

À paraître dans :

Études de Communication n° 30, Entre information et communication,
Les nouveaux espaces du document, Université de Lille 3, décembre 2007.

La redocumentarisation, un défi pour les sciences de l'information

Jean-Michel Salaün

Directeur de l'École de bibliothéconomie et des sciences de l'information
Université de Montréal

Les sciences de l'information viennent de passer un cap. Elles s'étaient construites sur un ordre documentaire relativement stable durant le XXème siècle. Même s'il subissait une explosion quantitative, même s'il intégrait de nouveaux supports comme l'audiovisuel, sa structure générale suivait des règles et canaux qui étaient repérés. Les sciences de l'information pouvaient alors se construire sur un corpus de documents soumis à des contraintes connues et proposer des méthodes et des outils pour en tirer l'information utile comme le classement, l'indexation, des équations de recherche de plus en plus raffinées, la scientométrie, etc. pour les traditions issues de la bibliothéconomie, et d'autres encore, comme l'évaluation, l'élimination, le calendrier de conservation de son autre racine professionnelle, plus méconnue en France : l'archivistique. L'informatique était dans cette perspective un outil de plus, sans doute révolutionnaire par sa puissance, mais qui ne remettait pas en cause les fondements du raisonnement.

Pour ma part, j'avais montré que les sciences de l'information se distinguaient des sciences de la communication par la perspective prise : celle de la recherche et de l'accès, c'est-à-dire la perspective de l'utilisateur et non celle de la création et de la diffusion, perspectives du producteur propres aux sciences de la communication (Salaün, 1993). Yves Le Coadic (1994), de son côté, pouvait alors légitimement proposer d'enlever un «s» à sciences, parlant de «la science de l'information» au singulier, même si elle faisait appel, en particulier pour former aux professions afférentes, à d'autres disciplines, comme l'histoire, la gestion, la sociologie ou le droit.

Dans cet article, construit comme une rapide proposition, je m'efforcerai de montrer que cette période est derrière nous et qu'il faut, de nouveau au XXIème siècle, s'exprimer au pluriel. Le numérique transforme, en effet, le cœur même du raisonnement, l'objet sur lequel était basé la construction de la dite science de l'information, même s'il lui était arrivé de l'oublier : le document. Et dès lors, l'ordre documentaire est remis en cause. Il ne faut pas s'en plaindre, c'est une période nouvelle passionnante qui s'est ouverte, un défi pour les chercheurs du domaine, pour peu qu'ils acceptent de ne pas rester enfermés dans leurs anciennes chapelles.

La proposition sera bâtie en trois temps. Tout d'abord, je m'efforcerai de replacer rapidement la bascule en cours dans son évolution historique. Puis, je présenterai la notion de «redocumentarisation», en m'appuyant très largement sur les réflexions menées

au sein du collectif Roger T. Pédaque (cette partie paraphrasera largement Roger et le lecteur familier de ses écrits en retrouvera des traces que je n'ai pu matériellement isoler sous forme de citations, il me pardonnera sûrement !) (2006). Enfin je laisserai la parole, pour conclure, à des chercheurs qui, d'un côté et de l'autre de l'Atlantique, rappellent que les fondements, les racines historiques des sciences de l'information sont pertinents pour comprendre et maîtriser les mouvements actuels du numérique.

