

Du cycle de vie des données au cycle de vie des objets

Alexandre Monnin

► **To cite this version:**

Alexandre Monnin. Du cycle de vie des données au cycle de vie des objets. Lisette Calderan and Pascale Laurent and Hélène Lowinger and Jacques Millet. Le document numérique à l'heure du web, ADBS, pp.221-228, 2012, Sciences et techniques de l'information, 978-2-84365-142-7. <hal-00843825>

HAL Id: hal-00843825

<https://hal.inria.fr/hal-00843825>

Submitted on 12 Jul 2013

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.

Du cycle de vie des données au cycle de vie des objets

Alexandre Monnin

Responsable Recherche web et métadonnées à l'Institut de recherche et d'innovation (IRI) du Centre Pompidou où il dirige plusieurs projets innovants et séminaires de recherche, Alexandre Monnin achève une thèse de philosophie à l'université Paris 1 (équipe Phico, composante EXeCO) consacrée à la philosophie du web. Discipline autour de laquelle il organise chaque année depuis 2010 les événements scientifiques « PhiloWeb », et à laquelle un numéro spécial de la revue *Metaphilosophy* a été récemment consacré. Il est collaborateur extérieur de l'INRIA (membre associé de l'équipe Wimmics) sur le projet *SemanticPedia* (incluant le *DBpedia francophone*) qu'il a contribué à lancer, également associé au CNAM (laboratoire DICEN), et chair et co-fondateur du « *Philosophy of the Web* » Community Group du W3C. <http://web-and-philosophy.org>, aamonnz@gmail.com

En nous appuyant sur notre expérience au sein du projet *DBpedia francophone*¹, nous entendons examiner la manière dont s'opère la transition entre un wiki ouvert à la contribution, tel que Wikipédia, et une base de connaissances faisant office de référentiel de fait pour le web de données, telle que *DBpedia*. À la lumière de l'architecture du web, quelle interprétation donner du couplage Wikipédia/*DBpedia* ? Encyclopédie (texte) et notice (données), ou plateforme de tri des objets ouverte à la controverse et rassemblement d'entités composant un monde commun ? Nous entendons ici déplacer le débat d'une perspective traditionnellement documentaire vers une perspective politique et philosophique ; en un mot comme en cent, *ontologique*.

1 Wikipédia et la pluralité des régimes du document : entre sources tertiaires, sources secondaires et sources primaires

Bâtir une encyclopédie telle que Wikipédia aura nécessité l'établissement de procédures nombreuses et parfois complexes. Pour quiconque souhaite percer le sens d'une telle entreprise, il est utile de se pencher sur les rares lieux où sont explicitées ces procédures. On entraperçoit alors l'effort de théorisation qui se lit derrière les (bonnes) pratiques défendues par les Wikipédiens.

L'opposition entre « sources secondaires » et « sources primaires » doit ici nous servir de premier fil conducteur [17]. Le réquisit de neutralité, constamment démenti à la lecture des articles de Wikipédia, apparaîtrait comme une fable contredite par les faits s'il ne s'ancrait dans cette distinction entre deux types de sources. Les sources secondaires, pour commencer, fournissent aux contributeurs leur principal matériau. Par sources secondaires, entendons tout document dont le témoignage est censé éclairer une situation : article de presse, article scientifique, monographie, etc. À l'inverse, la source primaire a perdu cette capacité à rendre compte d'autre chose que d'elle-même : elle est devenue, à son corps défendant, *son propre sujet*.

Pour illustrer cette dichotomie, l'ouvrage désormais ancien de l'historien Fustel de Coulanges, *La Cité Antique*, composé en 1864, nous servira d'exemple². Autrefois source secondaire, témoignage scientifique au sujet de la cité antique, il ne put échapper – comme toute chose – à l'érosion du temps et acheva sa carrière universitaire de livre d'histoire, ravalé au rang de document historique, témoignage (dé)passé d'une Cité Antique reconstruite par un historien du XIX^e siècle, jugé incapable, pour un regard contemporain intraitable, de s'abstraire de son milieu « culturel ». Témoignage à situer dans son contexte – *donc* hors du champ scientifique !

