HAL open science

Documents et numérique

Jean-Michel Salaün

To cite this version:

Jean-Michel Salaün. Documents et numérique. Contribution au rapport du Conseil d'Analyse Economique sur la Société de l'information., 2003. sic_00000610v1

HAL Id: sic_00000610

https://archivesic.ccsd.cnrs.fr/sic_00000610v1

Submitted on 29 Sep 2003 (v1), last revised 1 Dec 2003 (v2)

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.

Documents et numérique[1]

Jean-Michel Salaün
Ecole nationale supérieure des sciences de l'information et des bibliothèques (Enssib)
Contribution au rapport du Conseil d'Analyse Economique sur la Société de l'information.
29-09-2003

Résumé

Il peut être éclairant de regrouper les recherches en cours selon trois approches qui construisent progressivement un nouveau compromis entre une multiplicité d'acteurs pour, en réalité, ré-inventer des documents au travers du numérique : le signe (ou la forme), le texte (ou le contenu) et le médium (ou la relation).

Le document traditionnel repose sur un support, un texte et une légitimité. Une première phase de numérisation, celle dans laquelle nous sommes sans doute encore, a fait ressortir ses structures internes, l'importance des métadonnées pour le traitement et la puissance et les incertitudes du réseau. Une seconde phase, sans doute commencée mais dont l'aboutissement reste incertain, insiste sur le format XML qui intègre la structure mais que partiellement la forme, voudrait s'appuyer sur les ontologies pour retrouver et reconstruire les textes, et met en avant l'accès personnalisé.

Sommaire

Documents et numérique 1
Document comme signe ou forme 2
Support et structure 2
Formes éphémères 3
Document comme texte ou contenu 5
Langages documentaires et ontologies 5
Connaissances déroutées 6
Document comme médium ou relation 7
Sélection et réseau 8
Economie oubliée 10
Conclusion 11

Texte

Les documents sont omniprésents dans notre vie courante ; nous en emportons dans notre poche ou notre sac ; ils défilent constamment sous nos yeux ; ils accompagnent la production et la gestion des multiples biens et services qui font notre quotidien ; comme preuve ou

[1]Même si l'auteur de ce texte en assume l'entière responsabilité, il est redevable des réflexions, contributions, corrections de plus d'une cinquantaine de chercheurs francophones. Ce texte s'inspire en effet très largement d'un travail collectif réalisé au sein du Réseau thématique pluridisciplinaire 33 du département STIC du CNRS (http://rtp-doc.enssib.fr) et publié en ligne :
Document : forme, signe et médium, les re-formulations du numérique.
Roger T. Pédauque. Article. 08 juillet 2003. Working paper.
http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/05/11/index_fr.html
Une version anglaise est accessible à partir de la même Url.

renseignement, ils règlent nombre de nos relations sociales. Ils nous sont tellement familiers que nous ne les remarquons plus et nous ne ressentons évidemment pas le besoin de réfléchir à leur matérialité ou leur fonction.
Certains considèrent même la notion comme obsolète. Entrés dans la "société de l'information", nous nous débarrasserions des inutiles documents remplacés par une sorte d'énergie fluide : l'information. Il est vrai que du papier, support le plus courant, à l'électronique, les transformations se repèrent facilement, par exemple dans la matérialité, le traitement cognitif, la perception ou encore l'usage. Cette mutation profonde d'un élément essentiel de notre vie quotidienne, annoncée par les textes de quelques visionnaires, est toute récente, encore chaotique et sans doute sans retour. Elle procède d'une vigoureuse activité de recherche-développement, vite traduite en outils professionnels et domestiques par des industriels intéressés et des pionniers enthousiastes sans vraiment être vraiment pensée socialement, et encore moins politiquement maîtrisée.
Puisque le mouvement prime sur la réflexion, c'est lui qu'il faut observer, et tout particulièrement le mouvement initié par les chercheurs, très actifs dans ce domaine. Il peut être éclairant de regrouper les recherches en cours selon trois approches qui construisent progressivement un nouveau compromis entre une multiplicité d'acteurs pour, en réalité, ré-inventer des documents au travers du numérique : le signe (ou la forme), le texte (ou le contenu) et le médium (ou la relation). Ainsi un document ne serait finalement, de notre point de vue, qu'une transaction entre des hommes dont les modalités anthropologiques (lisibilité-perception, signe), intellectuelle (intelligibilité-assimilation, texte) et sociales (sociabilité-intégration, médium) fonderaient une part de leur humanité, de leur capacité à vivre ensemble. Autrement dit nous voudrions mesurer comment le paradigme technique "numérique" transforme ce contrat à trois facettes qui fonde la notion de document.