Les ordres documentaires

Sans remonter à l'Antiquité mais simplement à l'apparition de l'imprimerie, il est possible de repérer à grands traits plusieurs périodes dans les médias. Bien des auteurs ont proposé leur découpage. Les historiens du livre ont l'habitude d'évoquer «les trois révolutions du livre» (Gutenberg, l'ère industrielle, la dématérialisation) (Barbier, 1998), mais si on raisonne en termes de documents imprimés et non de livres, ce découpage est trompeur. À partir de l'imprimé, A. Marshall, sur la base de son expérience de directeur du musée de l'imprimerie à Lyon a proposé dans une conférence à Montréal (2006) une autre périodisation comprenant quatre moments. J'adapte ici le vocabulaire à ma compréhension : l'âge du livre (Gutenberg au XIX^{ème} siècle), l'âge de la presse (XIX^{ème}), l'âge de la paperasse (XX^{ème}) et le quatrième moment que j'appellerais volontiers l'âge des fichiers (XXI^{ème}). Nous démarrons donc par l'ordre des livres (Chartier, 1992) où l'imprimé dominant est bien le codex qui débouche sur la bibliothèque ; puis la presse grand public s'installe avec l'imprimerie industrielle, la rotative, et l'organisation des États modernes ; avec le développement explosif du commerce et de l'État-providence, contrats, factures, formulaires, circulaires, bordereaux, réclames (A. Marshall (2003) parle de «paperasse») envahissent le monde du travail et de l'administration tandis que les machines à imprimer se font plus légères et permettent les petits tirages ; enfin le numérique explose avec la redocumentarisation que je définirai plus loin, une nouvelle modernité et la composition et l'impression mise à la disposition de tous par l'intermédiaire du réseau et des ordinateurs personnels. Ce phasage, ici grossièrement évoqué (il faudrait en effet dépasser l'imprimé et au moins lui adjoindre la montée de l'audio-visuel), montre assez la relation étroite entre un ordre documentaire et une société.

Mon objectif n'est pas de faire ici œuvre d'historien, mais de remarquer que les sciences de l'information ont été initiées à la charnière entre la seconde et la troisième périodes, dans un mouvement que l'on pourrait baptiser de «documentarisation» au moment où l'explosion du nombre de documents a conduit à construire ou développer des techniques de gestion documentaire : bibliothéconomie, documentation, archivistique, science administrative.

Pour en rester à la bibliothéconomie, c'est, par exemple, au tournant du XIX^{ème} et XX^{ème} siècle qu'un Paul Otlet en Belgique ou un Melvin Dewey aux États-Unis ont construit les classifications décimales universelles. Parallèlement et en étroite correspondance des utopistes entrepreneurs proposaient des réalisations, comme O. Lafontaine sur un plan politique ou W. Carnegie en inventant la philanthropie moderne et finançant la construction de bibliothèques sur toute la planète. À partir de ce moment, on s'est mis à classer et indexer systématiquement les documents. Par la suite, le mouvement

s'est poursuivi d'abord sous l'appellation de documentologie (Otlet, 1934), puis de sciences de l'information, voulant s'émanciper du support pour ne s'intéresser qu'au contenu, non sans ambiguïtés mais c'est une autre histoire.

La redocumentarisation

Pour définir la re-documentarisation, il faut commencer par s'entendre sur le terme «documentarisation». Documentariser, c'est ni plus ni moins traiter un document comme le font, ou le faisaient, traditionnellement les professionnels de la documentation (bibliothécaires, archivistes, documentalistes) : le cataloguer, l'indexer, le résumer, le découper, éventuellement le renforcer, etc. On préfère «documentariser» à «documenter», qui renvoie plutôt à la création d'un ou de plusieurs documents pour expliquer un objet ou une action, mais dans certains cas les deux activités se recoupent. L'objectif de la documentarisation est d'optimiser l'usage du document en permettant un meilleur accès à son contenu et une meilleure mise en contexte.

Le numérique, par nature, implique une re-documentarisation. Dans un premier temps, il s'agit de traiter à nouveau des documents traditionnels qui ont été transposés sur un support numérique en utilisant les fonctionnalités de ce dernier. Mais le processus ne se réduit pas à cette simple transposition. En effet, bien des unités documentaires du Web ne ressemblent plus que de très loin aux documents traditionnels. Dans le Web 2.0, dans la construction du Web sémantique ou tout simplement sur les sites dynamiques, la stabilité du document classique s'estompe et la redocumentarisation prend une tout autre dimension. Il s'agit alors d'apporter toutes les métadonnées indispensables à la reconstruction à la volée de documents et toute la tracabilité de son cycle. Les documents traditionnels eux-mêmes, dans leur transposition numérique, acquièrent la plasticité des documents nativement numérique et peuvent profiter des possibilités de cette nouvelle dimension.

