait donc judicieux de tenter de fédérer ces efforts, notamment dans le domaine des données géo-historiques, d'autant que le SIG du CIDOC CRM a entamé récemment l'élaboration d'une extension de l'ontologie dédiée à la vie sociale. Il s'agissait enfin de mettre à la disposition de la communauté des chercheurs dix ans d'expérience en matière de modélisation collaborative de données historiques propre au projet *symogih.org*. Et d'offrir en même temps un lieu d'apprentissage du CIDOC CRM permettant aux projets de s'approprier l'ontologie plus rapidement que par la lecture de la norme.

Cette volonté a entraîné une réflexion sur les outils disponibles proposant à la fois des fonctionnalités d'alignement d'ontologies et à la fois de discussion collaborative sur les modèles de données. Après avoir évalué les outils existants, et notamment WebProtégé[18], il a semblé opportun, au vu de leurs limites, de mettre en place une application en ligne permettant d'aligner les modèles de données des systèmes d'information avec le CIDOC CRM sous la forme d'un environnement de gestion d'ontologie baptisé OntoME (*Ontology Management Environment)*[19].

La fonction de cet outil en ligne est double. D'une part, permettre à un projet d'importer son propre modèle de données et de l'aligner avec le CIDOC CRM et ses extensions puis d'exporter l'alignement effectué et l'utiliser pour exposer les données du projet au format RDF, même si dans le système d'information du projet celles-ci continuent à être produites dans un autre format, notamment relationnel. D'autre part, permettre la création de sous-ensembles de classes et de propriétés à partir des ontologies présentes dans l'application, en définissant de cette manière des « profils applicatifs » qui pourront être exportés et servir directement comme modèles de données dans des systèmes d'information distribués permettant aux chercheurs de réaliser leur propre projet. Dans ces profils, il sera possible d'ajouter également des libellés ou des explications supplémentaires pour chaque classe et propriété, dans plusieurs langues, afin de les rendre plus intelligibles aux utilisateurs « locaux », voire de définir des sous-classes et sous-propriétés

[17] https://doc.biblissima.fr/ontologie-biblissima ; https://masa.hypotheses.org/500 .

[18] https://protege.stanford.edu/products.php#web-protege .

[19] http://ontome.dataforhistory.org/ .

spécifiques au sous-domaine de chaque projet, tout en les alignant, toujours dans OntoME, avec les classes et propriétés du CIDOC CRM. Ces classes et propriétés plus spécialisées permettront de produire des données directement interopérables, à un niveau plus élevé d'abstraction, grâce à l'inscription dans un arbre de spécialisation.

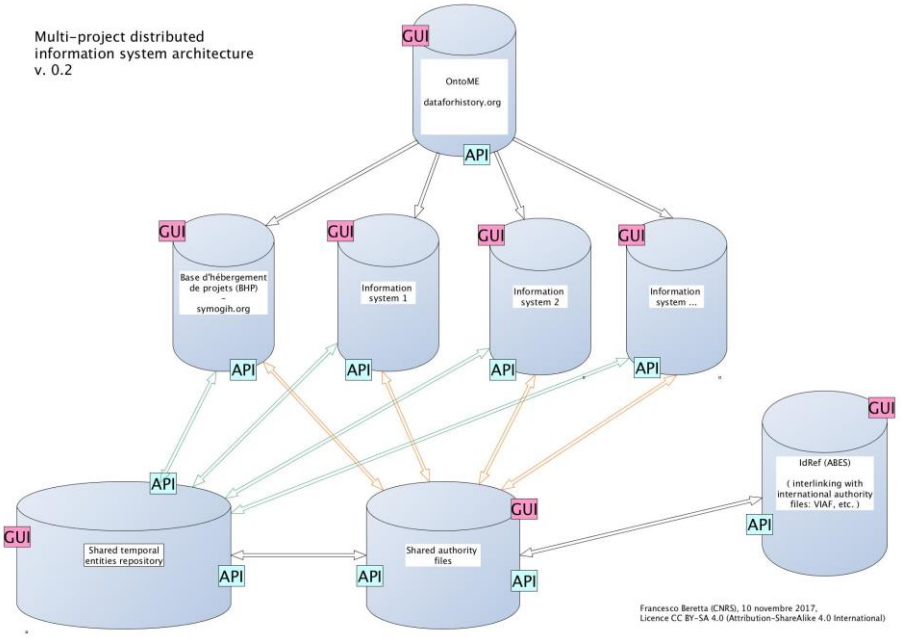

Figure 3. *Architecture d'un système d'information distribué utilisant OntoME*

La dimension collaborative de la production de modèles que permet OntoME invite les concepteurs de systèmes d'informations à participer à une démarche d'interopérabilité en vue de créer un système d'information distribué (Fig. 3) au sein duquel le modèle des données est défini dans l'application, de manière flexible et au niveau de spécialisation souhaitée par chaque projet, tandis que les instances des données elles-mêmes sont produites localement. Un alignement avec les notices d'autorité des IdRef de l'ABES pourra être mis en place afin d'aligner non seulement le modèle mais encore, dans la mesure du possible, les individus que contiennent les systèmes d'information distribués, par ex. les lieux, les personnes, les concepts, etc. Enfin, l'ensemble des données des projets pourra être mis à la disposition des autres chercheurs et du public, au format RDF, soit directement *via* une API locale par projet, soit en moissonnant les données rendues accessibles et en les versant dans un entrepôt commun sous forme de *triplestore* ou de moteur de recherche indexé.

