



**HAL**  
open science

# Document : forme, signe et relation, les re-formulations du numérique

Roger T. Pédauque

► **To cite this version:**

Roger T. Pédauque. Document : forme, signe et relation, les re-formulations du numérique. 2003.  
sic\_00000413

**HAL Id: sic\_00000413**

**[https://archivesic.ccsd.cnrs.fr/sic\\_00000413v1](https://archivesic.ccsd.cnrs.fr/sic_00000413v1)**

Preprint submitted on 28 Apr 2003

**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.

# Document : forme, signe et relation, les re-formulations du numérique

Document de travail - version provisoire  
Roger T. Pédaque, STIC-CNRS  
Contacts pedauque@enssib.fr

Version 1 - 28-04-2003

## Résumé

Ce texte présente un travail collectif de réflexion en cours au sein du réseau thématique pluridisciplinaire 33 du département STIC du CNRS. Il se propose de préciser la notion de document dans son passage au numérique à partir de recherches qui privilégient plutôt la forme (comme un objet matériel ou immatériel), le signe (comme un porteur de sens) ou la relation (comme un vecteur de communication).

Ces entrées montrent chacune que des transformations radicales sont en cours. Leur superposition souligne l'importance de la pluridisciplinarité pour une analyse lucide et complète de la notion et de son évolution.

## 1. Contexte

2. Très peu d'articles scientifiques proposent une définition du document, encore moins la discute. Le document se repère directement comme objet d'analyse dans deux communautés scientifiques : les chercheurs des sciences de l'information (ou Information science), issus des travaux concernant les techniques documentaires largement renouvelés par le traitement informatique, et des chercheurs s'intéressant à la numérisation qui ont souvent élargi leurs réflexions à la gestion électronique de documents. Par ailleurs, parce qu'il est un outil indispensable à la construction et l'avancée de la discipline, on le discute en histoire et tout particulièrement en archéologie, ou en géographie spécialement concernant les cartes, ou encore en droit pour les textes et articles de lois, règlements ou circulaires, mais sous un angle instrumental et rarement directement.
3. Nombre de dictionnaires, de répertoires de normes, d'encyclopédies présentent des définitions, qui relèvent plus de la désignation ou de la description que d'une réflexion approfondie sur la notion.
4. Pourtant, on en trouve trace dans un très grand nombre d'autres travaux de recherche qui utilisent un vocabulaire différent, parfois rigoureusement défini, souvent sujet à des interprétations diverses. Ainsi les chercheurs en réseau, en base de données, en fouille de textes, en recherche d'informations, en traitement automatique de la langue, en ingénierie des connaissances ou encore les linguistes de corpus, les sémiologues, les psychologues de l'apprentissage, les sociologues de la culture ou de l'organisation, les économistes des médias ou de l'information, les juristes de la propriété intellectuelle et, d'une façon générale "les humanités" usent de vocables divers comme information, donnée, ressource, fichier, écrit, texte, image, papier, article, oeuvre, livre, journal, feuille, page, etc. qui, bien entendu, ne sont pas synonymes, qui ont chaque fois une justification dans le contexte particulier de la recherche concernée, mais qui, tous, ont un rapport (la plupart du temps non assumé) avec la notion de document.

5. Enfin, les documents sont omniprésents dans notre vie courante (même, et peut être surtout, dans l'activité scientifique). Ainsi la notion est intuitive pour chacun d'entre nous sans que nous ressentions le besoin de la préciser.
6. Ce flou fait aujourd'hui problème. En effet le numérique bouscule profondément la notion de document sans que l'on puisse clairement en mesurer les effets et les conséquences faute d'en avoir au préalable cerné les contours. Cette remise en cause, même si elle fut annoncée par les textes de quelques pionniers et préparée par la convergence de plus en plus manifeste entre l'écrit et l'audiovisuel est toute récente, encore chaotique et sans doute sans retour. Il est probable que les nombreux chercheurs qui abordent ces questions sous de multiples facettes gagneraient à une vue d'ensemble leur permettant de se positionner plus lucidement.