La redocumentarisation prend alors un sens beaucoup plus large. Cette nouvelle forme de documentarisation reflète ou tente de refléter une organisation post-moderne de notre rapport au monde, repérable aussi bien dans les sphères privée, collective et publique, qui d'ailleurs se superposent de plus en plus. Comme dans la précédente modernisation, le document participe au processus et y joue même un rôle clé, mais il s'est transformé au point que l'on peut se demander s'il s'agit encore de la même entité.

Pourquoi alors reprendre le même terme, en ajoutant juste le préfixe re-, s'il s'agit d'un changement de paradigme ? En réalité, s'il y a bien une rupture, celle-ci est dans une continuité historique qu'il est d'autant plus important de souligner que les professions de la documentation y ont leur place, ou devraient y prendre une place plus grande qu'elle n'est aujourd'hui car leurs compétences y sont essentielles.

La première documentarisation ne s'est pas développée par hasard, nous l'avons dit. Elle s'inscrit clairement dans les quatre âges de l'imprimé entre le deuxième (la presse) et le troisième (la paperasse). La redocumentarisation marque le passage du troisième au quatrième (fichiers). Ces âges accompagnent des organisations sociales et idéologiques différentes. Le tout peut se résumer sur le tableau ci-dessous.

	Documentarisation	Redocumentarisation
Dates	Tournant XIXe-XXe	Tournant XXe-XXIe
Quelques figures	M. Dewey, P. Otlet, O. Lafontaine, W Carnegie	T. Berners-Lee, T. Nelson, B. Gates, S. Brin
Quelques techniques	Classification, Indexation, Langages documentaires, Thésaurus..	Protocoles Web (Html, Url) Web 2.0, Web sémantique Ontologies..
Quelques réalisations	Réseau mondial de bibliothèques	Google, Wikipédia
Les modernités	L'esprit scientifique, la raison- logique, l'État-nation, les votes, l'industrie, l'auteur ..	Le savoir limité, la raison-statistique, l'individu, les opinions, les services, la réflexivité ..
Quelques objets documentaires concernés	Les revues, les règlements, les contrats, les brevets, les œuvres, les médias et l'imprimerie.	Les pré-publications, les formulaires, les sources ouvertes, les wikis, les blogues et le web

Figure 1 : Les deux bascules documentaires

L'intérêt d'une telle proposition est de s'extirper du brouillard produit par l'extrême rapidité des changements actuels en les plaçant dans la longue durée et aussi de souligner la place que doivent y prendre les sciences de l'information, à condition de bien en assumer le caractère interdisciplinaire.

Reste à rendre la proposition opérationnelle. À cette fin, le collectif Roger T. Pédaque a proposé de classer les recherches en cours sur le document numérique selon trois dimensions, qui sont apparues comme les trois dimensions constitutives d'un document : la forme ou le signe, le contenu ou le texte, le médium ou la relation. Les modalités **anthropologiques** (lisibilité-perception, forme-signé), **intellectuelles** (intelligibilité-assimilation, texte-contenu) et **sociales** (sociabilité-intégration, médium-relation) du document doivent non seulement être efficaces prises chacune séparément, mais encore être cohérentes entre elles.

Les sciences de l'information ont pris en compte, sans toujours le conceptualiser clairement, ces trois dimensions en classant les objets documentaires (anthropologique), en indexant leurs contenus (intellectuel) et en organisant des systèmes d'accès (sociales). Ainsi on ordonne le système documentaire. En permettant de retrouver les objets-documentaires on évite le chaos, en permettant d'y retrouver l'information on réduit la cacophonie (le silence et le bruit documentaire) et en donnant au document une vie qui

dépasse un usage local et immédiat, on lui donne sa légitimité. L'objectif est chaque fois de mettre de l'ordre, un ordre documentaire, afin de pouvoir retrouver les documents utiles. Le tout peut être résumé sur deux schémas, le premier rappelle les dimensions constitutives du document, le second montre comment celles-ci sont documentarisées pour bâtir un ordre documentaire. Chacune des dimensions est alors annotée, on dirait aujourd'hui qu'on lui ajoute des métadonnées.