Document comme signe ou forme

Une première définition du document pourrait être représentée par l'équation : *Document traditionnel = support + inscription*, c'est à quelques variantes près la plus courante. Dans un premier temps en effet, l'accent est mis sur un support manipulable (au sens premier) sur lequel est fixée une trace interprétable selon sa forme par un oeil, une oreille humaine, ou par le toucher pour la lecture en braille et, pourquoi pas demain, d'autres sens, avec ou sans prothèse. Cette trace représente le contenu, matérialisé par une inscription. Le support traditionnel dominant (mais pas exclusif) est le papier, la trace l'écriture, manuscrite ou imprimée

Support et structure

La transformation des supports a été si souvent commentée qu'il est inutile d'y trop insister. Entre la valse des appareils, les réseaux, les "caches", etc. les documents numériques, aujourd'hui multimédias, basculent sans cesse d'un support à l'autre. Le lecteur doit utiliser un appareillage individuel sophistiqué pour accéder à la lecture. L'attachement permanent du document traditionnel à son support est ainsi rompu. On dématérialise (numérise) même couramment le document traditionnel par une opération de traitement d'images et de reconnaissance de formes. Notons que même si l'objectif est de reproduire une perception similaire ou homologue à celle de l'objet d'origine, il s'agit bien néanmoins d'une traduction nouvelle qui pourra occulter des éléments signifiants ou au contraire en faire découvrir ou redécouvrir de nouveaux, selon les choix technologiques effectués et l'utilisation ultérieure des fichiers.

Mais, dans notre équation (*support + inscription*) les conséquences du numérique sur le second terme, l'inscription, sont au moins aussi radicales. L'inscription relève du codage, une opération familière des informaticiens qui ont alors cherché à en isoler les éléments pour les modéliser, automatiser les opérations et réagencer les différents éléments ainsi perfectionnés. On peut représenter leur raisonnement par une seconde équation *document numérique = structure + données* qui ressemble comme une sœur à la notion de programme telle qu'elle est souvent présentée en informatique : *programme = logiciel + données*.
A partir du cœur, donc, du raisonnement informatique : l'élaboration d'algorithme, on peut alors reconstruire les documents, remontant pas à pas leur logique ou structure interne pour déboucher sur une représentation lisible à l'écran. Cette réflexion a débouché d'abord sur la bureautique, puis a rencontré les chercheurs développant des outils pour la publication électronique, et enfin a été confrontée à la nécessité d'échanger les documents à grande échelle. Elle a alors pris une dimension explosive avec la révolution du Web. Sans refaire l'histoire de tout ce cheminement, disons que, du point de vue qui nous occupe ici, le Web peut être décrit comme une infinité de documents reliés entre eux. Son architecture suppose que l'on puisse accéder à des documents de partout, sur n'importe quel type de matériel et selon la spécificité des groupes d'utilisateurs.
Les informaticiens ont fait ressortir ainsi deux niveaux fondamentaux de structuration des documents : leur structure logique (la construction d'un document en parties et sous-parties articulées entre elles), et leur représentation formelle de la présentation, les "styles" au sens informatique (pour le texte, par exemple, les choix typographiques).
Ils ont, dans le même temps, insisté sur le caractère essentiel de la normalisation afin qu'un document soit lisible sur tout type d'ordinateur et déchiffrable par des applications variées. Il est probable que le succès de plus en plus manifeste de la norme XML, et de ses nombreux dérivés particuliers, marque une nouvelle étape sinon un aboutissement de ces mouvements. En effet, cette norme, issue à la fois de l'informatisation des techniques éditoriales (SGML) et de la sophistication des premiers balisages du Web (HTML), intègre dans un même fichier structure et contenu par un balisage normalisé du texte, permettant de retrouver et de dépasser très largement la plasticité et complexité des feuilles reliées dont on avait perdu quelques fonctionnalités en route. D'une certaine façon, on pourrait dire que la stabilité visuelle maintenue par l'inscription sur un support est remplacée par l'attachement à une structure.

Formes éphémères

On soulève alors des questions inédites en renouvelant les termes de l'ancien contrat de lecture où le lien entre représentation perçue et structure logique était pérennisé par le support. Avec une approche à la XML, on capture la structure et le contenu ; la forme peut en être dérivée de différentes façons. Elle n'est pas représentée intrinsèquement.
Une évolution possible, déjà engagée mais dont la généralisation n'est pas certaine, serait que les documents ainsi « rédigés » rejoignent des bases de données, centralisées ou distribuées, et que l'ensemble des fichiers s'apparente de plus en plus à un ou plusieurs vastes jeux de « legos » où des briques de différentes tailles, formes et usages seraient agencées selon des configurations très variées. Un dernier pas serait ainsi en train de se franchir : un document n'aurait de forme à proprement parler qu'à deux moments : celui de sa conception par son auteur qui devra le visualiser ou l'entendre, pour s'assurer qu'il correspond à ses choix (et encore ce n'est pas obligatoire si sa production relève du processus) et celui de sa re-construction par un lecteur. Il est peu probable que le document sera toujours identique dans l'un et l'autre cas. Une autre façon de concevoir cette évolution serait de considérer que le document est maintenant la base de