À cette division entre deux régimes des sources correspond une autre division, entre deux régimes documentaires. *L'hétérothéticité*, où le document puise sa valeur de sa capacité à enregistrer un témoignage quelconque, en particulier les protocoles et les expériences relatées dans le cas des articles scientifiques, ce que Bruno Latour nomme des « associations d'humains et de non-humains³ », et *l'autothéticité*, qui consacre la perte d'une telle capacité. À lui seul, le temps qui passe n'explique cependant pas tout. Si les historiens avaient cessé de mener des recherches au sujet de la cité antique (sans majuscules), jamais les historiens des idées ne se seraient emparés de *La Cité Antique* (avec majuscules). Le renouvellement des connaissances, et non le passage du temps, expliquent la transition d'un régime à l'autre, basculant invariablement du secondaire vers le primaire, de l'hétérothéticité vers l'autothéticité.

¹ <http://lab.wikimedia.fr>

² Cet exemple est tiré de [17].

³ Sur la notion de « chaînes de traductions », voir [1].

Cette conception des sources en constant renouvellement est en grande partie gagée sur les usages de la publication scientifique dans des domaines tels que les sciences de la vie ou de la matière. Il n'y rien d'étonnant à constater à quel point la conceptualisation des sources sur Wikipédia semble emprunter aux sciences « dures » : le profil de ses contributeurs n'y est sans doute pas étranger. Toutefois, si la neutralité factuelle est indexée sur le régime documentaire à double entrée qui la soutient, alors se pose bel et bien un problème de taille. Comment, en effet, les faits eux-mêmes en viennent-ils à perdre de leur acuité au fil du temps (ou, répétons-le, de l'évolution des publications qui présente d'autres « faits » concurrents) ?

La question posée ici est celle de ce que Bruno Latour nomme l'*objectivité de second degré* [16]. Cette expression désigne le régime d'objectivité qui sied aux controverses scientifiques, lorsque l'issue d'un débat demeure fondamentalement incertaine et qu'il est impossible de s'en tenir à une posture de neutralité symbolisée par la hauteur ou le retrait. De surcroît, il n'est pas toujours loisible, ni souhaitable, d'attendre que la controverse parvienne à son terme pour trancher. Du point de vue de Wikipédia, cela poserait d'ailleurs un problème insurmontable car c'est au moment même où prend fin une controverse, lorsqu'une querelle s'assagit, que les positions exprimées initialement, dont on peut *désormais* arguer après coup que les unes étaient objectives quand les autres ne l'étaient pas, échouent sur le rivage des sources primaires. En ce sens, avec sa définition des sources secondaires, Wikipédia est bel et bien condamnée à indexer la factualité sur des sources disputées et toujours sujettes à controverse. Partant, et pour le dire en un mot, à déployer une objectivité de second degré, caractérisant davantage la recherche *in medias res* que la figure traditionnelle d'une Science appelée, par définition, à faire cesser toute discussion.

En tant que « source tertiaire » (c'est-à-dire appuyée sur des sources secondaires), la politique éditoriale de Wikipédia ne conduit d'ailleurs pas à opérer *a priori* un lissage ou une réconciliation asymptotique des points de vue. Ce rôle ne pourrait d'ailleurs échoir qu'aux seuls administrateurs⁴, les contributeurs ayant tous un point de vue à faire valoir, qu'ils n'abandonnent certainement pas sur le seuil avant de pénétrer sur l'agora. Ils ne contribuent, il est vrai, qu'à la condition expresse de solliciter des « porte-parole », pour reprendre l'expression désormais passée à la postérité de Michel Callon [3]. Qu'est-ce après tout que la citation de sources secondaires si ce n'est la mobilisation de porte-parole pour affirmer une position ?

Étrange mais féconde équation qui voit une double absence de neutralité (des contributeurs et des sources) passer pour un sommet de neutralité (sans doute la subjectivité des contributeurs apparaît-elle quelque part tempérée par le recours obligé à des *publications*) ! Accordons-nous plutôt sur l'expression tantôt mentionnée d'objectivité de second degré pour dissiper ces apparents paradoxes.

