



HAL
open science

Documentarité et données, instrumentation d'un concept

Arthur Perret, Olivier Le Deuff

► **To cite this version:**

Arthur Perret, Olivier Le Deuff. Documentarité et données, instrumentation d'un concept. 12ème Colloque international d'ISKO-France : Données et mégadonnées ouvertes en SHS : de nouveaux enjeux pour l'état et l'organisation des connaissances?, Oct 2019, Montpellier, France. hal-02307039

HAL Id: hal-02307039

<https://hal.science/hal-02307039v1>

Submitted on 7 Oct 2019

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.

Documentarité et données, instrumentation d'un concept

Arthur Perret

Doctorant en Sciences de l'information et de la communication

MICA, Université Bordeaux Montaigne

arthur.perret@u-bordeaux-montaigne.fr

Olivier Le Deuff

MCF HDR en Sciences de l'information et de communication

MICA, Université Bordeaux Montaigne

oledeuff@gmail.com

Résumé

Cet article constitue une synthèse sur le concept relativement récent de documentarité et propose un exemple de son application aux données. La synthèse montre que la documentarité ne se confond pas avec la notion de document ni avec la théorie de la documentalité. La documentarité correspond à « ce qui fait document ». Elle contribue à la théorie de notre perception des objets informationnels. Elle éclaire notamment la dimension processuelle et interprétative des données, dans le même esprit que les termes d'obtenues (Latour) et de capta (Drucker). L'exemple d'un portail de données bibliographiques permet d'illustrer dans la pratique des composantes de la documentarité telles que les dispositions et les affordances.

Mots clés

Documentation, données, interface

Title

Documentarity and data, instrumentation of a concept

Abstract

This article provides an overview of the relatively recent concept of documentarity along with an example of its application to data. The literature review shows that documentarity is distinct from the notion of document and from the theory of documentality. Documentarity is what acts as a document. It allows us to further theorize our perception of information. It shows in particular the procedural and interpretative dimensions of data, in the spirit of alternative terms such as Latour's *sublata* and Drucker's *capta*. The example of a bibliographic data portal puts in practice components of documentarity such as dispositions and affordances.

Keywords

Documentation, data, interface

INTRODUCTION

L'essor du Web avait occasionné en France une vaste réflexion sur l'articulation entre ancien et nouveau régime documentaire, notamment autour de la notion de « redocumentarisation » (Pédauque, 2007) et de ses critiques (Courbières, 2008). Une nouvelle période de questionnements s'est ouverte depuis, polarisée par la notion de données, laquelle interroge comme précédemment les fondamentaux de la gestion de connaissances, ainsi que ceux des sciences de l'information et de la communication (SIC) de façon plus globale. De la même façon que se dessinent les contours d'un imprimé post-numérique, au sens d'un paradigme commun aux deux supports, le document résiste épistémologiquement à une certaine forme de discours techno-prophétique qui aurait voulu lui substituer la donnée sans autre forme de procès. C'est ainsi qu'après avoir longuement interrogé ce que devient le document au 21^e siècle, un nouveau questionnement émerge : qu'est-ce qui est susceptible de « faire document » dans les données et les mégadonnées ? En réponse à cela, nous proposons d'examiner le concept de documentarité, dont les origines sont multidisciplinaires. Il s'agit de situer ce concept vis-à-vis de la recherche sur les propriétés de la notion de document ainsi que sur l'épistémologie des données, puis d'examiner son instrumentation.

1 – DOCUMENTATION ET AGENTIVITÉ

De nombreuses études sur la documentation menées ces dernières années se basent sur une redécouverte de ses pionniers européens, avec notamment les publications essentielles de Paul Otlet (1934), Suzanne Briet (1951) ou encore Robert Pagès (1948). Il s'agit en particulier d'un renouveau théorique important pour la bibliothéconomie anglo-saxonne. Les concepts fondamentaux de la documentation sont réinterrogés à l'aune de problématiques nouvelles, dans une perspective double : éclairer des phénomènes contemporains et en renouveler la théorie. Ces phénomènes découlent en partie du développement de l'informatique. Confrontés à de nouvelles terminologies et typologies d'objets, les chercheurs en documentation consacrent une attention particulière à l'élaboration de cadres conceptuels qui permettent d'en étudier les propriétés.

