

Le Web sémantique et l'histoire

Par Philippe Michon

Janvier 2016

Mise en contexte:

L'historien qui souhaite se lancer dans l'aventure du Web sémantique est confronté à une quantité impressionnante de ressources disparates puisqu'il n'existe aucun ouvrage qui traite de l'association entre les disciplines historiques et cette nouvelle conception du Web. Ce court article en deux parties est une première tentative visant à combler cette lacune. Le présent texte décrit la structure du Web sémantique et les différents concepts qui l'entourent. L'historien acquerra ainsi les connaissances nécessaires pour entamer le dialogue avec les informaticiens qui pourront mettre en place un système de données liées efficace s'appuyant sur les recommandations du *World Wide Web Consortium* (W3C). Un second texte (en préparation) présentera l'application d'une démarche sémantique à un corpus de données historiques et patrimoniales.

Première partie:

L'idée du Web sémantique découle d'un texte de Tim Berners-Lee publié en 2001 sous le titre *The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*¹. Dans ce texte de prospective, l'auteur rappelle que le Web (qu'il avait inventé une décennie plus tôt²) s'était développé en silos. Il propose plutôt une approche fondée sur un contenu dont le sens s'exprime de manière explicite : « The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users.³ ».

Dans le Web traditionnel, la spécificité de chaque page web impose un travail colossal pour créer des ponts entre les institutions gérant des bases de données. Une telle situation peut sembler

¹ Tim Berners-Lee, James Hendler et Ora Lassila, « The Semantic Web », *Scientific American*, 17 mai 2001, <http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>, consulté le 27 août 2015.

² World Wide Web Foundation, *History of the Web*, <http://webfoundation.org/about/vision/history-of-the-web>, 2012, consulté le 8 janvier 2016.

³ Berners-Lee *et al*, *loc. cit.*

sans importance pour l'utilisateur ordinaire, mais elle pose un défi fondamental pour les institutions culturelles qui sont appelées à fournir un accès intégré à leurs collections⁴. Le Web sémantique, tel que conçu par Berners-Lee, s'est donc imposé depuis quelques années comme étant la seule avenue qui permet cette agrégation⁵. Les principes du Web sémantique nous amènent donc à revoir la manière de concevoir les bases de données – présentement structurées en fonction d'un contenu ou d'une problématique spécifique – en tenant compte de l'ensemble des possibilités de notre champ disciplinaire et même au-delà.

1. Les cinq critères des données ouvertes et liées

Le W3C, – organe de référence pour l'élaboration de standards ouverts visant à assurer le développement du Web⁶ – a pris en main la définition d'un Web sémantique dont les principaux fondements sont le système des données ouvertes et liées cinq étoiles, chacune de ces étoiles étant essentielle pour qu'un système réponde à cette définition. Nous reprenons ici la traduction de ces cinq niveaux proposées par le gouvernement du Canada⁷.

- Première étoile : fournir des données avec une licence ouverte. Selon Statistique Canada, une licence ouverte « contient peu de restrictions quant à la façon dont les données peuvent être utilisées et permet spécifiquement leur distribution ultérieure dans le cadre d'activités commerciales ainsi que non commerciales. Il n'y a aucun frais pour ce genre d'utilisation.⁸ » Une licence ouverte peut toutefois imposer des restrictions. Par exemple, elle peut spécifier que la source ne peut être modifiée et que son utilisation doit se limiter à des activités non commerciales. Néanmoins, l'utilisation du contenu demeure toujours gratuite.

⁴ Online Computer Library Center, *Library, Archive and Museum Collaboration*, <http://www.oclc.org/research/activities/lamsurvey.html>, 30 novembre 2011, consulté le 27 août 2015.

⁵ De nombreuses organisations internationales, dont le *World Wide Web Consortium* (W3C), développent des méthodologies et des logiciels visant à faciliter l'implantation des données comme on peut le constater dans la bibliographie qui accompagne ce texte. Aucune alternative ne semble pouvoir concurrencer ce modèle pour le moment ou dans un avenir proche.

⁶ World Wide Web Consortium, *World Wide Web Consortium (W3C)*, <http://www.w3.org/>, 2016, consulté le 8 janvier 2016.