2 Architecture du web et web sémantique : au-delà du document, repenser l'objet

L'architecture du web, on le sait désormais, met en avant la notion abstraite de ressource au cœur du système. La ressource peut être « n'importe quoi », définition fixée sous la plume de Roy Fielding dans sa thèse, où il définissait le style d'architecture REST⁵. Par son contenu, elle fait écho à la définition de l'objet en philosophie, le « quelque chose en général ». À ceci près que les objets du web tirent leurs caractéristiques de leur place au sein du réseau, de ce que l'on dit à leur sujet, des liens dans lesquels ils sont enserrés : leur nature, par conséquent, est foncièrement *relationnelle* et *discutable*. Il est impossible pour de tels objets de dissocier *ce que l'on dit ou connaît d'eux* (les propriétés épistémiques) de *ce qu'ils sont* (propriétés ontologiques)⁶. Ce n'est guère étonnant si l'on considère que le web est un milieu associé parcouru de relations en tous genres, nombre d'entre elles étant matérialisées sous la forme de liens hypertextes ou de triplets RDF.

Le web de données s'appuie sur cette architecture pour proposer des langages de représentation permettant de décrire ces objets. À cet égard, il est bon de rappeler qu'un projet de recherche européen baptisé OKKAM⁷ (nom inspiré de Guillaume d'Ockham, auquel il emprunte le célèbre rasoir), visait à mettre en place un « web d'entité » (*Web of Entities*). Ce qui constitue, en un sens, une bien meilleure appellation que l'expression « web de données » – les données étant du côté des représentations (au sens du protocole HTTP), qui illustrent les ressources, et non des ressources elles-mêmes, qui se situent à un niveau plus fondamental⁸.

Avec le web et le web de données, c'est bien le type et la nature de ces entités qui est désormais au cœur des préoccupations. Dans cette perspective, la sémantisation d'une encyclopédie telle que Wikipédia qui, via ses entrées, recense les composants de notre monde, constitue un enjeu de premier ordre. Là où nous avons affaire à des entrées textuelles, nous avons désormais affaire, avec DBpedia, aux entités mêmes que visent ces entrées (en ceci qu'elles sont bien hétérothétiques, non réductibles à leur statut de document, encore moins de support). Du point de vue des machines, applications et autres services, c'est très clair car elles sont nombreuses à tirer de DBpedia la liste des entités qu'elles manipulent.

⁴ Qui s'en dispensent explicitement en refusant d'intervenir directement sur le contenu.

⁵ Voir le chapitre 10. Voir également [15].

⁶ Voir en particulier [10] et [12].

⁷ <http://www.okkam.org>

⁸ Sur ces différents niveaux, voir [11] et [13].

Nul retour à l'évidence de la chose du sens commun cependant. Cette filiation fait dépendre ces entités de la contribution des wikipédiens partout sur la planète : l'objectivité sur laquelle repose nos objets est bien une objectivité de second degré. La contribution des utilisateurs de Wikipédia, sans cesse soumise à la contradiction et aux ajustements des pairs, nous ramène ainsi à la définition de la *res* ou de la *substantia* posée par les romains, comme l'illustre cette très belle citation de Jean-François Courtine : « Le terme de *substantia* s'impose, selon l'esprit même de la langue, pour nommer ce que vise la question conjecturale *an sit* ? Qu'en est-il de la réalité du fait en question, peut-on faire fond sur lui, y a-t-il bien là une *res certa et solida* ? [...] Ce qu'Aristote, après Platon, nommait en terme de présence et de présentification est devenu sujet de débat, point litigieux, fait à établir. [...] La rhétorique latine s'assigne expressément pour objet de persuader (*facere fidem*⁹), de "stabiliser", ce qui est d'abord donné comme *res dubia*. [4] »

D'où l'importance accordée à la confiance. Le web de données ne s'appuyant en aucun cas sur la notion de vérité. Le graphe des objets mis en relation n'est pas établi en vertu de leur adéquation première avec « les objets du monde réel ». Le web n'est pas le lieu où se *représentent* des objets, il est le lieu où se *constituent* des objets. Objets instables, polémiques, « chevelus » et « non modernes », dirait Bruno Latour. Il n'y a plus d'un côté les représentations et de l'autre la vérité/réalité objective. Le concept de source secondaire, on l'a vu précédemment, ne s'accommode guère de ce lit de Procuste. Pas plus, d'ailleurs, que celui de ressources.