Cette vision, initiée par le Pôle histoire numérique du LARHRA, est portée et partagée par une communauté de plus en plus large en voie de structuration dans le cadre du consortium *Data for History*[20]. En préambule à sa constitution, deux ateliers à périmètre français ont d'abord été organisés à ce sujet, l'un en juin 2016 à Lyon et l'autre en mars 2017 à Brest. Le consortium a ensuite été officiellement constitué en novembre 2017, lors d'un workshop organisé à l'École normale supérieure de Lyon avec la participation d'une trentaine d'historiens, d'historiens de l'art, d'archéologues et de spécialistes des sciences de l'information venus de six pays européens. L'équipe du référentiel IdRef de l'ABES, ainsi que celle du Service interministériel des Archives de France ont manifesté leur intérêt et leur soutien pour cette initiative qui permet de rapprocher le travail des producteurs de métadonnées des documents conservés dans les dépôts d'archives et les bibliothèques à celui des historiens, dans une logique *open data*. Porté par des rencontres régulières, le consortium se structure et se construit progressivement. Le deuxième atelier de *Data for History*

[20] http://dataforhistory.org/ .

s'est déroulé à l'université Jean-Moulin Lyon 3 les 24 et 25 mai 2018 en parallèle de la rencontre du SIG du CIDOC CRM. Une rencontre d'un groupe plus réduit de membres du Consortium, mais augmentée de nouvelles personnes intéressées, a eu lieu à la suite de la présentation d'un panel *Data for History* sur l'interopérabilité des données présenté au congrès de l'EADH à Galway en décembre 2018[21]. Une prochaine rencontre est prévue à Leipzig le 4 et 5 avril 2019.

5. Conclusion

L'expérience du projet *symogih.org* a consisté à réaliser un système d'information multi-projet qui permet la production et le stockage collaboratif de données sous la forme de textes, de bases de données et d'informations géo-historiques, et leur publication sur le web. Cette plateforme offre ainsi un environnement virtuel de travail adapté à toutes les étapes du traitement des données : saisie à partir du dépouillement des archives, visualisation et interrogation via des outils numériques (statistiques, analyse de réseaux, cartographie, etc.), exportation des résultats et publication sur internet. C'est au stade de l'interrogation et de l'analyse des données que s'applique le questionnement d'une recherche historique, et qu'est produite la connaissance. Il s'agit en quelque sorte d'une surcouche qui se greffe sur le système d'information tout en étant distincte de celui-ci alors qu'il a été conçu précisément pour stocker des données produites selon un modèle le plus objectif possible (et donc le moins possible influencé par les problématiques des chercheurs) afin que ces données puissent être réutilisées pour de nouvelles recherches.

Le fait de gérer les projets au sein d'un environnement de recherche mutualisé permet en même temps d'assurer la pérennisation des données après la fin du financement, leur réutilisation par d'autres projets, leur publication sur internet et leur sauvegarde sur les infrastructures mises à disposition par la TGIR Huma-Num. La plateforme permet ainsi de mettre à la disposition des projets de recherche un plan de gestion des données de plus en plus en accord avec les principes FAIR qui permet aux utilisateurs de se concentrer sur leur recherche en histoire en les dispensant de la conception de ce plan. La plateforme est également susceptible de permettre la mise en place d'un modèle économique dans lequel l'apport de chaque projet (en termes de ressources humaines ou financières) permet de développer et de maintenir un système d'information modulaire au service de l'ensemble des projets, y compris ceux qui ne disposent pas ou plus —en l'état— de financements.

Dans le contexte du développement du web sémantique et de la publication d'un volume croissant de données, le consortium *Data for History* fait un pas de plus et propose de constituer un système d'information multi-plateformes autour d'un environnement de gestion collaborative d'ontologie, mis à la disposition du consortium par le Pôle histoire numérique du LARHRA, l'application OntoME. Centrée autour de l'ontologie CIDOC CRM et d'autres ontologies standardisées favorisant l'interopérabilité, cette application permettra aux projets de recherche, notamment en histoire (ANR, ERC, projets collectifs ou individuels, thèses, etc.), de documenter et d'expliciter le modèle utilisé pour la production des données. Cette démarche facilitera l'interopérabilité des données, leur réutilisation pour de nouvelles recherches et leur pérennisation, en accord avec les principes FAIR.

L'organisation en consortium vise aussi à promouvoir l'activité d'une communauté internationale de chercheurs en SHS et d'informaticiens intéressés par un modèle partagé de données, dans lequel pourront s'inscrire les spécificités de chaque projet. L'application OntoME se trouve au cœur de ce processus en permettant de faciliter la compréhension des différents modèles de données liés au domaine de la recherche en SHS et, parallèlement, de soutenir un processus de mutualisation et d'échange de données.

