



HAL
open science

Vers l'objet documentaire (re)contextualisé

Evelyne Broudoux

► **To cite this version:**

Evelyne Broudoux. Vers l'objet documentaire (re)contextualisé. Objets documentaires numériques : nouvel enseignement ?, 9e congrès des enseignants documentalistes de l'Éducation nationale, Mar 2012, Paris, France. pp. 47-51. sic_00715868

HAL Id: sic_00715868

https://archivesic.ccsd.cnrs.fr/sic_00715868

Submitted on 9 Jul 2012

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Vers l'objet documentaire (re)contextualisé

Evelyne Broudoux

9e congrès des enseignants documentalistes de l'Éducation nationale

Objets documentaires numériques : nouvel enseignement ?

22 mars 2012

INTRODUCTION

Dans une période pré-numérique, pour répondre aux demandes du public, les professionnels de la documentation et des bibliothèques choisissent un ensemble de documents à extraire d'un centre de ressources selon une classification documentaire adaptée aux supports analogiques. Les documents rangés selon la classification Dewey, CDU, ou entêtes matières Rameau sont décontextualisés. C'est-à-dire que les contextes de leur production, de leur diffusion, de leur influence nécessitent une reconstruction éditoriale a posteriori pour pouvoir être analysés et compris. Les choix des éléments méritant de figurer dans une bibliothèque ou un centre de ressources documentaires s'appuient essentiellement sur les sélections opérées en amont par les *gatekeepers* des industries culturelles.

Dans un contexte de numérique en réseau, les acteurs médiotechniques que sont les moteurs de recherche horizontaux comme Google orientent les résultats selon des algorithmes conçus pour servir des stratégies d'influence. Les graphes des réseaux formés par les renvois entre url révèlent des univers fortement connectés d'où vont émerger des objets informationnels classés selon leur notoriété calculée à partir de leur pagerank. D'un autre côté, le folkrank est venu ajouter une couche « sociale » d'avis personnels d'internautes sur la valeur de protodocuments, de documents, de conversations, ajoutant en cela une facette « critique » à la publication. Du web de documents, on est ainsi passé au web social pour arriver au web de données avec une orientation « sémantique ».

Sont donc entrés en compétition un système « lent » de sélection qualitative d'objets nécessitant des fabrications industrielles (livres, cd, dvd, etc.) et un système « rapide » fournissant des résultats « à la volée » d'objets qui connaissent souvent des délais très courts entre leur fabrication et leur publication, consommation. Entre les deux systèmes, l'étape de filtrage *a priori* caractéristique de l'édition imprimée a disparu, ou plutôt a été déplacée *a posteriori* avec la critique exercée par les commentaires, les évaluations de type (j'aime) ou par étoile (*). L'édition s'est ouverte à l'autorativité, à l'agrégation-curation et à la publication pour tous et se côtoient sur le web des documents aux nombreuses versions modifiées, des brouillons, des vidéoblogs, des articles de journalistes, des podcasts de chercheurs et de multiples plagiats... Cette diversification qui fait se côtoyer protodocuments et documents est rendue visible par les moteurs de recherche qui utilisent des algorithmes aux variables aisément manipulables.

Ce désordre a suscité bien des critiques négatives émanant des autorités traditionnelles et a provoqué dans un premier temps le rejet des élites. L'exemple emblématique de Google qui soutient Wikipédia en classant de manière systématique en tête de liste les résultats de la recherche liés aux contenus de l'encyclopédie est indicateur de la dynamique d'innovation poursuivie par le célèbre moteur de recherche. Une stratégie est-elle à l'œuvre dans cette recomposition de l'organisation des connaissances ? Assiste-t-on à la décomposition progressive d'un monde qui laisse progressivement sa place à de « nouvelles » manières de faire ?

Rappelons que l'ordre alphabétique de l'Encyclopédie de Diderot et d'Alembert a été interprété comme le signal de l'émancipation des intellectuels inféodés aux autorités religieuses, pendant les siècles des Lumières. Pour la première fois se trouvaient ordonnées des connaissances selon le classement arbitraire alphabétique alors que celles-ci par un astucieux système de renvois formaient les premières arborescences des classifications scientifiques actuelles. Cette objectivation du savoir trouve-t-elle ses limites aujourd'hui avec les algorithmes des moteurs de recherche qui servent surtout des stratégies de communication ?