données elle-même dont les différentes sorties ne seraient qu'une interprétation partielle de la richesse. Nous percevons très clairement les prémisses de cette dernière étape ; nous entrevoyons aussi bien des problèmes qu'elle soulève ; il est beaucoup plus aventureux d'en prévoir l'évolution et donc les conséquences, sinon pour dire qu'elles seront à coup sûr très importantes et durables. Néanmoins, donnons quelques pistes qui représentent autant de recherches en cours ou à venir.
Dès lors qu'un document n'est plus figé par un support, il faut gérer les temporalités diverses d'un et de plusieurs documents et de son écriture, de son enrichissement ou de sa ré-écriture par des intervenants variés. Il s'agit d'inventer les procédures permettant de rattacher un texte à un auteur (ou à un collectif d´auteurs), tout en permettant à chacun de s´approprier - de se ré-approprier - tout ou partie de documents produits par d´autres ou par eux-mêmes afin de limiter les proliférations «bruyantes».
Par ailleurs, il y a du sens dans la forme et il n'est donc pas anodin de séparer structure et données, encore moins de traiter à part les modes de représentation. Pour ne prendre qu'une illustration, ces séparations détruisent les bases de la diplomatique archivistique dont l'un des buts est d'authentifier le fond du document par l'analyse de sa forme, il en découle que l'authentification (la validation) doit (devra) être assurée par d'autres moyens, de type technique (filigrane électronique) ou organisationnel (tiers archiveurs certifiés). Pourrait -on imaginer que les exigences de forme - et entre autre la forme authentique - soient imposées et/ou validées par l'existence d'une feuille de style "signée" ou "filigranée" ?
Ces questions sont d'autant plus délicates qu'un même document peut se lire couramment sur différents appareils de lecture. Ainsi doit-on raisonner comme si les terminaux n'avaient pas d'influence sur la perception ? Il suffit pour se convaincre du contraire de comparer les écrans d'un micro-ordinateur, d'une tablette, d'un agenda électronique ou d'un téléphone portable. On retrouve alors le poids du support de lecture dont on avait cru pouvoir s'affranchir. Ces questions font l'objet de débats nourris, notamment au sein du consortium W3C (*device independence*).
Une autre série de questions concerne la pérennité des documents numériques. Les fichiers XML, pour peu qu'ils soient régulièrement rafraîchis et conservés dans de bonnes conditions, sont en théorie inaltérables puisqu'ils contiennent la totalité de leurs informations sous forme numérique. Ainsi certains considèrent que, sous peu, les problèmes de pérennité seront résolus. Mais, ces fichiers ne représentent pas, loin s'en faut, la (ou les) forme(s) sous lesquelles sont lus les documents. Sans même parler d'une mémoire du Web, dynamique et éphémère par construction, une mémoire complète de chaque document supposerait de conserver la totalité des matériels et systèmes de lecture successifs qui permettent d'y accéder. Là encore, un gros travail théorique et pratique reste à mener.
Enfin, sans prétendre épuiser les problèmes, relevons une dernière série de questions. Le document traditionnel est un objet matériel manipulable. Cet objet s'efface dans le numérique, jusqu'à, dans l'avancée ultime, n'être même plus qu'une sorte de puzzle dont les morceaux sont agencés à la demande du lecteur. Ne peut-on faire l'hypothèse que la stabilité visuelle du papier, la maniabilité et la co-existence des feuilles jouent un rôle important dans la cognition ? Quel impact les nouveaux régimes de lecture vont-ils avoir sur nos régimes de savoir ? L'élaboration d'un document peut-elle se détacher de sa forme perceptible et donc, est-il simplement concevable d'envisager une rupture formelle entre l'élaboration par l'auteur (qui est aussi le premier lecteur) et la proposition faite aux lecteurs ? Le succès des formats de "fac-similé" (PDF), est souvent analysé comme une résistance momentanée au changement. Ne s'agit-il pas plutôt d'un attachement à une stabilité perceptive ?
On pourrait résumer ces questions par une seule : en effaçant le support, n'a-t-on pas trop délaissé justement la forme ?

Document comme texte ou contenu

La définition du document traditionnel selon cette dimension pourrait être symbolisée par l'équation : *Document = inscription + sens*. Ici, le support est accessoire, y compris pour le document traditionnel du moment que l'inscription est préservée. L'important est le contenu, matérialisé par cette dernière, qui est porteur de sens. Le sens, lui-même se construit par rapport au contexte de production et de diffusion du document qui conditionne l'interprétation du contenu. Pour bien comprendre l'évolution en cours selon ce point de vue, il faut avoir en tête le raisonnement qui le fonde.

"Penser c'est classer", en réalisant des documents nous isolons et rangeons des discours pour nous aider à penser le monde. La mise en document passe nécessairement par un système qui permettra de placer le document dans un ensemble et de l'y retrouver, une indexation au sens strict ou figuré, et donc des systèmes de classification concrets ou abstraits. D'autre part, un document n'a de sens que s'il est lu ou interprété par un lecteur. Cette interprétation dépend largement du contexte dans lequel elle est pratiquée. Un même document pourra prendre des sens différents, voire opposés, selon l'époque et la situation sociale ou individuelle de l'interprétant. Ainsi un document est pris dans une double relation, relation au monde documentaire (classement) et relation au monde naturel (interprétation). Ces relations se réalisent grâce à un "horizon d'attente", un ensemble de signes familiers qui construit le contrat de lecture entre le lecteur et le document lui permettant d'en décrypter sans difficulté le sens, car il sera placé d'emblée dans son contexte d'interprétation.