La question la plus travaillée reste celle du document et de ce qui l'asseoit ontologiquement en tant que concept. Une propriété ayant fait l'objet de travaux importants est la *documentalité*. Bernd Frohmann (2012) définit la *documentalité* comme une capacité à générer des traces. Maurizio Ferraris (2013) a quant à lui proposé une théorie de la *documentalité* définie comme l'inscription des actes sociaux sous forme de documents. Ainsi que le relève Claire Scopsi (2018), les deux approches se rapportent à l'agentivité du document. Toutefois, Frohmann s'inscrit dans une critique de l'anthropocentrisme qui donne à sa définition une plus grande généralité. Il applique à la documentation l'argument de Bruno Latour suivant lequel la division entre les propriétés primaires (intrinsèques) et secondaires (ressenties) d'un objet doit être remise en cause. Frohmann montre notamment qu'une trace ne dépend pas de son enregistrement par l'homme et généralise l'affirmation de Suzanne Briet suivant laquelle une antilope dans un zoo constitue un document :

« *L'argument sous-jacent de l'antilope de Briet est que quelque chose devient un document en vertu de ses arrangements avec d'autres choses, et non d'une forme privilégiée de ces arrangements, telle que sa fonction de preuve* » [trad. libre] (Frohmann, 2012, p.173).

On retrouve dans cette approche une logique hyperdocumentaire qui ne procède pas de l'hypertexte mais de la vision de Paul Otlet sur l'évolution possible de la documentation. Otlet suggère en effet qu'une fusion document-instrument permettrait une réticularisation du réel avec la connaissance. Dans cette perspective, les propriétés intrinsèques des choses et leur

perception par nos sens se fondent en une « Hyper-Documentation » ainsi qu'une « Hyper-Intelligence » contigües (Otlet, 1934, p.429).

Ainsi que le montre Michael Buckland (2017), de telles possibilités avaient été envisagées par Robert Pagès, contemporain et élève de Suzanne Briet dont il a peut-être inspiré l'exemple de l'antilope. Pagès (1948) utilise un gorille pour illustrer la notion d'autodocument. Ron Day (2018) mobilise cette idée pour décrire une auto-documentalité qui s'inscrit selon nous dans le raisonnement de Frohmann. Day rappelle que la documentation traditionnelle envisage peu l'agentivité du document :

« *Contrairement à notre façon de penser, tous les documents, qu'ils soient iconiques ou figuratifs, sont utilisés de manière pragmatique sur le mode de l'index [...] La documentalité est prescriptive, la documentation est descriptive* » [trad. libre] (Day, 2018, p.8).

La documentalité, entendue comme « force documentaire », peut donc être formulée ainsi : il s'agit de *ce que fait* un objet du point de vue documentaire, le mot document ayant alors une acception plus large que dans le sens traditionnel de document imprimé ou numérique.

2 – CE QUI FAIT DOCUMENT

Nous souhaitons discuter ici une autre propriété fondamentale, dans une formulation proche et pourtant radicalement différente : *ce qui fait* document, et que nous appellerons ici documentarité. La documentarité est ce qui fait document *aux yeux de* — elle explicite le caractère central de la dimension interprétative dans ce qui constitue, lors sa réception, un document. Il s'agit d'une certaine façon de réintroduire la question de la subjectivité, tout en prenant en compte celle de l'agentivité. Notre démarche rejoint celle exprimée par Ron Day, qui a consacré un ouvrage à la documentarité comme philosophie de l'évidence : « *Au lieu d'aborder l'information sous un angle ontologique (« qu'est-ce que l'information ? »), ce livre aborde l'ontologie sous l'angle informationnel, en examinant comment quelque chose devient évident et peut être pris comme preuve de ce qui est* » [trad. libre] (Day, 2019, p.1).

Selon lui, on peut distinguer historiquement une documentarité « forte » basée sur des catégories de jugement a priori et une documentarité « faible » basée sur l'inscription socio-culturelle. Day emprunte à l'ontologie une autre distinction qui vient recouper et approfondir la première : celle entre les dispositions (c'est-à-dire les pouvoirs d'expressions intrinsèques) et les affordances (soit l'influence de l'environnement). L'ensemble sous-tend ce qu'il nomme des « technologies du jugement ».

En SIC, un « principe de documentarité » a été proposé par Stéphane Crozat (2016, 2019) : « *La documentarité est une mesure de ce que permet un contenu en terme de contrat de lecture du point de vue de ses propriétés documentaires* ».