Pour tenter de répondre à cette question, nous proposons de regarder les facteurs à l'origine de ce processus d'évolution, ce qui va nous permettre d'envisager de nouvelles orientations de développement aux « futurs » systèmes d'organisation des connaissances.

1) Caractéristiques des documents numériques

Tout d'abord, notons que la fin du document en tant qu'entité dépendante d'un enregistrement sur un support est à la base de la réorganisation de son identification. La libération du document de son support matériel provoque un ensemble de modifications que l'on peut tenter de regrouper.

- a. Se généralise la **générativité documentaire** qui consiste en un processus calculatoire visant à produire une occurrence de document en provenance de plusieurs sources distribuées (web de données). La durée de vie du document généré sera directement dépendante de son usage. Il en résulte un déplacement spatio-temporel qui perturbe les limites traditionnelles attribuées au document et à son cycle de vie.
- b. L'accroissement de la **diversité informationnelle** des documents rend leur typologie dépendante de leur caractérisation informatique. Les formats des documents autorisent des opérations, des calculs, de l'interactivité, de l'immersion, de la fouille, des mises à jour et des modifications. Par exemple, les données statistiques, les cartes interactives géolocalisant de l'information périodiquement tenue à jour, les infographies dynamiques, sont typiquement des objets informationnels pouvant faire tout ou partie de documents.
- c. Une **mutation des systèmes d'identification** des documents est en cours. Elle se caractérise par une externalisation de la représentation documentaire à l'aide d'attributs rassemblés sous le terme de métadonnées. Cette représentation par des métadonnées autorise le traitement algorithmique des documents et leur gestion globale par des systèmes d'informations documentaires susceptibles d'entrer en liaison. Les métadonnées peuvent pré-exister et servir à la description documentaire et/ou bien être générées à partir des contenus et servir à la caractérisation et à la fouille de textes.
- d. Un document numérique est à même de conserver un certain nombre d'empreintes à tout les stades de son cycle de vie : les métadonnées sont aptes à prendre en compte les différentes facettes de la conception, de l'évolution intrinsèque, du versioning, du reviewing, de l'indexation, du référencement, des traces de lecture et d'adoption (annotations, commentaires, modifications ou remix) de chaque document numérique pris dans un réseau sémantique ou référé par de « simples » liens hypertextes. Il s'agit donc véritablement d'une **représentation enrichie par les empreintes**.

Revenons sur la caractéristique qui concerne l'évolution des systèmes d'identification des contenus sur le web, des documents numériques et de leurs métadonnées. Chaque ressource du web (document, service, logiciel, etc.) décrite par le modèle formel RDF, normalisé par le W3C, est désignée par une URI qui la rend interopérable. Manipulables et interprétables par des programmes et des humains, les objets documentaires encodés en RDA combinent des informations d'adressage et d'identification en un seul modèle unifié capable de s'adapter à l'évolution dynamique du web.

On observe une convergence entre les techniques de description documentaires et les techniques d'adressage et d'identification des ressources sur le web permettant de construire des systèmes d'organisation des connaissances (cf. ontologies évolutives).

Afin de mieux comprendre l'ensemble de possibles ouvert pour les prochaines années, il apparaît nécessaire de différencier un certain nombre de concepts.

2) Objets documentaires, informationnels, documents, systèmes d'informations

- a. Différences entre un document analogique, numérique et un objet informationnel

Un « objet informationnel numérique » peut être caractérisé par un ensemble de couches invisibles, observables et actables. Celles-ci sont constituées par la partie logicielle des nouveaux objets de communication qui modèlent différentes mises en formes à partir desquelles scénarisation auctoriale et hiérarchisation éditoriale vont pouvoir s'organiser (Broudoux et al., 2005, 2007). Un objet informationnel numérique peut être un service, un document, un portail, etc.