⁷ Gouvernement du Canada, *Cote de degré d'ouverture des données*, <http://ouvert.canada.ca/fr/cote-degre-douverture-des-donnees>, 25 avril 2013, consulté le 8 janvier 2016.

⁸ Gouvernement du Canada et Statistique Canada, *Entente de licence ouverte de Statistique Canada - Foire aux questions (FAQ)*, <http://www.statcan.gc.ca/fra/reference/licence-faq-fra#a1>, 25 janvier 2013, consulté le 8 janvier 2016.

- Deuxième étoile : présenter les données dans un format structuré. Cette étape est nécessaire pour qu'un programme puisse lire les données compilées. Les données non structurées (texte libre) posent d'autres types de défis aux ordinateurs et ajouteraient d'autres couches de complexité à un système sémantique.
- Troisième étoile : miser sur des formats ouverts. Par exemple, *Excel* étant un logiciel propriétaire produit par *Microsoft*, il faut le posséder (ou un autre logiciel disposant d'un module de conversion) pour lire les formats *.xls* ou *.xlsx*. Un contenu doit être offert avec un format qui ne sera pas modifié au gré d'une entreprise pour en limiter l'accès et qui peut être lu par divers logiciels.
- Quatrième étoile : proposer un *Uniform Resource Identifier* (URI) pour chaque élément de notre base de données et l'inscrire dans un modèle *Resource Description Framework* (RDF). À titre d'exemple, chaque donnée d'une cellule d'un chiffrier électronique (sauf pour les dates) devrait pointer vers un URI, comme nous le verrons plus loin. Rappelons qu'un ordinateur ne comprend pas le sens des chaînes de caractères. Si deux personnes insèrent le terme « Jacques Cartier » dans leurs bases de données, il n'existe aucun lien entre ces deux chaînes sauf une équivalence de caractères. Pour créer un lien, il faut utiliser un URI, un identifiant unique, qui permet de référer au même élément sur le Web. Il peut s'agir d'un concept, d'une idée, d'une personne, d'un événement, d'un objet, d'un document, etc. Si l'on veut identifier « Jacques Cartier » dans notre réseau de données, on se doit d'utiliser un URI qui fait référence exclusivement à l'explorateur. Par exemple, il est possible d'utiliser l'URI proposé dans le *Library of Congress Subject Heading* (LCSH), catalogue qui identifie les différentes composantes des collections de l'institution. Dans ce système, l'explorateur Jacques Cartier est identifié par l'URI <http://id.loc.gov/authorities/names/n50080987>⁹. Il est aussi possible de créer ses propres URI et les héberger sur un serveur en assurant leur pérennité. De son côté, RDF propose une syntaxe simple et uniforme¹⁰ qui repose sur des triplets. Cette forme permet rapidement et

⁹ Library of Congress, *Cartier, Jacques, 1491-1557 - LC Linked Data Service*, <http://id.loc.gov/authorities/names/n50080987.html>, consulté le 27 août 2015.

¹⁰ Ruben Verborgh, *Federated SPARQL queries in your browser*, <http://ruben.verborgh.org/blog/2015/06/09/federated-sparql-queries-in-your-browser/>, 9 juin 2015, consulté le 28 août 2015.

simplement d'expliciter le sens d'un concept, de créer des liens entre eux et s'applique à des contenus de toutes les disciplines¹¹. Un triplet RDF prend toujours la forme *Sujet-Prédicat-Objet*. Pour reprendre l'exemple de Jacques Cartier, on pourrait proposer le triplet suivant : *Jacques Cartier-fonction-explorateur*. C'est l'accumulation de triplets autour d'un sujet qui donne du sens à celui-ci. Par exemple, le triplet précédent permet de s'assurer qu'il s'agit de Jacques Cartier l'explorateur et non du « député de la circonscription de Surrey à la Chambre d'assemblée du Bas-Canada de 1804 à 1809¹² » qui aurait comme triplet *Jacques Cartier-fonction-député*. Il est aussi possible de différencier les deux hommes par leurs années de naissance et de décès.