Partant du constat que l'information est nécessairement incarnée par des objets (il n'y a pas d'information sans support), Crozat soulève le problème suivant : dans un environnement numérique, la documentarité de ces objets — c'est-à-dire la façon dont ils font sens — n'est pas évidente. Plus précisément, leur documentarisation n'est pas explicite ou bien se déplace au niveau de la réception par le lecteur.

L'utilisation du mot documentarité dans un contexte théorique a une existence antérieure et extérieure aux SIC, dans un texte d'André Gaudreault et Philippe Marion (1994) sur le documentaire en tant que genre cinématographique. Les auteurs y questionnent ce qui peut « faire document » (p.13) ; selon eux, le récit documentaire s'oppose au récit de fiction, et c'est à sa lecture (à son visionnage) que se jauge sa « documentarité » (p.17). Les termes clés de notre problématique sont donc bien présents mais l'intérêt du texte dépasse une simple

proximité de vocabulaire. En effet, les auteurs mobilisent ces termes pour faire une analyse de la vérité comme rapport au réel qui fait écho aux problématiques que nous avons soulevées. Selon eux, la documentarisation n'est pas qu'un simple mécanisme d'enregistrement, puisqu'elle produit un « effet de trace [...] empreinte non pas du réel mais, plutôt, du sujet graphiateur, sujet-énonciateur » (p.17). Elle constitue donc un processus d'éditorialisation par lequel on produit une vérité qui tend vers le réel. La documentarisation diffère ainsi profondément de la documentalité. La documentarité, quant à elle, est déterminée par la réception : « *la vérité documentaire se pose en contrepoint du vraisemblable fictionnel [...] En invitant le spectateur à une sorte de confiance référentielle, l'effet documentaire ainsi produit constitue un puissant vecteur de vraisemblable* » (Gaudreault, Marion, 1994, p.15).

Dans leur acception, la documentarité tient à la fois de la propriété et de la qualité — c'est l'*Eigenschaft* allemande. Elle n'est pas liée à un acte d'interprétation générique : elle occasionne un jugement de valeur. Les auteurs évoquent un *degré de documentarité*, c'est-à-dire « un nombre plus ou moins grand d'indices du réel » (p. 19) ; la documentarité représente donc une qualité documentaire quantifiable, mesurable.

Cette approche a été appliquée au domaine du Web par Sophie Beauparlant (2017). Analysant la documentarité du webdocumentaire dans les termes de Gaudreault et Marion, elle montre que l'interface met en œuvre un jeu d'écriture du réel qui s'articule au contenu lui-même. Elle illustre la pertinence de leur proposition théorique en lui soumettant un cadre techno-sémiotique concret, confirmant l'intuition que les interfaces numériques contribuent à produire une compétence de lecture.

Ni la documentalité ni la documentarité ne correspondent à une reformulation de ce qu'est un document. Tout d'abord, la notion de document est fonctionnelle et non substantielle, ainsi que l'indiquent Otlet et Briet :

« *Le livre est le moyen d'enregistrement intégral de la pensée en vue de sa transmission [...] comme instrument intellectuel le livre sert non seulement à énoncer des théories, mais à les construire ; non seulement à traduire la pensée, mais à la former* » (Otlet, 1934, p. 426).

« *Un document est une preuve à l'appui d'un fait. [...] Tout indice concret ou symbolique, conservé ou enregistré, aux fins de représenter, de reconstituer ou de prouver un phénomène ou physique ou intellectuel* » (Briet, 1951, p.7).

La plupart des théoriciens ultérieurs ont reconduit cette idée :

« *Un objet qui supporte de l'information, qui sert à la communiquer et qui est durable [...] un objet faisant fonction de mémoire pour une instance réceptrice* » (Meyriat, 1981).

« *L'évolution de la notion de « document » chez Otlet, Briet, Schürmeyer et les autres documentalistes a insisté de plus en plus sur tout ce qui fonctionne comme un document plutôt que sous une forme matérielle traditionnelle. Le passage au numérique semble rendre cette distinction encore plus importante* » (Buckland, 1997, p.808).

« *Un document est un objet intentionnel. Cela implique qu'il soit considéré pour ce qu'il signifie, et non pour ce qu'il est physiquement* » (Bachimont, 2004, p.192).