Le document numérique se distingue du document ancré sur un support analogique par un changement en termes de maniabilité – dans le sens où l'objet n'est plus conçu pour répondre aux sollicitations directes des mains (ex : pages reliées en feuillets que l'on peut tourner, changements de disques, appareils de réception réagissant à des impulsions par

boutons, curseurs ou manettes interposées, etc.). S'interpose entre le lecteur et le document numérique des « périphériques » maniables : souris, clavier, écran tactile, trackpad qui agissent sur les couches invisibles de l'objet informationnel numérique. Sa maniabilité est au contraire tout entière conditionnée par ces couches codées qui vont fournir aux usagers des instruments de lecture, d'exploration et d'extraction de contenus, et cela de manière intuitive, contrainte ou fortement structurée.

b. Différences entre un objet documentaire numérique et un objet informationnel

Par « objet documentaire », est entendu tout élément porteur d'informations quel que soit le support (documents textes, images, multimédia, prototypes, matériaux, etc.) susceptible d'être pris en compte dans un système d'information (Boulogne et Dalbin, 2004).

Un objet documentaire numérique se distingue d'un objet informationnel numérique par une fonction d'identification documentaire réglée par un système d'information. Celui-ci répertorie l'ensemble des conventions qui organisent les objets afin de gérer leur cycle de vie, le repérage des contenus, leur classification et vise à répondre à la recherche et à la diffusion par leur identification.

Alors que dans l'objet documentaire analogique, cette identification est en grande partie dépendante du support qui le donne à lire, pour l'objet numérique, l'identification se réalise par des métadonnées de description et de calcul de mise en relations qui sont démultipliées dans le contexte du web.

3) Evolution des systèmes d'informations en fonction de ces différences

a. Les lieux de distribution, de diffusion et de stockage des documents se sont diversifiés : l'apparition d'un univers de dissémination qui n'existait pas auparavant provoque des remous et un mélange des rôles, professions, des activités confrontées au numérique, des genres de documents et de leurs modalités d'appropriation.

i. Avant le web : les acteurs des industries culturelles se distribuaient entre édition, diffusion, distribution avec de secteurs délimités comme par exemple l'audiovisuel télévisuel, les librairies-bibliothèques, les centres culturels, et le stockage patrimonial s'organisait en parallèle à ces activités.

ii. Avec le web : l'édition s'enrichit de la circulation entre amateurs, les documents perdent en stabilité avec leurs améliorations successives ou au contraire leur appauvrissement. Les moteurs de recherche en tant qu'acteurs majeurs de cette circulation ont créé un univers de règles échappant à celles héritées des autorités institutionnelles. Les documents circulent à toutes les étapes de leur cycle de vie, y compris celle de leur stockage patrimonial.

b. Nouveaux systèmes d'information documentaire (SID)

Selon la définition traditionnelle, les SID sont des systèmes permettant d'identifier des documents et de les gérer en collections. La gestion implique le repérage des sources et leur validation, la collecte, le traitement, le stockage et la diffusion des documents. Selon la finalité des systèmes d'information documentaires, des objectifs précis sont atteints : conservation pour les bibliothèques, archives pour le patrimoine, référentiels pour les entreprises, etc.

Les métadonnées concernent les documents – et peuvent inclure des descripteurs servant à l'indexation – et des collections de documents. Se sont ajoutées aux composantes classiques de description des documents analogiques (Titre, auteurs, année de publication, etc.), un ensemble de caractéristiques spécifiques aux documents numériques (ex : versionning de documents, champs destinés à l'identification de personnes, etc.).

Compte-tenu du fait que les métadonnées susceptibles d'être prises en compte dans les SID sont *a priori* en nombre non fini, il est tout-à-fait possible d'imaginer des SID adaptés à des projets de collections de bibliothèques numériques, c'est-à-dire réellement personnalisables selon les objectifs envisagés. L'interopérabilité des SID qui permet l'enrichissement et la mutualisation de données bibliographiques ouvre un espace de mise en collectif d'expériences qui profitera à tous.