## **7. Rupture et continuité**

8. Le contraste entre la relative stabilité qui prévalait jusqu'à présent et la vitesse et radicalité des bouleversements d'aujourd'hui explique sans doute le retard de l'analyse. On n'avait pas besoin de s'interroger, sinon comme historien, sur un objet trop courant pour ne pas être évident, et aujourd'hui on n'a pas vraiment eu le temps de prendre du recul.
9. Le document a été construit comme un objet, dont la concrétisation la plus banale est la feuille de papier, au cours d'un processus séculaire où se sont entrelacés outils, savoirs et statuts. Depuis quelques dizaines d'années avec le numérique, nous sommes entrés dans une phase nouvelle dont certaines caractéristiques sont en filiation directe avec la période précédente, tandis que d'autres marquent au contraire un changement radical et peut-être l'émergence d'une notion différente reprenant tout ou partie de l'utilité sociale de ce que nous appelions "document".
10. Le questionnement entre rupture et continuité ne se pose pas seulement sur l'objet. Les méthodes d'analyses ou les épistémologies sont-elles aussi en évolution rapide.

## **11. Une réflexion pluridisciplinaire**

12. Notre sentiment est que ces difficultés ne peuvent être levées que par une réflexion résolument pluridisciplinaire. Nous avons été conforté dans cette opinion par le département STIC du CNRS qui a lancé un réseau thématique pluridisciplinaire, baptisé "Document et contenus : création, indexation, navigation" (<http://rtp-doc.enssib.fr>) qui regroupe une centaine de chercheurs venus d'horizons divers. Pour la plupart des disciplines auxquelles ils appartiennent, le document n'est pas une notion centrale et les chercheurs n'en ont qu'une appréhension partielle. L'objectif du réseau est donc de déplacer cette attention oblique pour faire du document un objet principal de recherche, au moins pour un temps, en croisant les apports partiels des uns et des autres.
13. Il n'est pas sûr qu'il existe, entre les disciplines ou même à l'intérieur de chacune d'elles, un consensus sur les questions que nous abordons. Notre objectif n'est pas d'uniformiser, de définir une ligne, un courant ou une école de pensée, mais d'éclaircir et de préciser les notions afin de lever les malentendus, d'ouvrir les perspectives et de repérer les éventuels désaccords. Notre conviction est que le dialogue entre les disciplines ne pourra être fructueux que si l'on a réussi à dégager les notions essentielles, pour les discuter ou pour s'en servir comme point d'appui. Cette tentative n'est pas sans péril. D'une part, le contresens, ou simplement la superficialité, est possible. D'autre part les bases différentes des disciplines ou des courants peuvent être contradictoires. Nous espérons que la collaboration des membres du réseau nous

permettra de corriger les erreurs dues à la première difficulté. Pour la seconde si elle se présente, nous noterons les points de vue différents sans trancher entre eux.