Cinquième étoile : utiliser les URI et RDF pour créer des ponts avec d'autres jeux de données. Comme il fut mentionné précédemment, il est plus simple de créer soi-même ses URI au lieu de pointer vers des bases de données externes. Cependant cette méthodologie reproduit la structure en silo à laquelle le Web sémantique tente d'échapper. L'utilisation de l'URI du LCSH pour identifier Jacques Cartier devient donc la pratique pour obtenir cette cinquième et dernière étoile. Cet exemple s'inscrit dans une approche sémantique puisqu'on associe une base de données avec celle de la *Library of Congress*. Par la suite, quand une autre institution associe une de ses données avec le même URI, un lien est automatiquement créé entre les données des deux institutions par le biais de cet URI commun.

2. Le passage vers les données liées

La majorité si ce n'est la totalité des institutions culturelles de grande ou de moyenne importance possèdent déjà des bases de données relationnelles qu'elles peuvent ajouter à l'ensemble des données ouvertes et liées¹³. Ce processus de transformation requiert la mise en place d'une stratégie que nous expliquerons à partir d'une étude de cas dans le deuxième volet de

¹¹ *Ibid.*

¹² Ministère de la culture et des communications du Québec, *Cartier, Jacques*, <http://www.patrimoine-culturel.gouv.qc.ca/rpcq/detail.do?methode=consulter&id=17323&type=pge#.Vd-sQK01uzx>, 2013, consulté le 28 août 2015.

¹³ Richard Cyganiak et Anja Jentzsch, *The Linking Open Data cloud diagram*, <http://lod-cloud.net/>, 2014, consulté le 31 octobre 2014.

cette présentation. Elle requiert aussi l'utilisation de différents outils que nous présenterons ici en privilégiant les outils libres et gratuits.

La gestion de données dans un objectif de sémantisation doit nécessairement passer par une analyse globale. *OpenRefine*¹⁴ est l'outil tout désigné pour cette tâche. Développé par *Google*, mais maintenant disponible sous licence libre, *OpenRefine* permet d'effectuer trois grandes étapes obligatoires dans le processus de liaison. Premièrement, il permet d'explorer efficacement nos données. On peut regrouper nos données selon différentes facettes et ainsi repérer facilement les incongruités ou les faiblesses de celles-ci. Deuxièmement, il permet de nettoyer et de transformer nos données. Troisièmement, en lien avec un outil complémentaire, il permet d'identifier des URI existants pour nos données.

L'analyse d'un corpus de données révélera presque assurément certaines erreurs dans la structure des données. Une des bonnes pratiques est d'atomiser les données, soit de veiller à ce que chaque cellule possède une seule information. Par exemple, au lieu de présenter le concept « Jacques Cartier » sous le champ « nom », il est préférable de le séparer dans des champs « prénom » et « nom », lesquels pourront être regroupés au besoin dans l'ordre requis (nom, prénom ou vice versa). Les champs « description », quoique souvent utiles pour synthétiser l'information sous forme narrative, ne doivent pas remplacer d'autres champs. Si l'information contenue dans la description doit être liée, il faut alors la reproduire dans des champs plus spécifiques. Enfin, l'analyse peut révéler des colonnes de notre tableau qui contiennent très peu d'informations. Il est alors préférable, si la logique du projet le permet, de fusionner ces colonnes avec d'autres ou de les retirer des données à fournir en données liées. Une quantité importante de champs vides peut causer différents problèmes d'interopérabilité.

¹⁴ Tom Morris, *OpenRefine*, <http://openrefine.org/>, 2015, consulté le 5 novembre 2015.

Google refine schoenberg csv Permalink

225901 records

Extensions: Named-entity recognition | Freebase | RDF

Facet / Filter Undo / Redo

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts

All	ID	Duplicates	Cat./Trans. date	Primary seller	Secondary seller	Institution or co.	Buyer	Catalogue ID	Lot or cat.	Price	Currency	Sold	Sc
1	219	219,340,117,538,189,000	19970617	Sotheby's				Western manuscripts and miniatures (LN7375)	62	71445			
2	220		19970617	Sotheby's				Western manuscripts and miniatures (LN7375)	61	27513.75			
3	221		19970617	Sotheby's				Western manuscripts and miniatures (LN7375)	60	30360			
4	222	22,263,919,885,994,100,000,000,000,000,000	19970617	Sotheby's				Western manuscripts and miniatures (LN7375)	69	60555			
5	223	11,223	19970617	Sotheby's				Western manuscripts and miniatures (LN7375)	71	5692.5			
6	224	224,877,241,162	19970617	Sotheby's				Western manuscripts and miniatures (LN7375)	66	13282.5			
7	226	226,193,868	19970616	Sotheby's	Beck Collection			The Beck Collection of Illuminated Manuscripts (LN7362)	10	347325			
8	227	227,384,510,137,147,000,000,000,000	19970616	Sotheby's	Beck Collection			The Beck Collection of Illuminated Manuscripts (LN7362)	11	4179975			
9	228	22,824,245,001	19970616	Sotheby's	Beck Collection			The Beck Collection of Illuminated Manuscripts (LN7362)	6	220275			
10	1		19700704	Sotheby's	Phillips Thomas			Phillips Thomas	4	70			