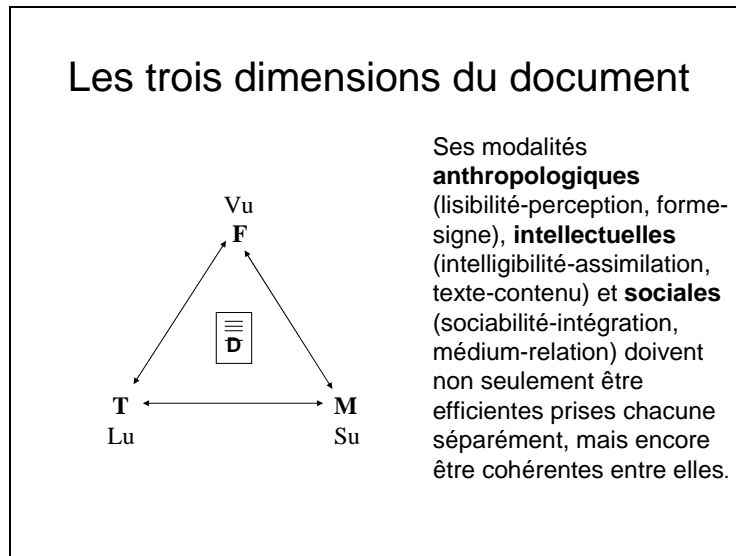


Figure 2 : Les 3 dimensions du document

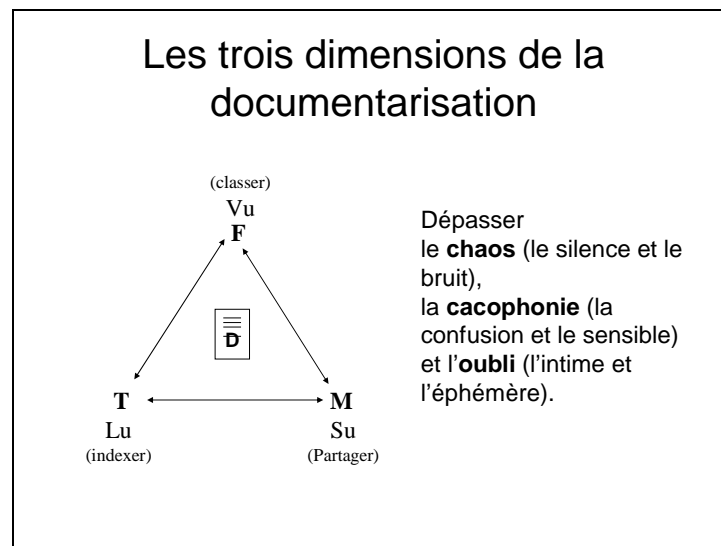


Figure 3 : Les 3 dimensions de la documentarisation

La période actuelle rebrasse brutalement les cartes sur chacune des trois dimensions et pour analyser correctement ces mouvements, il est indispensable de croiser des expertises venant de disciplines aux traditions bien différentes :

- La stabilité des supports a été remplacée par celle des structures dans des formats de fichiers qui séparent le contenu de sa mise en forme. Le numérique a déplacé la

question du support du document vers la problématique de sa structure en l'éclatant. La popularité croissante du langage XML marque une étape, laissant ouverte la question de la perception. Et, sans toujours le savoir, les chercheurs en traitement d'images ou encore les experts du consortium W3C retrouvent des problématiques anciennes des anthropologues, des historiens du livre ou encore des sémioticiens.