14. Outre les difficultés conceptuelles, l'objectif peut aussi rencontrer des obstacles plus terre à terre. Chaque spécialité développe naturellement sa propre culture et son propre vocabulaire, pour de bonnes (rigueur) ou de mauvaises (protection) raisons. Les mêmes mots prennent parfois des sens différents selon les communautés, souvent même ils sont étrangers aux extérieurs. Nous sommes contraints dans un texte transversal d'utiliser un vocabulaire commun, dans tous les sens du terme, au risque de caricaturer.
15. Concrètement, ce texte est le résultat provisoire d'un travail collectif encore en cours au sein du réseau qui se développe en plusieurs étapes :
  - Plusieurs actions spécifiques ont été d'abord lancées, en 2001, 2002 et 2003 regroupant des chercheurs de diverses équipes sur des thématiques particulières ;
  - Une réflexion transversale a été initiée, début 2003, sur la notion de document en utilisant l'intranet du réseau et un travail en séminaire. Celle-ci a permis de recueillir des contributions en nombre, ainsi que de réaliser une bibliographie ;
  - Ce texte est un document de travail qui capitalise une part des étapes précédentes, mis en ligne fin avril 2003 pour critiques et amendements.
  - Parallèlement un travail collectif de "percolation" est mené sur la bibliographie, afin de sélectionner une dizaine d'articles ou chapitres de livre, jugés fondateurs.
  - Une version corrigée du texte de synthèse, accompagnée des articles sélectionnés sera mise en ligne fin juin 2003.
  - Elle servira de point de départ à la dynamique de réflexion pour la Semaine du document numérique qui regroupera colloques, séminaires et journées d'études sur le sujet fin juin 2004.
16. Compte tenu de la méthode employée pour la rédaction de ce texte et des nombreuses contributions qui lui ont donné corps, nous avons choisi de ne noter aucune citation ou référence bibliographique directe. L'inverse conduirait en effet à biaiser la dynamique collective en induisant une concurrence entre auteurs ou écoles de pensée. Néanmoins, le texte sera accompagné d'une dizaine d'articles, communications, ou chapitres de livres, choisis comme représentatifs par les contributeurs. Ces derniers, qui ne sont évidemment pas soumis à la même règle de non-citation, orienteront le lecteur vers de multiples autres publications.

## 17. Propositions

18. Pour organiser nos propositions, nous utiliserons une analogie avec la distinction en linguistique entre syntaxe, sémantique et pragmatique. Sans entrer dans une discussion sur la validité de cette analogie, nous constatons qu'elle permet un classement assez simple des recherches en cours et des courants qui les sous-tendent. Nous distinguerons :
  - Le document comme forme ; sous cette rubrique, nous rangerons les approches qui analysent le document comme un objet, matériel ou immatériel et qui en étudient la structure pour mieux l'utiliser ou le manipuler.
  - Le document comme signe ; pour ces chercheurs le document est perçu avant tout comme porteur de sens ; ainsi le document est indissociable du sujet qui le construit ou le reconstruit et lui donne sens ; en même temps, il est pris dans un système documentaire ou un système de connaissances.

- Le document comme relation ; cette dimension enfin pose la question du statut du document dans les relations sociales ; le document est une trace d'une communication qui s'est affranchie de l'espace et du temps ; en même temps, il est un élément de systèmes identitaires et un vecteur de pouvoir.
19. Chaque catégorie doit être vue comme une dominante et non comme une dimension exclusive. Ainsi, par exemple, les chercheurs qui abordent le document par la première entrée ne négligent pas nécessairement les deux autres, néanmoins leur analyse et raisonnement privilégient le premier éclairage, les deux autres restant des compléments ou des contraintes extérieures. Plus précisément, le terme "d'entrée" serait peut être le plus approprié, chacune de ces entrées est en effet une façon d'aborder l'objet de la recherche, le document, à partir de laquelle les autres dimensions seront retrouvées à l'occasion de développements, de contraintes, d'obstacles ou de limites apparues dans le raisonnement premier. Mais la tendance de chacune des approches est sans doute aussi de trop relativiser les autres.
20. Nous aborderons chaque catégorie selon le même schéma :
- tout d'abord, un repérage des principales disciplines, savoir-faire ou spécialités qui privilégient ce point de vue. L'objectif n'est pas de discuter leur validité ou leur scientificité, mais de passer en revue la diversité des travaux représentatifs de l'orientation sans jugement d'importance, ni de valeur.
  - puis, nous proposerons une interprétation de l'évolution des points de vue dans le passage du document traditionnel au document numérique.
  - nous construirons progressivement une définition du document à partir de chaque entrée.
  - nous pointerons quelques questions qui restent en suspens à l'intérieur de cette catégorie, au delà de l'approfondissement des recherches particulières en cours.
21. Chaque fois nous essaierons de dégager l'essentiel, sans trop nous arrêter aux nuances, exceptions et cas particuliers. L'objectif est de souligner les fondamentaux, pas d'être exhaustif. Concernant la définition, une méthode consisterait à rechercher systématiquement les cas ne lui correspondant pas et ainsi construire une définition universelle. Cette modalité ne nous paraît pas très opératoire. Notre objectif n'est pas de répondre à tout mais de construire une définition générique, quitte à repérer des exceptions qui représentent soit des cas très particuliers, soit des situations intermédiaires ou transitoires quand ils ne relèvent pas tout simplement d'une analyse incomplète ou erronée.
22. Nous ne concluons pas le propos. Le débat devrait en effet nuancer les propositions présentées, les contredire éventuellement et faire émerger des analyses alternatives. La synthèse des trois approches, à supposer qu'elle soit envisageable et pertinente, ne pourra valablement se construire qu'une fois la discussion arrivée à son terme.