Enfin, la structure de la langue écrite depuis la lettre d'alphabet dans les langues indo-européennes jusqu'au discours organise la plupart des documents. Ceux-ci sont fait de morceaux discrets, plus ou moins isolables et réagencables, analysables, soumis à des règles de syntaxe, de mise en discours et de style. Cette utilisation de la langue naturelle confère aux documents une très grande plasticité, qui va servir de base aux développements d'outils pour leur gestion.

Langages documentaires et ontologies

L'explosion documentaire, c'est-à-dire la brutale augmentation du nombre de documents manifeste dès la fin du 19ème siècle et sans rémission depuis, a conduit en effet à l'invention, difficile et laborieuse, par les documentalistes de ce qu'on a appelé les "langages documentaires" (références bibliographiques, index, thésaurus, résumés etc.), organisés de façon associative ou hiérarchique, qui sont directement issus du raisonnement exposé ci-dessus : il était possible en effet de construire à partir des textes des documents (ou des images, ou des objets eux-même) un langage artificiel ou formel permettant de les classer pour les retrouver à la demande. Dans un premier temps, le numérique a été utilisé par les documentalistes simplement comme un outil performant de classement des items des langages documentaires sous forme de bases de données bibliographiques. Rapidement le traitement informatique de la langue naturelle puis la production et la gestion directe de documents électroniques, le succès du Web et enfin la modélisation du raisonnement ont changé la donne.

Comme dans l'approche précédente, les informaticiens ont cherché à isoler les éléments logiques pour les modéliser. Mais ici, ils se sont attaqués directement au contenu. Soit pour de l'indexation automatique, soit pour des résumés, soit encore pour des systèmes de question-réponse, les informaticiens et les linguistes ont réuni leurs compétences en utilisant des outils statistiques et morphosyntaxiques. A leur manière, ils ont suivi un chemin parallèle à celui des documentalistes, reconstruisant à l'aide de filtres et de calculs, sinon un langage au sens informatique, du moins du

texte censé représenter d'une façon structurée le contenu du document et autorisant ainsi un traitement automatique par les machines. Les résultats furent, dans un premier temps, moins probants que ne le pensaient les zélateurs. Bien des questions sur la performance des outils restent encore non résolues. Du côté des langues, par exemple, on peut s'interroger sur l'application des outils aux langues dont la structure et l'écriture divergent des langues indo-européennes. Les meilleurs outils ont dû intégrer une part de travail humain, se présentant plus comme des outils d'aide que comme des outils automatiques. Néanmoins, leur efficacité est spectaculaire, pour l'internaute sinon pour l'initié, dans leur application au Web sous forme de moteurs.

L'arrivée annoncée du Web sémantique peut être comprise à la fois dans la continuité de ces résultats et comme, sinon une rupture, du moins un saut méthodologique. Pour la première interprétation, on notera, par exemple, la structure toujours plus formalisée des documents (XML) et l'insistance sur l'indexation (RDF). Il s'agit de ce point de vue de construire une bibliothèque distribuée multimédia à l'échelle du réseau des réseaux, intégrant des outils de recherche plus performants. Mais l'ambition est aussi plus large. L'objectif est de passer d'un Web constitué simplement d'un ensemble de fichiers reliés entre eux à un réseau utilisant pleinement les capacités de calcul des machines reliées, notamment sur les traitements sémantiques des textes. A cette fin l'utilisation de "métadonnées" que l'on peut modéliser et combiner est essentielle. Ainsi, à leur manière, les promoteurs du Web sémantique construisent à leur tour des sortes de langages documentaires, qu'ils ont baptisé "ontologies".

La rencontre des promoteurs du Web sémantique avec les chercheurs de l'ingénierie des connaissances, dont l'objectif est la modélisation des opérations de pensée et de gestion des savoirs, était alors inévitable. Ces derniers intègrent, notamment, la question des statuts des documents, de la modélisation des connaissances et surtout celle des ontologies. Les ontologies se focalisent sur l'essence d'un domaine (comme la médecine, ou une spécialité de la médecine par exemple), sur son vocabulaire et, au-delà, sur le sens dont il est porteur. Ce sens a deux facettes, celui compris par l'être humain et c'est la sémantique discursive et cognitive et celui « compris » par la machine et c'est la sémantique formelle de l'ontologie. Une des principales richesses des ontologies est donc leur formalisation qui permet leur usage par un programme informatique là où un thésaurus est en échec.