- Les thésaurus laissent la place au traitement statistique de la langue, aux ontologies, aux « tags » ou aux moteurs. Au-delà de la performance des outils, le « Web sémantique » pose la question de la construction des savoirs dans la relation entre les métalangages et les documents eux-mêmes. Là encore les chercheurs en informatique et ingénierie des connaissances croisent des préoccupations anciennes, très anciennes même des chercheurs en Sciences de l'homme depuis la philosophie jusqu'aux Sciences de l'information, en passant par les différents courants des Sciences du langage.
- Les bibliothèques ou la documentation s'effacent devant les réseaux, les collections en ligne, les moteurs ou les dépôts. La notion de publication se transforme brutalement avec celle d'espace public. Celle de bibliothèque, dont les contours étaient limités, explose. Celle d'archives, autrefois réservée à un enregistrement pour une préservation à long terme, doit composer avec le « record management » ou encore les « archives ouvertes ». Le document numérique appartient maintenant à un espace régi par les lois des grands nombres : parce qu'il est mis en relation avec un nombre quasi-infini de ses semblables et parce qu'il est potentiellement visible par un nombre non-fini de lecteurs. Mais les problèmes économiques, juridiques, politiques, moraux qui lui sont attachés ne sont pour autant effacés. Ils se posent simplement différemment et avec une plus grande urgence.

Le succès du document numérique, on le sait, a été explosif et mondial au travers du Web. La mission des bibliothèques et de la documentation de donner accès à l'information est aujourd'hui remplie par une multitude de services directement accessibles en ligne. Mieux ou pire, la rentabilisation commerciale qui était réservée, ou presque, aux médias et aux industries culturelles (qui en France relèvent des sciences de la communication) s'étend aujourd'hui aux industries de l'information. Les industries, dites du *fair use* pèseraient selon une étude (Rogers & Szamosszegi, 2007) pas moins du 1/6^{ème} du PIB des Etats-Unis ! Les médias et l'édition scientifique transforment leur offre en proposant des collections patrimoniales à leurs lecteurs. Grâce une telle politique, l'éditeur Reed-Elsevier est devenu le premier éditeur mondial (Livre-Hebdo, 2007).

Le repositionnement des sciences de l'information

Je laisserai la parole, pour conclure, à des scientifiques qui, de part et d'autre de l'Atlantique, plaident pour un retour aux sources des sciences de l'information afin d'éclairer les mouvements actuels.

Le titre du premier article, *Epistemic Infrastructure in the Rise of the Knowledge Economy* (Hesdrom & King, 2005), résume son propos. Par « infrastructure épistémique », les auteurs comprennent les bibliothèques, les centres d'archives, les

musées, toute institution qui collecte et organise des éléments signifiants. L'argument de l'article vise à démontrer que ces institutions ne sont pas la conséquence de la montée des échanges d'informations dans les sociétés, mais bien la cause, et donc qu'elles participent directement au développement des civilisations et des économies. Voici un extrait de la conclusion (trad JMS) :

Les infrastructures épistémiques se sont développées sur les processus de sélection élaborés par les conservateurs, bibliothécaires et archivistes pour filtrer les connaissances selon les normes et standards professionnels, les sujets et les domaines, en étant attentif aux besoins des communautés d'utilisateurs. Cette sorte de collecte systématique construit la confiance dans les ressources informationnelles. Une économie du savoir, bâtie sur de l'information numérique dépendra, de la même façon, d'indicateurs clairs de qualité, d'autorité, d'authenticité. Les leçons de la construction des infrastructures épistémiques au cours du XIX^{ème} siècle et au début du XX^{ème} sont des exemples forts pour cette évolution. L'économie du savoir aura sûrement besoin d'outil nouveau pour s'épanouir, mais elle dispose déjà d'un fort potentiel dans les traditions muséales, bibliothéconomiques et archivistiques. (p.16)

Les auteurs de cet essai sont impliqués dans le développement des Information-Schools (I-Schools) aux Etats-Unis dont l'objectif est de redonner une place majeure dans le développement académique aux écoles en sciences de l'information. Par ailleurs, la thématique de leur propos rejoint les débats et les initiatives très nourries autour de ce que l'on appelle de ce côté de l'Atlantique les « cyberinfrastructures » (ACLS 2006, NSF 2003), c'est-à-dire les plateformes qui permettent la mise en relation de l'ensemble des données scientifiques pour l'ensemble des disciplines.