[21] https://eadh2018eadh.wordpress.com/ .

Cette vision soulève, dans le même temps, la question du modèle de financement d'une infrastructure de systèmes d'informations collaboratifs et distribués ou mis en réseau. Elle remet en question le modèle actuellement hégémonique en Europe du financement d'infrastructures par projet et invite les acteurs à créer des synergies autour de centres régionaux tels, en Autriche, l'infrastructure GAMS à l'Université de Graz (http://gams.uni-graz.at) portée par le Centre de modélisation des connaissances (*Zentrum für Informationsmodellierung*) ou, en Allemagne, le Centre des données historiques du Land de Saxe-Anhalt à l'Université de Halle-Wittemberg (*Historisches Datenzentrum Sachsen-Anhalt*). C'est au niveau de ces centres de recherche régionaux qu'il paraît le plus judicieux de situer la gestion de systèmes d'information disciplinaires, tout en facilitant leur mise en réseau et en les adossant, pour l'infrastructure et dans la mesure du possible, aux centres de ressources numériques nationaux, tels DARIAH-DE pour l'Allemagne (https://de.dariah.eu/), CLARIAH pour les Pays-Bas (https://www.clariah.nl) et la TGIR Huma-Num en France (https://www.huma-num.fr/).

Bibliographie

Akoka, J. et Comyn-Wattiau I. (2001). Conception des bases de données relationnelles, Vuibert, Paris.

Alerini, J., Lamassé, S. (2011). Données et statistiques. L'avenir du travail en ligne pour l'historien. *Les historiens et l'informatique.* Un métier à réinventer, Rome, École française de Rome, p.171-187.

Allen, J. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26, p. 832-843.

Audibert L. (2009). *Bases de données : de la modélisation au SQL*, Ellipses, Paris.

Beretta F. (2014). Exploration du site web *scholasticon.fr* : une application de la méthode SyMoGIH (Système modulaire de gestion de l'information historique). *La prosopographie au service des sciences sociales.* Lyon, CEROR, p. 289-310.

Beretta F. (2016). Pour une annotation sémantique des textes: le projet *symogih.org* et la *Text encoding initiative*. *Bruniana & Campanelliana,* vol. 22, n°. 2, p. 453-465.

Beretta F. (2017). L'interopérabilité des données historiques et la question du modèle : l'ontologie du projet SyMoGIH. *Enjeux numériques pour les médiations scientifiques et culturelles du passé*, Paris, Presses Universitaires de Paris Nanterre, p.87-127.

Beretta F. et Butez C. (2013). Un SIG collaboratif pour la recherche historique. *Géomatique Expert*, n. 91, p.30-35 / n.92, p.48-54.

Beretta F, Letricot R. (2017). Le portail XML du projet symogih.org : un projet d'édition numérique collaborative de sources et d'informations historiques », *Humanités numériques et construction des savoirs*, London, ISTE Editions, p.125-143.

Beretta F., Vernus P. (2012). Le projet SyMoGIH et la modélisation de l'information : une opération scientifique au service de l'histoire. *Les Carnets du LARHRA*, n.1, p.81-107.

Bradley J., Pasin M. (2015). Factoid-based prosopography and computer ontologies: Towards an integrated approach. *Literary and Linguistic Computing*, vol. 30, issue 1, p.86–97.

Courtin, A., Minel, J.-L. (2017). Propositions méthodologiques pour la conception et la réalisation d'entrepôts ancrés dans le Web des données. *Enjeux numériques pour les médiations scientifiques et culturelles du passé*, Paris, Presses Universitaires de Paris Nanterre, p.53-86.

Dedieu J.-P. (2004). Les grandes bases de données : una nouvelle approche de l'histoire sociale: le système Fichoz. *HISTORIA. Revista de Faculdage de Letras da Universidate do Porto* s.3, vol.5, 2004, p.101-114.

Doerr M. (2003). The CIDOC CRM. An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine* vol. 24, number 3, p.75-92.

Domingue J., Fensel D., Hendler J. A., eds (2011). *Handbook of semantic web technologies*. Vol. 1. Foundation and technologies (Berlin / Heidelberg, Springer.

Eide Ø. (2014-2015). Ontologies, Data Modeling, and Tei. *Journal of the Text encoding initiative*, vol. 8.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A. (2003). *WonderWeb Deliverable D18 Ontology Library* (final), Trento, Laboratory For Applied Ontology.

Meroño-Peñuela, A., Ashkpour A., van Erp M., Mandemakers K., Breure L., Scharnhorst A., Schlobach S., van Harmelen F. (2015). Semantic Technologies for Historical Research: A Survey. *Semantic Web*, 6, p. 539-564.

Segaran T., Evans C., Taylor J. (2009). *Programming the Semantic Web*. Beijing e.a., O'Reilly.

Soutou Christian, *UML 2 pour les bases de données*, Paris, Eyrolles, 2007.

West M. (2010). *Developing high quality data models*, Morgen Kaufman, Burlington, MA.

Wilkinson M.D., Dumontier M., Aalbersberg I.J., Appleton G., Axton M., Baak A., Blomberg N., et al. (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 3 (March 15, 2016): 160018.