Ainsi, comme pour l'entrée précédente mais d'une façon sans doute moins avancée, nous sommes peut être au seuil d'une nouvelle étape pour le document numérique grâce aux apports du Web sémantique. Nous pourrions représenter cette étape par la transformation de l'équation en *Document WS = texte informé + ontologies.*

Connaissances déroutées

Le changement d'échelle introduit pourtant une rupture qui pourrait avoir d'importantes conséquences. Tant que les langages documentaires étaient affaire de spécialistes pour l'amélioration de services, limités même s'ils étaient essentiels et donc maîtrisables, comme ceux des bibliothèques et des centres de documentation, les enjeux cognitifs et économiques restaient confinés. A partir du moment où l'on raisonne à une échelle beaucoup plus large, inédite, la construction des outils d'ingénierie des connaissances, en particulier les ontologies ne sera pas sans conséquences lourdes sur notre rapport au savoir et il pourrait bien, de plus, se consolider une économie des métadonnées dont la position permet l'accès aux documents. A l'inverse des débats menés à la fin des années soixante-dix sur "l'industrie de l'information", il ne semble pas que l'on ait encore aujourd'hui pris vraiment la mesure de ces enjeux.

D'autre part, l'utilisation de plus en plus poussée de la langue naturelle comme d'un outil a tendance à souligner l'aspect communicationnel des textes en estompant la cohérence globale des documents. Ainsi pour ne prendre que l'exemple littéraire, plusieurs chercheurs ou essayistes ont vu dans l'avènement du Web, et plus précisément dans les techniques d'hypertextualité ou d'hypermédia, un effacement des documents. Ainsi la triade classique auteur-oeuvre-lecteur, à l'origine de la construction du document littéraire, pourrait laisser place à un processus interactif où les liens entre les pages accessibles joueraient un rôle plus important que le texte tel qu'il était auparavant construit par l'auteur. Néanmoins, même si d'intéressantes expériences d'écriture hypertextuelle ont été et sont toujours conduites avec des conséquences sémantiques et cognitives non négligeables, il apparaît que le développement explosif du Web a conduit au contraire à une multiplication exponentielle de documents mis en ligne. Les liens entre les pages paraissent se structurer progressivement pour construire de nouvelles normes de paratexte, renforçant au contraire l'aspect documentaire du Web.
Pour le dire autrement, par l'hypertextualité et les moteurs nous dérangeons l'ordonnancement traditionnel des documents, sans avoir vraiment encore reconstruit un classement que nous maîtriserions. Pourtant le contenu n'a de valeur que par rapport à un contexte. Comment rendre compte du statut du texte, des relations avec les ensembles auxquels il se réfère, de la confiance que le lecteur peut lui accorder, de l'articulation avec les informations orales, de sa globalité de sens, pour ne prendre que ces exemples, sans référence à la notion de document ?
Néanmoins l'augmentation du nombre de documents accessibles dans une forme ne comprenant pas de métadonnée est beaucoup plus importante que celle de documents "indexés". Pire, la concurrence sur le Web conduit à des stratégies opportunistes d'indexation, visant à tromper les moteurs de recherche. Ainsi, il est vraisemblable que, au moins dans un premier temps, deux dynamiques parallèles co-existent. D'un côté, pour les communautés auto-régulées qui ont intérêt à développer une recherche documentaire performante (experts, entreprise, médias, etc.), des "langages métiers" seront appliqués aux documents le plus en amont possible de leur fabrication, vraisemblablement d'une façon manuelle-assistée. D'un autre côté, des méta-langages automatiques plus légers, éventuellement adaptés aux comportements de recherche de grandes catégories, continueront à se perfectionner pour les outils ouverts largement aux internautes.
La question fondamentale pourrait être résumée ainsi : comment en ajoutant de "l'intelligence" aux outils de recherche documentaire modifions nous notre rapport aux savoirs et à la culture ?

Document comme médium ou relation

L'aspect communicationnel des documents est bien la troisième et dernière dimension qu'il nous reste à aborder. Un document donne en effet un statut à une information, à un signe matérialisé. Il est porté par un groupe social qui le suscite, le diffuse, le sauvegarde et l'utilise. C'est une preuve qui fait foi d'un état des choses et c'est une annonce qui prévient d'un évènement. C'est aussi un discours dont la signature le rattache à un auteur. C'est un témoignage, volontaire ou non. C'est une pièce de dossier. Le statut de document n'est pas acquis pour l'éternité, il se donne et il peut aussi se perdre définitivement dans l'oubli collectif ou encore se retrouver si quelqu'un redécouvre et re-légitimise un document disparu de la conscience collective, mais non détruit.
Un document n'est pas nécessairement publié. Bien des documents, parce qu'ils règlent des questions privées (dossier médical, transaction entre des personnes), ou parce qu'ils contiennent des secrets non-divulgables, ne sont consultables que par un nombre très limité de personnes. Néanmoins, ils ont un caractère social dans le sens où ils sont rédigés selon des règles établies qui