Capture d'écran d'Open Refine

Une fois le nettoyage et la réorganisation des données effectués, il est possible de lier ces données à des données équivalentes. *OpenRefine* utilise alors des outils complémentaires, soit des logiciels de *Named-Entity Recognition* (NER) qui permettent d'associer des chaînes de caractères à des URI's existants¹⁵. *AlchemyAPI*, *DBpedia Spotlight* et *Zemanta*, trois NER, ont été adaptés par le *Multimedia Lab* de l'Université de Ghent et *iMinds* en collaboration avec le groupe *MaSTIC* de l'Université Libre de Bruxelles afin d'être compatibles avec *OpenRefine*. Le chercheur doit néanmoins confirmer la pertinence des liens proposés, car les noms très courants ou les variantes orthographiques peuvent générer des liens erronés.

	PERSONNAGE_	nom complet	AlchemyAPI
21.	7222	Hector Authier	Hector Authier Choose new match

Capture d'écran d'une correspondance trouvée par *AlchemyAPI*

La dernière étape consiste à convertir nos données en format RDF. *Karma Tool* développé par Craig Knoblock et Pedro Szekely de l'Université de Southern California permet de transformer des tableurs (*Excel*, *GoogleSheets*, *OpenOffice Calc*) en réseau de liens suivant les

¹⁵ David Nadeau et Satoshi Sekine, « A survey of named entity recognition and classification », *Named Entities: Recognition, classification and use*, 2007, vol. 30, n° 1, (coll. « *Linguisticae Investigationes* »), p. 3.

principes du RDF. La prise en main du programme est rapide puisque l'interface est simple et compréhensible. Par exemple, la construction des triplets se fait par un système de schéma de connaissances intuitif qui vient s'attacher aux différentes colonnes de notre tableur. De plus, la conversion en fichier *.xml/rdf* se fait automatiquement lors de l'enregistrement. Un des projets phares de l'équipe de recherche associé au programme a été réalisé à partir d'une collection d'œuvres d'art de la *Smithsonian Institution*¹⁶.

The screenshot shows the Karma Tool interface. At the top, a hierarchical diagram shows the structure of a Cultural Heritage Object (CHO). It starts with 'Aggregation1' leading to 'aggregatedCHO', which then leads to 'CulturalHeritageObject1'. Below this, a diagram shows 'creator' linked to 'Person1', which is linked to 'classLink'. The main part of the interface is a table with columns for 'ConstituentUri', 'ObjectNumber', 'Title', 'Date', 'CreditLine', 'Dimensions', 'CreditLineRepro', 'Medium', 'Classification', and 'ClassificationUri'. The table contains three rows of data representing different artworks.

ConstituentUri	ObjectNumber	Title	Date	CreditLine	Dimensions	CreditLineRepro	Medium	Classification	ClassificationUri
icanart.si.edu/linkedata/person/2	1983.99	Cupid and Psyche	1927	Gift of Mrs. Gertrude Aarons	10 3/4 x 15 x 7 1/8 in. (27.2 x 38.0 x 18.1 cm.)		bronze	Sculpture	http://id.americanart.si.edu/linkedata/term/vci
icanart.si.edu/linkedata/person/2	1983.99	Cupid and Psyche	1927	Gift of Mrs. Gertrude Aarons	10 3/4 x 15 x 7 1/8 in. (27.2 x 38.0 x 18.1 cm.)		bronze	Sculpture	http://id.americanart.si.edu/linkedata/term/vci
anart.si.edu/linkedata/person/2692	1984.124.1	Iran, from the United Nations Series	1944	Gift of Container Corporation of America	sheet: 11 x 9 in. (28.0 x 22.7 cm)		gouache and metallic gouache on paper	Painting	http://id.americanart.si.edu/linkedata/term/vci