3 Wikipédia/DBpedia : encyclopédie (documents) et notices (données) ou forum (avec discussions, procédures et controverses) où s'élabore un monde commun ?

Jean-Michel Salaün établit une comparaison, dans le premier chapitre de cet ouvrage, entre les entrées de DBpedia et les notices documentaires dont, dans le chapitre 2, Gautier Poupeau retrace l'évolution à la lumière des transformations induites par le numérique, et plus spécifiquement par le web. Rappelons la définition de G. Poupeau : « Nous utiliserons ici [le] terme [de notice] dans un sens plus générique, pour désigner tout outil ayant pour fonction de décrire un objet dans une collection suivant des règles définies par le contexte métier dans lequel se déroule l'activité de signalement : notice de bibliothèque, inventaire d'archive, fiche descriptive d'une œuvre dans un musée, etc. » [...] « Cette fiche prend son sens dans un contexte spécifique : celui de l'institution qui détient une collection. »

Une telle définition rappelle avantagement la place qui échoit au contexte dans la gestion, la production et l'interprétation des notices. Dans le cas d'espèce qui nous intéresse, quelle serait l'institution chargée de donner du sens aux notices de DBpedia en les constituant en collections ? Qu'il s'agisse des objets de l'architecture du web (plateforme ouverte et décentralisée qui n'est pas liée *a priori* à une institution donnée) ou du couple Wikipédia/DBpedia, dans tous les cas l'institution au sens étroit du terme fait défaut. Wikipédia est-elle seulement une encyclopédie dotée d'un éditeur ?

On sait qu'en interne la question fait débat, les seuls éditeurs reconnus étant les contributeurs. Peut-être faut-il accepter d'aller au-delà d'une conception documentaire traditionnelle pour saisir la portée réelle du travail accompli depuis plus d'une décennie. Avec lui, le sens même de la collection bascule. Encyclopédie ou terrain d'expérimentation hautement politique de constitution d'un monde commun, reposant sur des procédures complexes, la ligne de fracture distingue les questions purement documentaires des questions ontologiques et cosmopolitiques¹⁰ que charrie le web :

- ce dernier permet en effet, nous l'avons vu, d'ouvrir à nouveaux frais la question ontologique (qu'y a-t-il ?) en permettant à tout un chacun d'*identifier* (grâce aux URIs) et d'*individuer* les objets (ou ressources) qu'il souhaite ;
- à cet égard, ce n'est sans doute pas l'effet du hasard si DBpedia constitue la principale application du web de données. Au-delà des processus traditionnels d'édition, le couple Wikipédia/DBpedia a permis l'incubation de nombreuses procédures de *tri* (sur Wikipédia) portant sur les constituants d'un monde partagé (les entités de DBpedia), toujours à négocier, par la mobilisation de sources secondaires.

En ce sens, la controverse n'est nullement un mal nécessaire dans l'attente d'une résolution qui ne doit pas manquer d'intervenir à terme – lorsque nous serons tous morts, comme le disait justement Keynes ! La forme même des articles, l'espace partagé de publication qui est le leur, garantissant une superposition de points de vue pluriels et hétérogènes – à plus forte raison lorsqu'on explore les couches en forme de palimpseste de leur historique. Au contraire, donc, elle est la forge même où se « com-posent » des objets partagés, discutables et discutés, en un mot des objets provisionnels. Un « parlement des choses » ou une « démocratie d'objets », en somme, pour reprendre les belles expressions de Bruno Latour¹¹ et Levi Bryant [2].