fondent leur légitimité, qu'ils sont utilisés dans des relations formelles, et que, en cas de disfonctionnement, ils auront valeur de référence.
Inversement, la publication, large ou limitée, constitue souvent un moyen simple de légitimation. En effet, une fois rendu public, c'est-à-dire consultable par un nombre conséquent de personnes, un texte fait partie du patrimoine commun. Entré dans une mémoire partagée, il ne peut plus être modifié sans difficulté, sa valeur est appréciée collectivement. Mais la publication ouvre aussi la voie à une résonance entre lecteurs, qui échangent leurs avis, et entre documents, qui se relient les uns les autres. Ainsi par un effet statistique la notoriété et donc la valeur marchande, directe ou indirecte, se concentre sur quelques éléments au détriment des autres.
Les documents sont utilisés pour la régulation des sociétés humaines en assurant une communication et une pérennisation de la norme et des connaissances nécessaires à leur survie ou leur continuité. D'une certaine façon, nous pourrions dire que le contrat de lecture trouve ici sa troisième dimension : la sociabilité, l'appropriation par laquelle le lecteur en prenant connaissance d'un document marque sa participation à une société humaine ou, inversement, l'inscription sur un artefact d'une représentation du monde naturel et son insertion dans un patrimoine collectif.
La multiplication des documents est donc liée à l'évolution des sociétés qui en font un usage intensif. L'organisation politique et sociale s'appuie sur la production et l'échange de documents. Les religions et leurs clercs, les Etats et les administrations, les organisations productives et le commerce, la société civile dans leurs différentes composantes, leur évolution historique, leurs géographies et cultures propres, leurs fonctions changeantes se sont servies et se servent encore largement des documents pour leur régulation interne comme pour l'affirmation concurrentielle de leur identité et position. Des métiers se sont spécialisés sur tel ou tel moment du processus de mise en document ou tel ou tel domaine d'application. Des systèmes se sont construits et formalisés pour répondre à la régularité de la production. Des entrepreneurs, petits ou gros, se sont lancés dans l'aventure ou des organisations l'ont prise en charge. Ces dispositifs ont un coût de mise en place, d'entretien et une inertie, donc une économie. L'importance sociale des documents peut aussi être mesurée au juridisme qui les accompagne dans les sociétés organisées, ou simplement aux nombreux rituels qui les entourent. La majorité des documents sont soumis à des règles précises pointilleuses sur leur présentation, leur rédaction, leur contenu, leur mode de conservation, de gestion, de divulgation. Même les plus libres en apparence sont attachés par un droit moral (inaliénable !) à leur auteur. L'économie ouverte des documents a développé des modes de régulation bien particuliers, comme le droit d'auteur (droit de l'œuvre) ou le copyright (droit du document).

Sélection et réseau

Les documents prolifèrent aussi bien dans les sociétés fermées que dans les sociétés ouvertes pour des usages strictement utilitaires ou au contraire d'enrichissement culturel et, chaque fois, l'intégration du numérique se traduit par des mouvements contradictoires dont l'interprétation n'est pas aisée.
Dans les organisations, le premier constat pourrait être celui de l'effacement, difficilement mesurable, car il s'est effectué en ordre dispersé non sans contradiction, d'un nombre important de documents qui, tenus sous une forme traditionnelle, rendaient compte des procédures. De plus en plus souvent, coïncidant avec la disparition de cadres intermédiaires, le remplacement est total : formulaires, tableaux fiches, pilotages, modes d'emploi qui faisaient les beaux jours de la bureaucratie publique ou privée sont remplacés par des bases de données et de l'échange de

données informatisées. Mais, il se constate conjointement une montée d'une mise en écrit et en documents, amplifiée notamment de façon exponentielle par la démarche qualité ou par le travail collaboratif. Dans cette perspective, moins que jamais, le document ne ferait sens isolément, mais serait constitué par l'enregistrement informatique de transactions préalablement définies. L'affichage des informations, éphémère et nécessairement dépendant de technologies évolutives, ne constituerait pas à lui seul le document ; il devrait être validé par des procédures certifiées. Si l'on ajoute les progrès de la signature électronique, bien des transactions pourraient à l'avenir se réaliser sans les formalismes adoptés pour le papier.

Du côté des médias traditionnels, alors que le numérique envahit la fabrication, la recherche de modèles économiques viables sur le Web n'a pas abouti et ce dernier n'est qu'un outil de promotion ou de communication avec les lecteurs ou spectateurs. Seuls quelques secteurs, qui avaient déjà des affinités avec les réseaux ont trouvé des modalités de financement : l'information financière et l'information scientifique, non sans contestation pour la seconde. Il est possible aussi que la musique à la suite du mouvement d'échanges entre internautes puisse redéfinir son mode de distribution et de valorisation, mais la mutation sera radicale et douloureuse pour les firmes traditionnelles.