Capture d'écran de *Karma Tool*

3. Au cœur des composantes du Web sémantique : URI et triplets

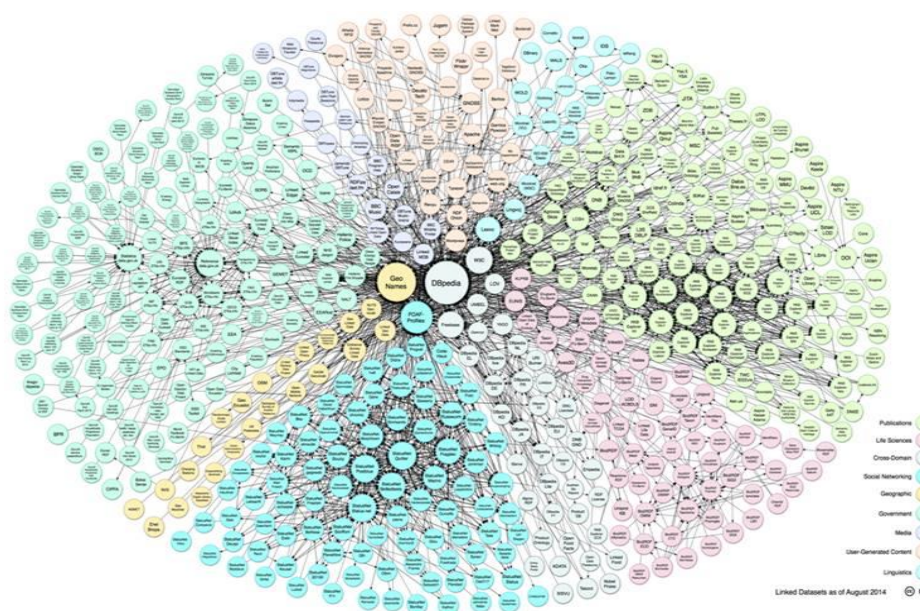
Les URI et les triplets sont au cœur des structures du Web sémantique. Toutes les parties d'un triplet peuvent être associées à un URI. Dans un premier temps, il sera question des URI des sujets et des objets.

Puisque l'objectif principal du Web sémantique est de créer de la cohésion et du sens par des liaisons entre les données, l'architecte de l'information historique se doit de créer des ponts entre sa base de données et d'autres acteurs du nuage des données liées. Un des problèmes majeurs du Web sémantique est la multiplication des URI pour une même entité. Évidemment, il est plus simple de créer son propre URI pour Jacques Cartier que de chercher ceux qui existent

¹⁶ Craig Knoblock et Pedro Szekely, *Karma: A Data Integration Tool*, <http://usc-isi-i2.github.io/karma/>, 2015, consulté le 6 octobre 2015.

déjà. Toutefois, il est préférable de s'appuyer sur des URI existants, même si cette pratique est laborieuse, car elle permet d'atteindre les cinq étoiles. De plus, cette pratique inscrit directement vos données dans le nuage partagé et les rend disponibles pour des requêtes fédérées. Il existe plusieurs acteurs qu'un historien se doit de consulter pour associer ses données à des URI existants.

La plus importante base de données sémantique est *DBpedia*, une version des articles de *Wikipedia* sous forme de triplets¹⁷. Une large communauté travaille à l'automatisation de la transformation de l'information relationnelle vers des triplets. *DBpedia* est actuellement l'acteur central du Web sémantique selon la représentation de Richard Cyganiak¹⁸, puisqu'il s'agit du plus large ensemble de données et de la base qui possède le plus de liens entrants.



Nuage de données liées de Richard Cyganiak

Source : Cyganiak et Jentzsch, *The Linking Open Data cloud diagram*, <http://lod-cloud.net/>

Rapidement, on peut trouver l'URI *DBpedia* de Jacques Cartier grâce à une simple recherche classique¹⁹. *DBpedia* étant un joueur central, il est essentiel de faire référence à ce dernier, mais il ne faut pas négliger d'autres joueurs plus près de la sensibilité historique. C'est le cas de la *Library of Congress* dont il a été question précédemment. À titre de bibliothèque nationale des

¹⁷ DBpedia, *About / DBpedia*, <http://dbpedia.org/about>, 2014, consulté le 28 août 2015.