L'auteur du second papier (Ross, 2007), qui est en réalité une conférence, est européen. Il est impliqué dans les programmes sur les bibliothèques numériques et la conservation de l'Union européenne. Cet extrait de la conclusion de sa conférence résume clairement la proposition (trad JMS) :

Je suggèrerais, avec modestie, que les bibliothèques numériques adoptent une position théorique. Comme je l'ai suggéré plus haut la science des bibliothèques est dépourvue de fondements théoriques et des connaissances de base pertinentes pour le monde touffu du numérique. L'archivistique, avec ses principes d'unicité, de provenance, de classification et de description, d'authenticité, d'évaluation et ses outils comme la diplomatique pourrait nous donner un cadre pour une fondation théorique des bibliothèques numériques. (..) (p. 10)

Ainsi, de chaque côté de l'Atlantique, des voies s'élèvent pour redonner leur place aux sciences de l'information en s'appuyant sur leurs racines historiques. Même si le vocabulaire n'est pas le même, il s'agit bien d'éclairer le mouvement de redocumentarisation en cours. Il ne reste plus qu'aux chercheurs en sciences de l'information à relever le défi avec l'aide de leurs collègues des autres disciplines concernées, ou bien à ces derniers à s'approprier les savoirs des premiers..

Bibliographie :

ACLS (2006), *Our Cultural Commonwealth*, rapport de l'American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences, 43 p.

<http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf>

Barbier F. (dir) (2001), Les trois révolutions du livre, actes du colloque international de Lyon/ Villeurbanne, 1998, N° spécial de la Revue française d'histoire du livre, 2000, n° 106-109, Nouvelle série. - Genève : Droz, 343 p.

Chartier R. (1992), L'ordre des livres, lecteurs, auteurs, bibliothèques en Europe entre XIVe et XVIIIe siècle : Alinéa, 117 p.

Marshall A. (2006), Que faire du XXème siècle ?, Journée d'études de L'Association québécoise pour l'étude de l'imprimé, Montréal, 13 oct 2006.

Marshall A. (2003), Une brève histoire de la paperasse, papier de travail pour le RTP-DOC, 8 p.
http://rtp-doc.enssib.fr/IMG/pdf/marshall_contribution.pdf

NSF (2003), Revolutionizing Science and Engineering Through Cyberinfrastructure, rapport de la National Science Foundation, Janvier 2003, 84 p.
<http://www.nsf.gov/od/oci/reports/atkins.pdf>

Le Coadic Y.-F. (1994), La Science de l'information, Que sais-je ? : PUF, 127 p.

Livre-Hebdo (2007), *Le classement 2007 de l'édition mondiale*, Spécial Francfort, 9 p.
<http://www.livreshebdo.com/cache/upload/pdf/FrancFra.pdf>

Pédauque R. (2006), Le Document à la lumière du numérique : forme, texte, médium : comprendre le rôle du document numérique dans l'émergence d'une nouvelle modernité : C&F éditions, 218 p.

Rogers T. & Szamoszegi A. (2007), *Fair Use in the US Economy: Economic Contribution of Industries Relying on Fair Use*, Washington DC: Computer & Communications Industry Association, 84 p.
<http://www.ccianet.org/artmanager/uploads/1/FairUseStudy-Sep12.pdf>

Ross S. (2007), *Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries*, ECDL 2007, 10 p.
http://www.ecdl2007.org/sross_draft_ECDL_paper.pdf

Salaün J.-M. (1993), Les sciences de l'information en questions, le point de vue du lecteur, *Réseaux* n°58, mars-avril 1993, p.11-25.