Dans ce cadre, ce sont essentiellement les propriétés caractérisant le fonctionnement du document qui font l'objet de nouvelles recherches. Or une propriété ne peut à elle seule résumer l'ensemble des propriétés, ou alors elle constitue une nouvelle théorie. On notera que la documentalité selon Frohmann possède 4 aspects — « fonctionnalité, historicité, complexité, agentivité » — qui recourent en partie la notion de document, ce qui interroge sur leur articulation. Quoiqu'il en soit et pour toutes ces raisons, la documentarité n'a pas

vocation à réaliser en elle-même et par elle-même une vaste synthèse théorique qui reste, à notre sens, encore à venir.

La documentarité n'est pas une propriété exclusive au document traditionnel : elle s'applique à tout objet que l'on analyserait dans une perspective documentaire. C'est pourquoi elle peut s'inscrire dans une démarche philosophique, ainsi que le fait Ron Day, ou plus spécifiquement info-communicationnelle, comme nous allons l'illustrer ici avec la question des données.

3 – ÉPISTÉMOLOGIE DES DONNÉES

La notion de données est plus ancienne que son usage en informatique. Le terme lui-même n'émerge qu'au 17^e siècle, dans une phase du développement de l'instrumentation scientifique qui voit se multiplier les objets d'observation systématique. Bien que non formulée comme telle, c'est l'essor d'une pratique de documentation du travail expérimental qui va faire naître la donnée. La logique en est déjà définie dans l'Encyclopédie de Diderot et D'Alembert :

« Données, adj. pris subst. terme de Mathématique, qui signifie certaines choses ou quantités, qu'on suppose être données ou connues, & dont on se sert pour en trouver d'autres qui sont inconnues, & que l'on cherche. Un problème ou une question renferme en général deux sortes de grandeurs, les données & les cherchées, data & quæsita. »

Une certaine catégorie de *quæsita* procède de la certitude épistémologique, quasiment philosophique, suivant laquelle les données cherchées sont moins inconnues que non encore connues. Le mouvement des sciences sociales prédictives illustre à l'extrême ce positionnement méthodologique. L'autre catégorie correspond aux données inférées, produites par déduction à partir de données existantes. C'est le principe du Web sémantique (*Linked Open Data, LOD*) : permettre d'interroger des données exposées pour faciliter la production de nouvelles connaissances. La structuration et les métadonnées constituent les principaux leviers d'enrichissement de ces données.

En France, la distinction entre document et données constitue un thème de recherche important au tournant du 20^e siècle, que les SIC ont abordé notamment par le concept de redocumentarisation. Ce terme porte à la fois l'idée d'un passage de l'analogique au numérique et d'une atomisation du document dans ses modes de production (Pédaque, 2007). Le débat sur la pertinence de la notion de document est alors axé sur deux aspects. Le premier, toujours d'actualité, concerne le bouleversement des valeurs documentaires traditionnelles, au premier rang desquelles la preuve. Le second aspect du débat a trait à la granularité ou à l'échelle documentaire. Paul Otlet avait théorisé le dépassement du livre sur la base d'une unité informationnelle abstraite, le *biblion*, et d'une unité documentaire concrète, la *fiche* (Robert, 2015), une vision en partie vérifiée par la généralisation de l'informatique bureautique.

En revanche, la science de l'information anglo-saxonne se focalise plutôt sur la triade donnée-information-connaissance que sur le rapport entre données et documents. Une enquête très complète de Chaim Zins sur les approches conceptuelles de cette triade a montré que les définitions données pour *data* intègrent souvent le mot *record* ; celle de Michael Buckland en est un bon exemple : *« The word "data" is commonly used to refer to records or recordings, statistical observations, collections of evidence »* (Zins, 2007). On peut y voir un phénomène de repli de l'ancienne bibliothéconomie (américaine notamment) face à l'émergence d'une « science des données » (*data science*).

Si la promesse du Web sémantique avait incité les chercheurs à se poser de nouveau la question des différents niveaux de documents, l'émergence des mégadonnées (ou *big data*) rend cette question en partie caduque, ou du moins la déplace. Il est acquis que la donnée est

nécessairement plus petite que le document, ce qui facilite une économie de l'information basée sur la raison computationnelle (calcul, recombinaison) avec les outils informatiques classiques développés durant la seconde moitié du 20^e siècle. En revanche, l'agrégation de données sous forme de masses volumineuses et hétérogènes remet ces approches en question : les mégadonnées constituent des éléments infra-informationnels (selon l'expression de Bruno Bachimont) qui excèdent les capacités d'analyse à la fois méthodologiques et technologiques existantes. Le débat sur l'articulation entre document et données se déplace alors des questions de granularité — influencées par l'orientation positiviste des premières théories documentaires — à celles d'architecture. De nouvelles technologies sont développées pour monter en charge sur l'analyse brute, dont l'apprentissage profond. Il est intéressant de noter que la logique sous-jacente à l'analyse ne change pas ou peu : l'esprit humain tend à rechercher des motifs réguliers dans des phénomènes désordonnés, nous développons simplement des techniques nouvelles adaptées à la complexification croissante de nos objets d'étude.