¹⁸ Richard Cyganiak et Anja Jentzsch, *loc. cit.*

¹⁹ DBpedia, *About: Jacques Cartier*, http://dbpedia.org/page/Jacques_Cartier, consulté le 28 août 2015.

États-Unis, l'institution maintient à jour son catalogue de vocabulaires et d'autorités, comme le LCSH par exemple, depuis 1898²⁰.

Si un chercheur remarque une erreur dans la formation d'un triplet, il est recommandé d'en informer la personne ou l'organisation responsable afin que la correction soit apportée. Si la demande reste sans réponse, le chercheur peut choisir un URI d'un autre partenaire ou créer son propre URI. Dans les deux cas, on devrait aussi créer un lien vers l'URI erroné en l'accompagnant d'une mention afin d'avertir d'autres chercheurs. Un système robuste de validation des URI devra d'ailleurs être développé pour que le Web sémantique puisse s'imposer plus largement.

La partie centrale du triplet, le prédicat, permet d'associer le sujet et l'objet. Tout comme les deux autres parties, le prédicat doit être normalisé pour créer des ponts entre les différentes bases de données. La hiérarchisation du vocabulaire et la définition de liens explicites entre les concepts forment un schéma des connaissances que l'on qualifie d'ontologie. Les prédicats servent ainsi à décrire la nature d'une entité (personnage, événement, concept) et à relier deux entités ensemble. Par exemple, si l'on s'intéresse aux relations de Jacques Cartier, il serait possible de placer sous sa propre personne Jacques Noël et Thomas Fromont²¹. Par contre, les relations entre Jacques Cartier et ces deux personnes sont différentes. Le premier est le neveu de l'explorateur tandis que le second est l'assistant de celui-ci sur la Grande Hermine. Une ontologie permet précisément d'effectuer la distinction entre des liens familiaux et des liens professionnels. Jacques Cartier serait associé à Jacques Noël par le triplet *Jacques Cartier – oncle de – Jacques Noël* et à Thomas Fromont par *Jacques Cartier – supérieur de – Thomas Fromont*.

L'ontologie *Friend of a Friend* (FoaF), développée par Dan Brickley et Libby Miller ainsi qu'une grande communauté de membres, permet de décrire des personnes et de créer un réseau entre celles-ci. La construction du réseau se réalise grâce au prédicat « connaît » (en anglais : *knows*). FoaF est aussi intégré dans l'ontologie BIO, un « vocabulary for describing biographical

²⁰ Library of Congress, *Library of Congress Subject Headings - LC Linked Data Service*, <http://id.loc.gov/authorities/subjects.html>, consulté le 28 août 2015.

²¹ Université Laval et University of Toronto, *Biographie – CARTIER, JACQUES (1491-1557)*, http://www.biographi.ca/fr/bio/cartier_jacques_1491_1557_1E.html, 2015, consulté le 28 août 2015.

information about people, both living and dead.²² ». Bien que BIO soit présentement moins utilisé que FoaF, le premier est plus performant en histoire que le second puisqu'on peut structurer la vie et les relations d'une personne sous forme d'événements. Par exemple, la relation entre Jacques Cartier et Catherine Des Granches²³ pourrait être représentée comme un intervalle entre le mariage et la mort de l'explorateur. Les triplets permettant de décrire cette situation seraient:

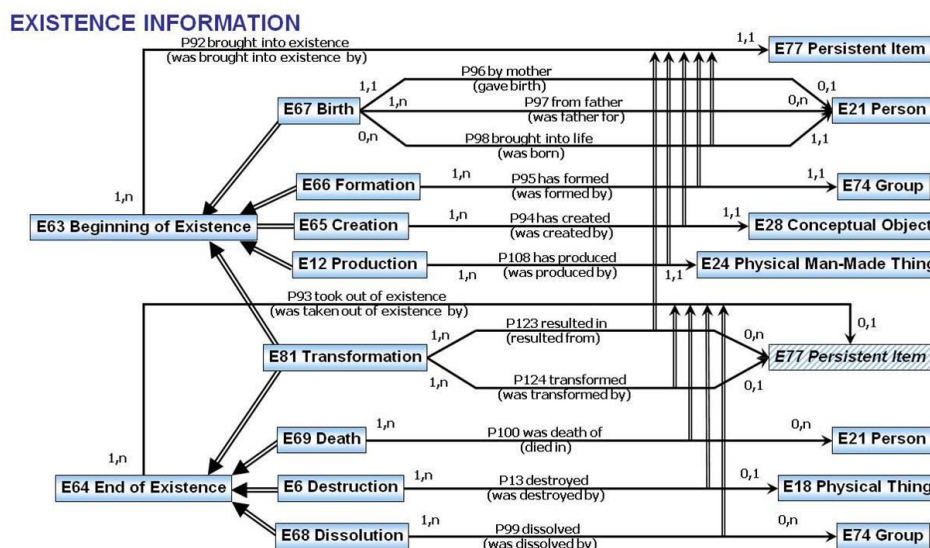
- *Jacques Cartier – est un participant de la – Relation mari et femme de Jacques Cartier et Catherine Des Granches*
- *Catherine Des Granches – est une participante de la – Relation mari et femme de Jacques Cartier et Catherine Des Granches*
- *Relation mari et femme de Jacques Cartier et Catherine Des Granches – est une intervalle de temps nommée – Période de vie conjugale entre Jacques Cartier et Catherine Des Granches*
- *Période de vie conjugale entre Jacques Cartier et Catherine Des Granches – débute avec le – Mariage de Jacques Cartier et Catherine Des Granches*
- *Période de vie conjugale entre Jacques Cartier et Catherine Des Granches – se termine avec – le décès de Jacques Cartier²⁴.*

Le concept d'événement est très prometteur pour le travail historique, mais il faut aussi aller au-delà de la biographie afin d'intégrer d'autres types de sources et d'informations dont les archives et les objets muséaux. Le domaine muséologique fournit un vocabulaire favorisant le plus l'interdisciplinarité entre les bibliothèques, les centres d'archives et les musées. L'*International Council of Museums (ICOM)*, par sa branche documentaliste nommée le *Comité international pour la documentation (CIDOC)* a défini l'imposant *Conceptual Reference Model (CIDOC CRM)*.

²² Ian Davis et David Galbraith, *BIO: A vocabulary for biographical information*, <http://vocab.org/bio/0.1/.html>, 14 juin 2011, consulté le 28 août 2015.

²³ Louis-Marie Le Jeune, « Jacques Cartier » dans *Dictionnaire Général de biographie, histoire, littérature, agriculture, commerce, industrie et des arts, sciences, mœurs, coutumes, institutions politiques et religieuses du Canada*, Ottawa, Université d'Ottawa, 1931, vol.1, p. 313-314, <http://faculty.marianopolis.edu/c.belanger/QuebecHistory/encyclopedia/JacquesCartier-Naissancejeunesseetmariage-HistoiredelaNouvelle-France.htm>, consulté le 28 août 2015.

²⁴ Pour approfondir le mécanisme de BIO: Ian Davis et David Galbraith, *loc. cit.*

Schéma des informations sur l'existence de CIDOC CRM²⁵

Cette ontologie semble la plus prometteuse pour les sciences historiques puisqu'elle associe tous types de connaissances liées à l'histoire et au patrimoine par la construction d'événements. L'historien souhaitant utiliser une ontologie pertinente doit nécessairement connaître les recommandations de CIDOC CRM. Dans la seconde partie de cet article, nous présenterons une étude de cas d'utilisation de CIDOC CRM afin de démontrer comment on peut y intégrer des données historiques et patrimoniales variées, le genre de recherches qui sont nécessaires, le type de liens qui peuvent être créés, et les limites actuelles d'une telle démarche.

²⁵ International Council of Museums, *The CIDOC CRM*, <http://cidoc-crm.org/>, décembre 2014, consulté le 26 mars 2015.