### **23. Document comme forme**

24. Le terme de "forme" peut être contesté car il prête à confusion, nous l'utilisons car nous n'en avons pas trouvé de meilleur. Il faut l'entendre ici à la fois dans le sens de "contour", autrement dit le document est appréhendé comme un objet dont on repère les frontières, et en référence au "formalisme", c'est à dire que cet objet répond à des règles qui le constituent.

### **25. Spécialités concernées**

26. Dans cette catégorie l'ingénierie domine.

27. Les savoir-faire professionnels qui privilégient ce point de vue sont nombreux, parfois très anciens comme la calligraphie, la typographie, et aussi la bibliothéconomie dont le coeur est le catalogage, le classement et la gestion des documents ou bien, pour d'autres formes de la représentation, les techniciens de la musique, de la vidéo ou du cinéma..
28. Il est logique que les informaticiens qui partent d'objets matériels pour les numériser, ceux de l'image donc, aient de fortes parentés avec ces premières spécialités. Ils s'intéresseront vite à la structure interne des documents avec les systèmes automatiques de reconnaissance des formes, en tout premier la reconnaissance automatique des caractères, puis de l'écriture manuscrite ou encore de la forme des mises en page ou des images. Dans leur domaine, ils retrouveront les problèmes de formats, d'échange, de stockage, de description, d'adressage, de conservation, ou de traitement des grands nombres.
29. De la même façon, tous ceux qui s'intéressent aux caractères typographiques, à la normalisation internationale de ces domaines, au traitement de texte, ceux qui construisent des systèmes numériques de vidéo déclinent en les renouvelant profondément de vieux savoir-faire.
30. D'autres spécialités informatiques ont opté pour ce premier point de vue. La conception de systèmes de gestion électronique de documents, comme le nom l'indique, part bien de l'idée que le document préexiste comme objet repérable, même s'il peut être virtuel. Si cette fois, le point de départ n'est plus un objet concret, mais un fichier informatique, bien des problèmes posés relèvent des mêmes interrogations fondamentales.
31. Enfin, un changement brutal d'échelle est apparu avec l'invention et le succès explosif du World Wide Web. Autour de ce réseau de réseaux, une intense activité de recherche, conception, négociation, normalisation, développement s'est déployée, notamment mais pas seulement, au sein du consortium W3C. Même si ces chercheurs emploient peu le mot de "document" et lui préfère celui de "ressource", les questions posées par les concepteurs du Web, dans sa version actuelle (c'est à dire avant le "Web sémantique"), relèvent aussi largement de cette première approche : il s'agit de savoir comment relier, à une échelle planétaire, des ressources entre elles et donc de définir des normes et des systèmes applicables sur toutes les machines et de donner une adresse repérable aux dites ressources. Celles-ci sont des fichiers HTML qui ont les caractéristiques du document tel qu'on l'entend dans cette première dimension.