Le Web, conforme aux orientations des concepteurs de l'Internet, s'appuie en effet sur un réseau de communication de plusieurs à plusieurs où chaque pôle, grand ou petit, doit disposer des mêmes outils et être à la fois producteur et consommateur. Le Web n'est pas une simple technique, il suppose, en fait, une conception sociale, ou plutôt une conception de la communication sociale, proche de la "République des sciences", des "communautés virtuelles" ou du mouvement des logiciels libres. Dans une telle société, chaque personne est acteur et responsable devant la communauté de ses actes. Si l'on traduit cela dans notre domaine, nous dirons que chacun est capable de lire ou d'écrire des documents qui concernent la vie collective et chacun aura à coeur de ne rendre public que des documents qui enrichissent la collectivité. Cette idée est très présente dans nombre de discours et d'initiatives du domaine à commencer par ceux du consortium W3C.

L'industrie du contenant (industrie des logiciels et des télécommunications), non sans débats, batailles et compromis, est très attentive à ces développements qui confortent ses positions puisqu'ils favorisent l'augmentation du trafic et des traitements au détriment d'une industrie du contenu fondée au contraire sur une sélection et un filtrage *a priori*. Ainsi cette dernière, court-circuitée par le réseau s'appuie sur le droit de la propriété intellectuelle, en multipliant menaces et procès, pour défendre le capital d'œuvres qu'elle construit et thésaurise dans sa relation avec les auteurs et créateurs. D'une certaine façon, on pourrait dire que deux modes de régulation sont en concurrence : la régulation par la technique et la régulation par le droit.

Il y a donc malentendu, volontaire ou innocent, au départ. Ce malentendu est néanmoins d'une grande fertilité, car il permet aux collectivités dont la communication est bridée par le système traditionnel de trouver un espace pour échanger. Il donne aussi à nombre d'institutions, à commencer par celles d'intérêt général, un outil simple pour communiquer avec la population. Ainsi, le Web est un vaste bazar où l'on trouve une multitude de documents consultables gratuitement pour le lecteur et reliés entre eux.

Certains considèrent qu'il ne s'agit que d'une organisation provisoire, illustrant la jeunesse du média. Cette analyse ne prend peut-être pas assez la mesure de la rupture que produit le Web. Il est possible que le filtrage ou la sélection ne se fasse plus dans le Web *a priori* comme dans les médias traditionnels, mais *a posteriori*. La résonance jouerait alors comme un système de "percolation" où les documents les plus pertinents seraient progressivement repérés et mis en valeur, par le nombre de liens et par le jeu des moteurs. L'important alors serait la toile elle-même

qui, par son mouvement continuel (les liens qui se font et se défont, les moteurs qui tournent et "aspirent", les pages qui apparaissent et disparaissent), autoriserait un repérage des documents. L'implication d'un nombre conséquent d'internautes, jusqu'ici à l'écart du petit monde fermé des médias, et le succès manifeste du média dans les pratiques donnent consistance à cette hypothèse en procurant une dimension et une rapidité inédite à la dynamique ordinaire de la légitimité par la notoriété.
Les développements annoncés du Web sémantique laissent présager d'autres développements, notamment dans les relations entre document et service. L'orientation générale est à la personnalisation et au service sur mesure. Ainsi chacun verrait s'afficher sur son terminal le document répondant à ses questions, besoins ou envies du moment. Mais, même si certains services et formulaires préfigurent cette orientation, les applications restent encore des démonstrations en laboratoire.

Economie oubliée

Mais ces avancées laissent néanmoins dans l'ombre le problème du financement des contenus. La loi du moindre effort appliquée à l'accessibilité du Web fait que l'internaute préfèrera éviter tous les obstacles et barrières à sa navigation plutôt que les affronter. Ainsi, il contournera toute demande de financement direct. Dans la même dynamique, un mouvement militant fait de la gratuité une sorte de qualité naturelle du Web ou plutôt le Web serait l'occasion d'une accessibilité au savoir et à la culture libérée des coûts de diffusion et des contraintes commerciales. Le Web serait une sorte de bibliothèque universelle.
Il se pose alors la question de la limitation de l'attention, sans laquelle un document ne saurait avoir de lecteur. L'attention humaine est limitée, par le temps disponible, par la fatigue du lecteur ou par les compétences techniques ou intellectuelles dont il dispose. Cette problématique est bien connue des radiodiffuseurs et des documentalistes. L'internaute étant nécessairement actif, il ne saurait être "capté" comme peut l'être l'auditeur de la radio ou le téléspectateur. Autrement dit le Web marie la liberté de choix de l'édition avec l'accessibilité de la radiodiffusion, ou élargit les services de bibliothèque à toute la planète pour la collection et le domicile pour la consultation. Mais les lois bibliométriques et les effets de résonance risquent de jouer à une échelle inédite par secteurs : l'attention se concentrant très fortement sur un nombre réduit de documents et se dispersant sur un très grand nombre. Du côté du public, les services pourraient ne profiter, voire être confisqués, par une minorité d'internautes, plus malins, disponibles, compétents ou puissants. Ces phénomènes et leurs conséquences sont connus à une échelle réduite dans le monde des bibliothèques traditionnelles et encore bien peu étudiés dans le Web.
Opportunisme et politique se conjuguent pour que, progressivement, l'économie du contenu sur le Web se configure comme une économie du don et du troc, et sans doute à terme un marché institutionnel, du "B2B". Est-on vraiment sûr que cette structure de financement est viable dans un réseau ouvert et même qu'elle garantit à moyen terme la diversité et la pluralité des documents mis en ligne ? Est-on sûr simplement qu'elle contient, passé l'enthousiasme des pionniers, les ressources suffisantes, secteur par secteur, pour alimenter la production et gestion de documents ?
Dès aujourd'hui se pose le problème de la transition entre une époque fondée exclusivement sur le papier et celle où le numérique devient incontournable. L'objectif de traduire l'ensemble des documents existants d'un support traditionnel à un état numérique est hors de portée. Le développement explosif de l'informatique oblige néanmoins à envisager des traitements de très grande ampleur, sauf à s'accommoder d'une amnésie radicale de notre culture documentaire. Il faudra donc bien dans un avenir proche être en mesure de faire des choix raisonnés (que