BIBLIOGRAPHIE

- BERNERS-LEE, Tim, HENDLER, James et LASSILA, Ora. « The Semantic Web », *Scientific American*, 17 mai 2001, <http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>, consulté le 27 août 2015.
- CYGANIAK, Richard et JENTZSCH, Anja. *The Linking Open Data cloud diagram*, <http://lod-cloud.net/>, 2014, consulté le 31 octobre 2014.
- DAVIS, Ian et GALBRAITH, David. *BIO: A vocabulary for biographical information*, <http://vocab.org/bio/0.1/html>, 14 juin 2011, consulté le 28 août 2015.
- DBPEDIA. *About / DBpedia*, <http://dbpedia.org/about>, 2014, consulté le 28 août 2015.
- DBPEDIA. *About: Jacques Cartier*, http://dbpedia.org/page/Jacques_Cartier, consulté le 28 août 2015.
- GOUVERNEMENT DU CANADA. *Cote de degré d'ouverture des données*, <http://ouvert.canada.ca/fr/cote-degre-douvertrure-des-donnees>, 25 avril 2013, consulté le 8 janvier 2016.
- GOUVERNEMENT DU CANADA et STATISTIQUE CANADA. *Entente de licence ouverte de Statistique Canada - Foire aux questions (FAQ)*, <http://www.statcan.gc.ca/fra/reference/licence-faq-fra#a1>, 25 janvier 2013, consulté le 8 janvier 2016.
- INTERNATIONAL COUNCIL OF MUSEUMS. *The CIDOC CRM*, <http://cidoc-crm.org/>, décembre 2014, consulté le 26 mars 2015.
- KNOBLOCK, Craig et SZEKELY, Pedro. *Karma: A Data Integration Tool*, <http://usc-isi-i2.github.io/karma/>, 2015, consulté le 6 octobre 2015.
- LE JEUNE, Louis-Marie. « Jacques Cartier » dans *Dictionnaire Général de biographie, histoire, littérature, agriculture, commerce, industrie et des arts, sciences, mœurs, coutumes, institutions politiques et religieuses du Canada*, Ottawa, Université d'Ottawa, 1931, vol.1, p. 313-314, <http://faculty.marianopolis.edu/c.belanger/QuebecHistory/encyclopedia/JacquesCartier-Naissancejeunesseetmariage-HistoiredelaNouvelle-France.htm>, consulté le 28 août 2015.
- LIBRARY OF CONGRESS. *Cartier, Jacques, 1491-1557 - LC Linked Data Service*, <http://id.loc.gov/authorities/names/n50080987.html>, consulté le 27 août 2015.
- LIBRARY OF CONGRESS. *Library of Congress Subject Headings - LC Linked Data Service*, <http://id.loc.gov/authorities/subjects.html>, consulté le 28 août 2015.
- MINISTÈRE DE LA CULTURE ET DES COMMUNICATIONS DU QUÉBEC. *Cartier, Jacques*, <http://www.patrimoine-culturel.gouv.qc.ca/rpcq/detail.do?methode=consulter&id=17323&type=pge#.Vd-sQK01uzx>, 2013, consulté le 28 août 2015.
- MORRIS, Tom. *OpenRefine*, <http://openrefine.org/>, 2015, consulté le 5 novembre 2015.

NADEAU, David et SEKINE, Satoshi. « A survey of named entity recognition and classification », *Named Entities: Recognition, classification and use*, 2007, vol. 30, n° 1, (coll. « *Lingvisticæ Investigationes* »), p. 3-26.

ONLINE COMPUTER LIBRARY CENTER. *Library, Archive and Museum Collaboration*, <http://www.oclc.org/research/activities/lamsurvey.html>, 30 novembre 2011, consulté le 27 août 2015.

UNIVERSITE LAVAL et UNIVERSITY OF TORONTO. *Biographie – CARTIER, JACQUES (1491-1557)*, http://www.biographi.ca/fr/bio/cartier_jacques_1491_1557_1E.html, 2015, consulté le 28 août 2015.

VERBORGH, Ruben. *Federated SPARQL queries in your browser*, <http://ruben.verborgh.org/blog/2015/06/09/federated-sparql-queries-in-your-browser/>, 9 juin 2015, consulté le 28 août 2015.

WORLD WIDE WEB CONSORTIUM. *World Wide Web Consortium (W3C)*, <http://www.w3.org/>, 2016, consulté le 8 janvier 2016.

WORLD WIDE WEB FOUNDATION. *History of the Web*, <http://webfoundation.org/about/vision/history-of-the-web>, 2012, consulté le 8 janvier 2016.

ZENG, Marcia Lei. *Metadata*, New York, Neal-Schuman Publishers, 2008, 365 p.