## 32. Evolution

33. Une première définition du document pourrait être représentée par l'équation : *Document = support + contenu*.
34. Dans un premier temps en effet, l'accent est mis sur un support manipulable (au sens premier) sur lequel est fixé une trace interprétable par un oeil ou une oreille humaine, le contenu. Le support traditionnel dominant (mais pas exclusif) est le papier, la trace l'écriture, manuscrite ou imprimée. La feuille écrite, élément de base, peut être enrichie par les mises en page et le paratexte et allongée par les reliures, renvois etc. conférant à ce document une très grande plasticité et complexité. Le "codex" (le livre avec des pages reliées) est sans doute la forme la plus élaborée du document traditionnel. Sa qualité peut se mesurer à la robustesse de "ses spécificités", pratiquement inchangées depuis près d'un millénaire ! Au prix d'un important effort social (l'école), ce document est directement perceptible, c'est à dire sans outil intermédiaire, par une part plus ou moins importante de la population : celle qui a appris à lire.

35. Quand, dans l'histoire, cette notion (*support + contenu*) s'est étendue à d'autres formes de représentation, comme la musique enregistrée, le cinéma puis l'audiovisuel, le support n'a pas gardé sa faculté d'appropriation directe. Alors même que la représentation se rapprochait de la perception humaine immédiate (et donc nécessitait un apprentissage moins lourd pour être décryptée), le dispositif de lecture s'est sophistiqué. Il est indispensable de disposer de machines pour écouter un disque (gravé) projeter un film (inscrit sur une pellicule), ou une bande vidéo (enregistrée). L'objet est toujours nécessaire à la lecture, mais il n'est plus suffisant. Mieux, la radiodiffusion puis la télévision ont permis de séparer le décryptage de la transmission du signal, l'appareil de restitution d'une représentation perceptible par l'humain. Ainsi l'auditeur ou le téléspectateur écoute ou regarde un "programme" dont la transmission lui échappe. D'une certaine façon, le réseau en entrant dans son domicile l'a dépossédé d'une partie de l'autonomie temporelle qu'il avait gagné en manipulant des objets gravés ou enregistrés.
36. Le passage d'un signal analogique à un signal numérique, avec toutes les facilités de traitement informatique qui l'accompagnent, a des conséquences radicales pour l'ensemble de ces documents, écrits, images et audiovisuels. Les mutations se repèrent sur les dispositifs d'écriture-lecture et sur les documents eux-mêmes.
37. Concernant les dispositifs, on observe tout d'abord un étonnant chassé-croisé entre l'écrit et l'audiovisuel. Le premier intègre un dispositif familier au second. Il n'est plus possible de lire sans machine. Disques optiques ou magnétiques, bandes enregistrées, machines pour traiter et restituer le signal ainsi que connexion par réseau sont les outils de la lecture électronique. Le second, l'audiovisuel, peut retrouver dans le Web la maîtrise temporelle, perdue au travers du réseau de radiodiffusion qui transmettait un flot de programmes ininterrompable par l'auditeur ou le téléspectateur.
38. La seconde conséquence sur les dispositifs est l'entrelacement des supports et des signaux. La notion de support se complexifie. Est-ce le fichier, l'outil matériel qui l'héberge ou encore la surface de l'écran sur lequel il s'affiche ? Les mêmes « supports » d'ailleurs peuvent accueillir indifféremment n'importe quel type de représentation pourvue qu'elle soit numérique et même, sous réserve de compatibilité de format, les représentations elles-même peuvent se confondre : on pourra "lire", étroitement imbriqués du texte, de l'image, de l'audio ou de l'image animée.
39. Enfin, depuis le développement conjoint de la micro-informatique et de la télématique, les machines elles-même se multiplient et s'autonomisent, entre ordinateurs portables, tablettes, téléphones, et outils intégrés de diverses fonctions recherchant la meilleure adéquation entre les comportements des lecteurs et leurs besoins, génériques ou spécifiques. Ainsi, il n'est pas indifférent pour le devenir des documents que la diffusion du téléphone portable ait été beaucoup plus rapide et plus étendue que celle du micro-ordinateur.
40. La réflexion des informaticiens conduit à isoler les éléments logiques qui constituent le document, pour les modéliser, automatiser les opérations et réorganiser les différents éléments ainsi perfectionnés. On pourrait résumer le sens de la transformation en cours par celle de l'équation première : *Document = support + contenu* deviendrait *Document = structure + contenu*. Dans cette entrée, le premier terme de la fonction est modélisable, le second est repris sans traitement spécifique.
41. La notion de support a perdu sa pertinence, tandis que les algorithmes définissent l'objet en autorisant sa représentation. Ceux-ci doivent piloter le signal qui court sur la surface de l'écran (visuel) ou dans les haut-parleurs (audible) pour qu'il construise l'équivalent de ce qui fut à la fois le support et reste encore la trace concrétisée par le fichier, autrement dit le