À cette terminologie descriptive — données, cherchées, mégadonnées —, est venue s'ajouter une terminologie plus conceptuelle. Dans les années 1990, les sociologues des sciences ont critiqué l'usage du mot « donnée » pour désigner des objets en réalité construits, arrachés au terrain au prix d'un temps et d'efforts parfois considérables. Le mot « obtenue » est notamment suggéré par Bruno Latour comme une alternative souhaitable :

« La tentation de l'idéalisme vient peut-être du mot même de données qui décrit aussi mal que possible ce sur quoi s'appliquent les capacités cognitives ordinaires des érudits, des savants et des intellectuels. Il faudrait remplacer ce terme par celui, beaucoup plus réaliste, d'obtenues et parler par conséquent de bases d'obtenues, de sublata plutôt que de data »(Latour, 2007, p. 609).

Le mot *sublata* apparaît à diverses reprises dans les travaux de Latour mais il fait une occurrence particulièrement intéressante dans *Pandora's Hope*, où il est introduit suite à une remarque sur le rôle de la représentation graphique :

« In order for the botanical and pedological data to be superposed on the same diagram later, these two bodies of reference must be compatible. One should never speak of "data"—what is given—but rather of sublata, that is, of "achievements" » (Latour, 1999, p. 42).

Les termes ne sont pas anodins. *Achievement* signifie accomplissement et sera traduit dans l'édition française par « obtenues ». Avec ce mot, Latour insiste sur le fait que la connaissance est le fruit d'une construction : pour lui, ce que nous appelons donnée est une information de nature processuelle. Quant à *sublata*, il découle des verbes latins *tollo* et *suffero*, qui signifient tour à tour élever, porter ou supporter. On peut y voir un lien avec la métaphore des nains se tenant sur des épaules de géants, très connue en sciences depuis sa reprise par Newton, et qui illustre la nature cumulative du savoir : nos accomplissements sont tributaires de ceux de nos prédécesseurs. La proposition de Latour sur les données nous semble liée à cette vision.

Plus récemment, Johanna Drucker a également avancé une alternative conceptuelle au mot « data ». Il s'agit des « capta » :

« co-dépendantes, constituées de manière relationnelle entre l'observateur et les phénomènes observés, fondamentalement différentes du concept de la donnée comme phénomène indépendant de l'observateur » [trad. libre] (Drucker, 2011a, par. 50).

Ceci prolonge son travail sur l'interface, formulé dans des termes similaires :

« *Codépendance et contingence, l'expérience performative de la connaissance produite dans la relation entre environnement et sujet, tels sont les termes qui définissent l'interface interprétative* » [trad. libre] (Drucker, 2011b, p. 18).

Comme Latour mais de façon plus systématique, Drucker s'appuie la représentation graphique des résultats statistiques pour suggérer un glissement terminologique et conceptuel :

« *Croire que les données sont intrinsèquement quantitatives — évidentes, neutres sur le plan des valeurs et indépendantes de l'observateur — exclut la possibilité de les concevoir comme qualitatives, co-dépendamment constituées — en d'autres termes, de reconnaître que data sont des captas [...] Je suggère que nous repensions fondamentalement les données comme des captas en termes d'ambiguïté plutôt que de certitude, et que nous trouvions des moyens d'exprimer graphiquement la complexité de l'interprétation* » [trad. libre] (Drucker, 2011a, par. 49-50).

Cette « complexité » dont parle Drucker provient notamment du fait que la représentation graphique hérite elle-même de modes d'interprétation, ainsi que d'expression : elle mobilise une ou plusieurs façons de penser, que nous devons prendre en compte. Les *capta* de Drucker traduisent le caractère cumulatif de la connaissance et la dimension processuelle de l'information scientifique, tout comme les *sublata* de Latour.