numériser en priorité ?) et de construire des outils capables de traiter à un coût raisonnable de grandes masses de documents. Là encore, même si de nombreuses initiatives sont lancées, il s'agit de mesures ponctuelles et dispersées. Pourtant les déséquilibres de contenus sur le Web se mesurent à une échelle planétaire.
Enfin et sans prétendre une nouvelle fois à l'exhaustivité, relevons que le problème se pose, au jour le jour dans chaque organisation et chaque collectivité, avec le renouvellement de la notion d'archive dont la base est l'enregistrement et la conservation de documents. Le rôle de l'archive est de garder la mémoire d'une activité humaine. Un rôle nouveau, plus actif, émerge pour les archives avec le numérique. Archives ouvertes, récupération des programmes audio-visuels à la source ou des émissions de télévision diffusées, archivage du Web, bien des activités inédites se développent, tandis que les pratiques de l'archivistique se renouvellent, la mise en place du *records management* dans les organisations apparaissant comme une condition pour un bon archivage électronique. On y sent aussi bien des questions encore sans réponse définitive sur un rôle différent à assumer : hésitation entre le témoignage d'une action passée et l'enregistrement d'une action en cours ; confusion entre l'archivage et la publication ; simple enregistrement ou préparation d'une utilisation à venir. Et encore, comment garder la mémoire d'un mouvement continu de renouvellement de pages reliées entre elles ?
Toutes ces questions peuvent se résumer en une seule : quels choix de financement et quel équilibre trouver pour une économie viable du contenu numérique ?

Conclusion

L'Internet a été conçu d'abord comme un outil de communication de plusieurs à plusieurs, le succès foudroyant de la messagerie montre qu'il y avait là une attente massive. Le Web a hérité de cette conception première, mais il s'agit d'un outil documentaire : publication, stockage, classement, recherche, autrement dit les activités ordinaires des métiers du document, sont ses fonctions. Il a naturellement rencontré sur sa route d'autres outils numériques, nés des développements du document : bureautique, publication électronique, numérisation, recherche documentaire, lecture assistée etc.
Le numérique a aussi croisé le papier, et l'opposition souvent soulignée entre l'un et l'autre parait aujourd'hui vaine. Quasiment tous les documents d'aujourd'hui ont été à un moment de leur vie sous un format numérique, et ceux qui échappent à cette règle risquent de tomber dans l'oubli. Inversement, de très nombreux documents numériques sont à un moment ou un autre imprimés, sur une imprimante individuelle ou dans une imprimerie professionnelle. Ainsi, l'important est bien de mieux cerner la notion de document en général, dont le numérique est à la fois un révélateur et un facteur d'évolution.
Les trois entrées que nous avons développées successivement permettent un premier décryptage du mouvement en cours. Remarquons sans surprise que nous avons repéré dans chacune d'elles des étapes dans l'histoire des documents numériques que nous pouvons maintenant mettre en parallèle. Le document traditionnel repose sur un support, un texte et une légitimité. Une première phase de numérisation, celle dans laquelle nous sommes sans doute encore, a fait ressortir ses structures internes, l'importance des métadonnées pour le traitement et la puissance et les incertitudes du réseau. Une seconde phase, sans doute commencée mais dont l'aboutissement reste incertain, insiste sur le format XML qui intègre la structure mais que partiellement la forme, voudrait s'appuyer sur les ontologies pour retrouver et reconstruire les

textes, et met en avant l'accès personnalisé. Il y a dans cette évolution générale un sens dont il faudrait mieux comprendre l'orientation, les conséquences et les limites.
Dans chaque entrée nous avons insisté sur l'idée de contrat de lecture traduit par la lisibilité dans la première, par la compréhension dans la seconde et par la sociabilité pour la troisième. Il est probable que ce contrat à trois facettes présente, dans toutes les nuances que nous avons exposées, la réalité de la notion de document. Dans cette perspective, le numérique n'est qu'une modalité de multiplication et d'évolution de ces contrats. Mais l'importance qu'il a prise, sa performance, la rapidité de sa diffusion, et en même temps les questions inédites et les risques qui s'ouvrent, rendent d'autant plus nécessaire une fine et juste analyse.