4) Un document calculé pris dans un contexte redéfini

a. La notion de contexte

- i. La notion de contexte a comme caractéristique son évolutivité et ses toutes récentes diversifications liées à l'avancée des techniques de calcul dans la recherche d'informations.
- ii. Le sens commun attribue au terme contexte l'ensemble des circonstances qui coproduisent des événements ou à partir desquelles s'insèrent des faits. Cette signification naît au milieu du XIXe siècle et perdure aujourd'hui.
- iii. Si nous nous attachons à la définition latine de contexte : *con* (avec) *textus* (tissu), on peut décliner le contexte selon plusieurs points de vue scientifiques comme celui de la linguistique (un ensemble de relations entre les éléments d'un discours) ou celui de l'intertextualité littéraire (un entrelacement de significations insérant un texte dans un arrière-plan constitué par d'autres textes).

b. Caractéristiques du contexte

Une définition actuelle de la notion de contexte implique trois notions :

- i. Le cadre de perception qui est l'ensemble des constituants du décodage de l'environnement menant à son interprétation et à sa compréhension.
- ii. L'ancrage spatio-temporel : tout événement, toute action prend place dans un espace/temps conditionnant un ensemble de possibles et de représentations.
- iii. Le concept d'actions - qui produisent leur propre interprétation - dénommées « actions situées » par Lucy Suchman nous permet de d'envisager le contexte de manière phénoménologique : le contexte n'existe pas en soi, il est le prisme à partir duquel nous pouvons interpréter et agir sur notre environnement.

Dans le cadre de la documentation, la classification d'un titre consiste à insérer ses métadonnées de description en étant guidé par une arborescence systématique d'organisation des connaissances, selon des règles précises. Cette opération fait appel à une fonction de contextualisation que l'on pourrait qualifier de normée, dans la mesure où il existe des éléments de classement auxquels on fait appel de manière systématique comme le lieu géographique éventuel auquel référer le document et la période historique qu'il concerne.

Le processus d'indexation consiste donc pour une part à poser les bases d'une future contextualisation du document par les lecteurs, par une analyse de contenu et par les métadonnées existantes (titre, auteur, code d'identification, etc.).

c. Contexte calculé

- i. Dans le contexte du web actuel, les métadonnées renseignant les contenus sont quasiment inexistantes et le classement des résultats en tête de liste par les moteurs sont souvent le résultat de stratégies d'influences – rassemblées sous le terme de SEO¹ - échappant aux spécialistes du « document », du « livre » et de « l'archive ».
- ii. Parallèlement à cet état de fait, le web3 (« linked data ») en tant qu'écosystème informationnel en construction où les objets traitent et produisent des données qu'ils communiquent est un générateur de métadonnées. Les ressources concernées sont intra et extradocumentaires (documents et fragments documentaires, individus, entités, conversations, logiciels, etc.). Ce web des objets documentaires communicants est aujourd'hui pris en compte dans les stratégies de développement du moteur Google avec le *knowledge graph* qui regroupe les données sémantiques de Wikipédia. Sont ainsi automatiquement agrégées les informations contenues par Wikipédia comme résultat de la recherche d'informations dans un encart

¹ Search Engine Optimization

spécifique (infobox). Cette automatisation est une véritable fabrique d'auteurs comme le remarque justement Bernhard Rieder²...

d. Missions des centres de documentation

Dans cette phase transitionnelle du web, les missions des médiateurs des documents évoluent. Car il s'agit de prendre en compte les réseaux d'influences intervenant avant la médiation documentaire et d'objectiver le document en le recontextualisant selon les objectifs des usagers. D'abord, le professeur documentaliste doit apporter des explications sur le mécanisme de la générativité documentaire et être capable de déconstruire les circonstances qui font qu'un document est le produit d'un effet de viralité sur un réseau social. Dans un deuxième temps, il s'agira d'apprendre à se servir des nombreuses métadonnées renseignant les contenus du web³.

Plus généralement, c'est de la construction de systèmes d'information personnalisés dont il est question. Les années à venir seront placés sous le signe de la profusion de métadonnées interopérables qui vont caractériser de multiples façons les contenus et les systèmes d'information seront à même de trier des documents selon des critères renouvelés de sélection. Ceci permettra peut être le prélèvement d'une empreinte d'influence afin de restituer le contexte de l'émergence des documents mais aussi leur recontextualisation pour des besoins spécifiques : dossiers documentaires, archives thématiques, etc.