contenu tel qu'il est principalement appréhendé dans cette première approche. Cet effacement du support a deux conséquences qui donnent lieu à deux courants de recherche, les uns partent de l'analogique pour aller vers le numérique, les autres font le chemin inverse.

42. D'une part, si l'on veut utiliser les performances de l'outil, il va falloir traduire les anciens documents en documents numériques ; autrement dit, il faudra faire passer un document traditionnel d'une équation à l'autre. Cette opération peut se faire à partir du document premier que l'on numérise. On peut aussi raisonner simplement à partir de la représentation d'un document en reconstruisant directement, sans l'aide d'un ancien support, l'équivalent visuel de tout ou partie de sa représentation. C'est dans ce premier courant que l'on trouve les traiteurs d'image, comme leur nom l'indique leur recherche consiste à reconstruire l'image, donc la représentation, d'un document.
43. Le principe est la reconnaissance des formes. Pour être reconnue, une forme doit donc avoir été préalablement connue. Plus le document d'origine s'appuie sur des structures génériques, plus il sera facilement transposable. Ainsi la complexité s'accroît quand on passe des caractères typographiques au graphique puis à l'image et enfin aux objets en trois dimensions. Même si l'objectif est de reproduire une perception similaire ou homologue à 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.
44. Même si l'on renonce à l'objectif hors de portée de traduire l'ensemble des documents existants d'un support traditionnel à un support numérique, le développement explosif de l'informatique oblige à envisager des traitements de très grande ampleur, sauf à s'accommoder d'une amnésie radicale de notre culture documentaire. Il faudra donc 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.
45. D'autres chercheurs partent directement de l'équation finale (*document = structure + contenu*) ; autrement dit, ils font le chemin inverse. A partir du cœur du raisonnement informatique, l'élaboration d'algorithmes, ils reconstruisent les documents, remontant pas à pas leur logique ou structure interne pour déboucher sur une représentation lisible à l'écran. Ce second courant est issu de l'utilisation courante du texte dans les langages informatiques, il a débouché d'abord sur la bureautique, puis la télématique, et enfin sur la nécessité de pouvoir échanger les documents à grande échelle. Il a alors véritablement explosé avec la révolution du Web.
46. Ces informaticiens ont raisonné par couches, pour isoler et traiter séparément les éléments de la structure du document. Ils ont ainsi découvert ou redécouvert les différents niveaux logiques de cette dernière. Le niveau le plus bas étant celui du texte ou du signal analogique, que l'on a cherché à unifier sous forme d'unicode, MPEG etc.
47. Du point de vue qui nous occupe ici, le Web peut donc être décrit comme une infinité de documents reliés entre eux. Son architecture est bâtie sur trois piliers : des ressources identifiées par un schéma d'adressage universel (identification), qui sont représentées par un ensemble non-exclusif de schémas (représentation) et échangés selon des protocoles standardisés (interaction). Cette 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.
48. Les deux courants ne sont pas indépendants et partant de points différents, pour atteindre le même but, ils convergent. Pour notre propos, la révolution fondamentale est peut-être l'uniformisation progressive du format (au sens informatique) du document. 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 permettant de retrouver et de dépasser très largement la plasticité et complexité des feuilles reliées que nous avons soulignées au démarrage de cette partie et dont on avait perdu quelques fonctionnalités en route.

49. Progressivement, les documents ainsi « rédigés » rejoignent des bases de données, centralisées ou distribuées, et 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 pourront être agencées selon des configurations très variées. Un dernier pas est ainsi en train de se franchir : un document n'aura 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 soit 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.
50. Nous percevons très clairement les prémisses de cette dernière étape ; 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.