4 – LA DOCUMENTARITÉ DES DONNÉES

Nous avons établi que la documentarité est une qualité perceptible pouvant faire l'objet d'un jugement de valeur et que la donnée est une construction qui porte la trace de modes d'interprétation et d'expression. Afin de croiser les réflexions sur la documentarité d'une part et sur les données d'autre part, nous nous penchons sur un exemple concret de mégadonnées ouvertes (intersection de l'*open data* et du *big data*). Le portail Isidore moissonne, enrichit et expose des données bibliographiques issues de la recherche en SHS. Son utilisation nous permet de souligner trois logiques que nous pouvons relier aux éléments théoriques discutés précédemment.

La première logique est celle de la structuration. Elle est fondamentale, en ce que qu'elle conditionne les deux autres. Toute écriture numérique organise sa propre énonciation computationnelle, laquelle fait l'objet d'une textualisation par l'humain. Suite aux travaux de Goyet (2017) et Collomb (2017) sur la notion d'architexte, nous définissons celle-ci comme une technologie intellectuelle qui permet une écriture de l'écriture, mobilisant aussi bien la liste que l'algorithme ou le balisage.

La donnée s'inscrit dans ce schéma. Son encodage est la première caractéristique qui constitue l'objet de notre regard interprétatif. Tout jeu de données est structuré suivant des règles syntaxiques spécifiques. L'API d'Isidore fournit des données en XML et en JSON, deux formats pensés pour le stockage et le transport de l'information numérique. Dans les deux cas, les données sont considérablement enrichies par rapport à leur source, grâce au croisement de plusieurs référentiels. Il en résulte des fichiers texte dont le seul volume peut influencer notre appréciation de leur valeur informationnelle. Toutefois leur présentation diffère grandement. Le XML fait un usage classique du « blanc » (retours à la ligne et indentation) qui lui confère un aspect relativement lisible, avec une seule information par ligne. En revanche, le JSON livré par Isidore est « minifié », c'est-à-dire que le blanc en est retiré à des fins d'optimisation. Il en résulte un fichier qui, théoriquement, contient la même

information, avec des délimiteurs moins lourds que le XML, mais qui est beaucoup moins lisible.

La deuxième logique observée est celle de l'éditorialisation. Si la structuration est déjà une affaire de ligne éditoriale, l'éditorialisation suppose des choix de médiation de la donnée qui dépassent le simple format de stockage et mobilisent des programmes de conversion (telles les feuilles XSLT) ainsi que des feuilles de style (par exemple en CSS). Lorsque des données sont affichées dans un navigateur Internet, celui-ci propose une mise en forme par défaut. Des langages différents (tel XML et JSON) peuvent être traités différemment. Ainsi, Firefox (Mozilla) n'affiche pas directement le XML fourni par l'API d'Isidore mais une version sans balisage, ni retours à la ligne, ni indentation, au détriment de la lisibilité globale du fichier. En revanche, il propose une interface pour le JSON qui en facilite l'usage (moteur de requêtes, affichage des entêtes, copie rapide). Isidore comprend par ailleurs une interface graphique qui change entièrement l'expérience de la donnée par rapport à un usage « brut » via l'API. Notre perception repose alors en grande partie sur les caractéristiques de l'interface.

La troisième et dernière logique est celle de la réutilisation, qui correspond à la dimension combinatoire de la redocumentarisation, également exprimée par l'idée de raison computationnelle. Toute écriture humaine sur ordinateur entraîne des opérations de lecture, calcul et écriture par la machine. La récupération des données est en partie conditionnée par leur exposition et en partie par les compétences d'écriture du réutilisateur. Isidore propose un SPARQL *endpoint*, c'est-à-dire une interface vers le jeu de données structurées suivant le principe du Web sémantique. Il permet une plus grande liberté d'interrogation ainsi que l'automatisation des requêtes : un langage de programmation comme Perl ou Python peut combiner SPARQL et expressions régulières pour extraire certains champs d'un jeu de données, puis les inscrire dans un fichier avec une certaine syntaxe en vue d'autres utilisations, par exemple une analyse statistique.

Cet objectif de réutilisation peut rencontrer plusieurs obstacles. Le plus évident est la piètre qualité de certaines sources, que l'enrichissement ne permet pas de combler, et qui a un impact immédiat sur le traitement (absence d'informations, mauvais nommage de champs). On voit ici une limite majeure des services basés sur le moissonnage, fortement dépendants de la qualité de la structuration en amont. Mais il faut également songer à la manipulation volontaire, beaucoup plus aisée via des ordinateurs qu'aux temps de l'imprimerie. L'informatique a en effet généralisé un certain nombre de compétences éditoriales ; ainsi, comparée à celle de la monnaie, la falsification des statistiques apparaît comme triviale.