⁹ Toute la question de la confiance sur le web, ultime brique du « cake » du web sémantique, prend dès lors tout son sens.

¹⁰ « On reprend ici le sens grec d'arrangement, d'harmonie en même temps que celui plus traditionnel de monde. [8] »

¹¹ L'enjeu, pour Latour, est bien de procéder à une « redistribution du discutable et de l'indiscutable beaucoup plus originale, beaucoup plus démocratique » que par le passé [9]. Il nous apparaît que, du fait de son architecture, le web constitue une plateforme privilégiée où s'organise *de facto* une telle redistribution.

Bien sûr, une telle expérimentation, vieille d'à peine dix ans seulement, demeure largement perfectible. Nombreuses sont les voix à ne pas trouver d'écho, ne serait-ce qu'en l'absence de porte-parole disposant des moyens d'être publiés. D'où l'urgence, selon nous, d'acquérir une vision claire des enjeux techniques et politiques de ces plateformes qui constituent, plus que jamais, l'arrière-plan de nos « formes de vie technologiques [5] ». Celui-ci est en partie documentaire mais en partie seulement.

Références (hors de ce volume)

- [1] M. AKRICH, M. CALLON, B. LATOUR. *Sociologie de la traduction : textes fondateurs*. Presses de l'École des mines, 2006
- [2] L. R. BRYANT. *The Democracy of Objects*. MPublishing, University of Michigan Library, 2011
- [3] M. CALLON. « Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay ». *Power, action and belief: A new sociology of knowledge*, 1986, 32, 196–233
- [4] J.-F. COURTINE. *Les catégories de l'être*. Paris : PUF, 2003
- [5] S. LASH. *Critique of Information*. (First ed.). SAGE Publications Ltd., 2002
- [6] B. LATOUR. « Note sur certains objets chevelus ». *Nouvelle revue d'ethnopsychiatrie*, 1995, n° 27, p. 21-36
- [7] B. LATOUR. « Moderniser ou écologiser. À la recherche de la septième cité ». *Écologie et politique*, 1995, n° 13, p. 5-27
- [8] B. LATOUR. *Politiques de la nature*. La Découverte, 2004
- [9] B. LATOUR. « L'alternative compositionniste. Pour en finir avec l'indiscutable ». *Écologie et politique*, 2010, n° 40(2), p. 81
- [10] P. LIVET. « Web Ontologies as Renewal of Classical Philosophical Ontology ». *Metaphilosophy*, 2012, vol. 43, n° 4, p. 396-404
- [11] A. MONNIN. *La ressource et l'ontologie du Web*. 2011. <http://hal-paris1.archives-ouvertes.fr/hal-00610652>
- [12] A. MONNIN. « L'ingénierie philosophique comme design ontologique : retour sur l'émergence de la "ressource" ». *Réel-Virtuel*, 2012, n° 3
- [13] A. MONNIN. « The artifactualization of reference and "substances" on the Web. Why (HTTP) URIs do not (always refer nor resources hold by themselves) ». *American Philosophical Association Newsletter*, 2012, n° 11
- [14] A. MONNIN. « Entre sciences et contributions, du nouveau régime d'objectivité du Web et de quelques conséquences sur l'éducation ». *Causes toujours*, 2012, (26)
- [15] A. MONNIN., N. DELAFORGE, F. GANDON. « CoReWeb: From linked documentary resources to linked computational resources ». In : A. Monnin, H. Halpin, L. Carr (Eds.). *Proceedings of the WWW2012 conference workshop PhiloWeb 2012: "Web and Philosophy, Why and What For?"* (Vol. 859). CEUR Workshop Proceedings. 2012
- [16] T. VENTURINI, M. JACOMY, D. RICCI. *Blurring the Net: A Method to Visualise Second-Degree Objectivity*. S. d. http://www.medialab.sciences-po.fr/publications/Venturini-Second_Degree_Objectivity_draft1.pdf
- [17] WIKIPEDIA CONTRIBUTORS. « Wikipédia :Sources primaires et secondaires ». *Wikipédia*. Wikimedia Foundation, Inc., 2012, August 12