Penser que toute information publiée sur le web peut être envisagée sous la forme d'une « action située », s'ancrant dans un cadre de perception spatio-temporel précis, donne les éléments de base nécessaires à la construction d'une attitude critique, condition première de l'appropriation des connaissances.

La notion d'« information literacy » qui est de savoir « quand et pourquoi on a besoin d'information, où la trouver, et comment l'évaluer, l'utiliser et la communiquer de manière éthique »³ se retrouve convoquée dans toute sa complexité. Car la capacité à identifier, apprécier, rechercher, évaluer, adapter, organiser et communiquer l'information dans un contexte itératif d'examen et de réflexion⁴ suppose une connaissance précise des mécanismes techniques qui renouvellent les médiations documentaires.

CONCLUSION

Un document numérique sur le web apparaît pris dans un entrelac d'influences. Sa trouvabilité est conditionnée par un ensemble de marques indicatrices du contexte dans lequel il a été généré. Les nouveaux formats de description qui conditionnent la trouvabilité des documents sont à même de fabriquer plusieurs sortes de classement : statistiques, sémantiques, etc. mais surtout personnalisables...

Les rôles des documentalistes et des bibliothécaires évoluent. Chargés de proposer des documents matériels à des lecteurs émettant des souhaits, les bibliothécaires et les documentalistes dont le rôle

² Rieder Bernhard. Billet du 21 mai 2012. Hitler, l'auteur : <http://thepoliticsofsystems.net/2012/05/hitler-lauteur/>

³ **CILIP's definition of information literacy:**

Information literacy is knowing when and why you need information, where to find it, and how to evaluate, use and communicate it in an ethical manner.

This definition implies several skills. We believe that the skills (or competencies) that are required to be information literate require an understanding of:

- A need for information
- The resources available
- How to find information
- The need to evaluate results
- How to work with or exploit results
- Ethics and responsibility of use
- How to communicate or share your findings
- How to manage your findings

<http://www.cilip.org.uk/get-involved/advocacy/information-literacy/Pages/definition.aspx>

⁴ "the ability to identify, assess, retrieve, evaluate, adapt, organise and communicate information within an iterative context of review and reflection" . <http://www.rin.ac.uk/our-work/researcher-development-and-skills/what-is-information-literacy>

était d'extraire des documents rangés selon une classification documentaire adaptée aux supports matériels voient leurs compétences évoluer : à eux de construire des systèmes de classification modulable de documents numériques adaptés aux situations. Leur rôle de médiation évolue.

Bibliographie

Boulogne Arlette [et al.]. *Vocabulaire de la documentation*. Paris, ADBS éditions, 2004. 334 p. Sciences et techniques de l'information. ISSN 1762-8288. ISBN 2-84365-071-2

Broudoux E., Grésillaud S., Le Crosnier H., Lux-Pogodalla V. « Construction de l'auteur autour de ses modes d'écriture et de publication » in « *Créer, jouer, échanger : expériences de réseaux* ». H2PTM'05. p.123-142, Hermès/Lavoisier, novembre 2005.

Broudoux E. Construction de l'autorité informationnelle sur le web (dir. Skare R., Windfield Lund N., Varheim A.) in « *A Document (re)turn (contributions from a research field in transition)* ». Peter Lang/Europäischer Verlag der Wissenschaften, pp.265-278, mars 2007.

Broudoux E., Scopsi C. (dir.). Dossier « *Métadonnées du web : les enjeux autour des techniques d'enrichissement des contenus* » in Revue Etudes de communication. Numéro 36, 2011.

Calderan L., Hidoine B., Millet J. *Métadonnées : mutations et perspectives*. Paris, ADBS, 2008.

Moles Abraham. « Phénoménologie de l'action » in *Les théories de l'action*. Centre d'études et de promotion de la lecture, Paris, 1972.

Suchman Lucy. *Plans and situated actions : the problem of human-machine communication*, Cambridge University Press, New York, 1987.