Dans l'exemple que nous avons développé, les données fonctionnent essentiellement sur le mode de la documentarité « forte » au sens de Ron Day : c'est le mécanisme traditionnel de la référence, qui repose sur la structuration et les métadonnées. Nous en jugeons la qualité en fonction de l'adéquation à nos besoins d'information ou de réutilisation, en référence à des normes (syntaxe, nomenclature). Ce que nous appelons degré de documentarité de ces données procède de leurs pouvoirs d'expressions intrinsèques. Mais cette composante dispositionnelle est fortement contrainte par les affordances du support, en l'occurrence le navigateur. Notre interaction avec les données définit leur capacité à manifester l'évidence ; leur degré de documentarité dépend alors en partie de notre expérience informationnelle.

Par ailleurs, si on a pu dire que la documentarité en régime numérique n'est pas évidente (Crozat, 2016), c'est parce que l'inscription des pouvoirs d'expression se fait par un jeu d'écriture plutôt difficile d'accès. La culture technique n'est pas partagée par tous les acteurs impliqués dans l'exposition des données. En aval, c'est la même chose : le passage par des jeux d'écriture sophistiqués limite fortement la réutilisation. Interfaces, codes sources et algorithmes constituent un codage complexe qui n'appartient pas encore à nos référentiels partagés, contrairement à des supports tels que le livre. La dimension cumulative et

processuelle de l'information est d'autant plus complexe à interpréter. La part de documentarité « faible » rentre alors en jeu. Cet aspect transparaît peu dans le cas des données bibliographiques ; on peut imaginer d'autres terrains pertinents pour prolonger ces réflexions, par exemple les portails *open data* des administrations publiques.

CONCLUSION

Ce qui fait document influence ce que nous faisons avec les documents. Cette logique s'applique à tous les objets info-communicationnels. Les données, dont les problématiques en matière d'épistémologie commencent à être réorientées en direction des questions interprétatives, illustrent particulièrement bien les logiques complexes qui président à leur valeur documentaire. En tentant d'évaluer leur degré de documentarité, nous réalisons que celle-ci s'exprime le plus souvent par des jeux d'écriture qui, des principes de structuration aux possibilités de réutilisation en passant par les modes d'éditorialisation, façonnent leurs contours et leur devenir. Il en résulte un triple enjeu de formation pour les projets liés aux données et mégadonnées en SHS, dont l'appropriation reste suspendue au développement d'une véritable culture technique. À cela, nous pensons que la théorie peut apporter une certaine contribution, notamment lorsque les concepts éclairent les continuités indiscutables entre questionnements anciens et actuels.

RÉFÉRENCES BIBLIOGRAPHIQUES

BACHIMONT Bruno (2004). *Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle*. Mémoire d'habilitation à diriger des recherches. Compiègne : Université de technologie de Compiègne.

BEUAPARLANT Sophie (2017). Dieu n'a pas inventé le webdocumentaire. *Sens Public* [en ligne]. 15 décembre 2017. Disponible à l'adresse : <http://sens-public.org/article1276.html>.

BRIET Suzanne (1951). *Qu'est-ce que la documentation ?* Paris : Éditions documentaires, industrielles et techniques.

BUCKLAND Michael K. (1997). What is a "document"?. *Journal of the American Society for Information Science*. 1997. Vol. 48, n° 9, p. 804-809. DOI 10.1002/(SICI)1097-4571(199709)48:9<804::AID-ASI5>3.0.CO;2-V.

BUCKLAND Michael K. (2017). Before the Antelope: Robert Pagès on Documents. *Proceedings from the Document Academy* [en ligne]. 2017. Vol. 4, n° 2. Disponible à l'adresse : <http://ideaexchange.uakron.edu/docam/vol4/iss2/6>.

COLLOMB Cléo (2017). Faire compter les machines. *Communication* [en ligne]. 2017. Vol. 34, n° 2. DOI 10.4000/communication.7327. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-01633886>.

COURBIÈRES Caroline (2008). La question documentaire à l'épreuve du numérique : le recours aux fondamentaux. In : *Sciences de la Société*. 2008. n° 75, p. 40-51.

CROZAT Stéphane (2016). Proposition : principe de documentarité. In : *As we may...* [en ligne]. 2016. Disponible à l'adresse : <http://aswemay.fr/co/010013.html>.

CROZAT Stéphane (2019). Le document numérique n'existe pas, il faut l'inventer (principe de documentarité). *Masterclass DNHD #1* [en ligne]. Université Bordeaux Montaigne : s.n. 2019.

Disponible à l'adresse : <https://stph.scenari-community.org/pres/20190222-documentarite/co/20190219-documentarite.html> .

DAY Ronald (2018). Auto-Documentality as Rights and Powers. *Proceedings from the Document Academy* [en ligne]. 14 décembre 2018. Vol. 5, n° 2. Disponible à l'adresse : <https://ideaexchange.uakron.edu/docam/vol5/iss2/3> .

DAY Ronald E. (2019). *Documentarity: Evidence, Ontology, and Inscription*. Cambridge : MIT Press; 2019. ISBN 978-0-262-04320-5.

DRUCKER Johanna (2011a). Humanities Approaches to Graphical Display. *Digital Humanities Quarterly* [en ligne]. 2011. Vol. 5, n° 1. Disponible à l'adresse : <http://digitalhumanities.org/dhq/vol/5/1/000091/000091.html> .

DRUCKER Johanna (2011b). Humanities Approaches to Interface Theory. *Culture Machine* [en ligne]. 2011. Vol. 12. Disponible à l'adresse : <http://svr91.edns1.com/~culturem/index.php/cm/article/download/434/462>.

FERRARIS Maurizio (2013). *Documentality: why it is necessary to leave traces*. New York : Fordham University Press. ISBN 978-0-8232-4968-8. BD336 .F4613 2013

FROHMANN Bernd (2012). The documentality of Mme Briet's antelope. In : PACKER Jeremy et WILEY Stephen B. Crofts (éd.), *Communication Matters: Materialist Approaches to Media, Mobility and Networks*. London : Routledge. p. 173-182.

GAUDREAUULT André et MARION Philippe (1994). Dieu est l'auteur des documentaires... *Cinéma : Revue d'études cinématographiques / Cinéma : Journal of Film Studies*. 1994. Vol. 4, n° 2, p. 11-26. DOI 10.7202/1001020ar.

GOYET Samuel (2017). *De briques et de blocs. La fonction éditoriale des interfaces de programmation (API) web : entre science combinatoire et industrie du texte* [en ligne]. Thèse de doctorat. Paris : Université Paris-Sorbonne. Disponible à l'adresse : <https://tel.archives-ouvertes.fr/tel-01665406/> .

LATOURE Bruno (1999). *Pandora's hope: essays on the reality of science studies*. Cambridge : Harvard University Press, 1999. ISBN 978-0-674-65335-1.

LATOURE Bruno (2007). Pensée retenue, pensée distribuée. In : JACOB, Christian (éd.), *Lieux de savoir*. Paris : Albin Michel. p. 605-615.

MEYRIAT Jean (1981). Document, documentation, documentologie. In : COUZINET, Viviane (éd.), *Jean Meyriat, théoricien et praticien de l'information-documentation*. rééd., 2001. Paris : ADBS Éditions.

OTLET Paul (1934). *Traité de documentation. Le livre sur le livre*. 3. Bruxelles : Les Impressions nouvelles , 2015. ISBN 978-2-87449-299-0.

PAGÈS Robert (1948). Transformations documentaires et milieu culturel (Essai de documentologie). *Revue de Documentation*. 1948. Vol. 15, n° 3, p. 53-64.

PÉDAUQUE Roger T. (2007). *La redocumentarisation du monde*. Toulouse : Cepaduès-Éd.

ROBERT Pascal (2015). Le biblion et les substituts du livre. Théorie et pratique du dépassement du livre chez Paul Otlet. *Communication & langages*. 2015. Vol. 2015, n° 184, p. 3-23. DOI 10.4074/S0336150015012016.

SCOPSI Claire (2018). The Documentality of Memory in the Post-Truth Era. In : *Proceedings from the Document Academy* [en ligne]. 2018. Vol. 5, n° 2. Disponible à l'adresse : <https://ideaexchange.uakron.edu/docam/vol5/iss2/4>.

ZINS Chaim (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*. 15 février 2007. Vol. 58, n° 4, p. 479-493. DOI 10.1002/asi.20508.