## 51. Définition 1

52. L'observation de cette première dimension nous conduit à formuler une définition du document numérique, à ce stade incomplète mais représentative d'un important mouvement en cours :
53. *Un document est une représentation organisée selon une structure stable.*

## 54. Questions

55. Cette première approche, par la forme, laisse plusieurs questions en suspens. Nous insisterons ici sur celles qui se rapportent à la relation entre le monde perceptible et l'organisation numérique du nouveau monde documentaire.
56. Une première série de questions relèvent de l'affichage des documents. Alors même que la "bibliographie matérielle" a étudié de très près l'objet livre sous tous ses aspects, il semble que le passage au numérique ait surtout approfondi la question de la structure à partir de son entrée logique et à des fins de traitement. Ainsi, dans cette première dimension, les chercheurs considèrent volontiers que, puisque la structure est intégrée au fichier, n'importe quel affichage est alors possible, et donc les problèmes de perception relèveraient d'une autre problématique. Cette conception suppose que structure et contenu sont indépendants, ce qui est pour le moins discutable. Il y a sans doute un important travail encore à mener sur la lecture électronique pour mieux comprendre les mécanismes d'interdépendance entre les deux termes de l'équation. Il suffit pour s'en convaincre de comparer les écrans d'un micro-ordinateur, d'une tablette, d'un agenda électronique ou d'un téléphone portable.
57. Les progrès réalisés sur la mise en écran, suite notamment aux travaux de la firme Xerox puis au perfectionnement des outils bureautiques, s'en tiennent à une organisation visuelle qui en reste, pour le document, à une mise en page et une mise en dossier. Quelques graphistes

proposent d'intéressantes compositions. Mais ces efforts paraissent peu liés aux recherches précédentes. De même les livres électroniques ou les espoirs mis dans l'encre électronique sont plus probants pour le moment sous la plume des journalistes que dans la vie courante ou même celle des laboratoires, même si la possibilité de reconstruire un codex électronique fait rêver. Certains pour qui la représentation spatiale est fondamentale, comme les géographes ou les architectes, ont déjà mené une réflexion poussée sur ces questions. , mais ils sont l'exception et non le cas général. Une nouvelle fois c'est peut-être l'audiovisuel qui ouvrent des voies prometteuses avec la réalité augmentée où sont intégrés des aspects analogiques et des reconstructions numériques.

58. Une deuxième série de questions concerne la stabilité et la pérennité des documents numériques. Ces questions sont souvent débattues. D'un côté, elles ne diffèrent pas des très anciens problèmes d'archivage et de conservation, simplement transposés sur d'autres techniques. D'un autre côté, des problèmes radicalement nouveaux sont posés : les fichiers XML, pour peu qu'ils soient régulièrement rafraîchis et conservés dans de bonnes conditions, sont en théorie inaliénables puisqu'ils contiennent la totalité de leurs informations sous forme numérique. Ainsi certains considèrent que, sous peu, les problèmes de pérennité seront résolus. Mais à l'inverse, ces fichiers ne représentent pas, loin s'en faut, la (ou les) forme(s) sous lesquelles sont lus les documents. Ainsi une mémoire complète de ces documents 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.
59. Enfin, sans prétendre épuiser les problèmes, soulignons une troisième 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. Néanmoins, le lecteur a toujours accès au document à partir d'une machine, le terminal où ce dernier s'affiche. Ira-t-on vers une version extrême de cette idée où le document ne serait plus que l'avatar moderne d'une ardoise magique où s'afficheraient à la demande des éléments signifiants multimédias, seulement contraints par une logique de sens ? Ou aura-t-on une restructuration de documents types répondant à des besoins ou situations particulières dont l'éventuelle dynamique sera confinée à des plages strictement définies ? Et 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 ?
60. On pourrait résumer ces questions par une seule : en effaçant le support, n'a-t-on pas trop délaissé justement la forme ?

## **61. Document comme signe**

62. Comme pour l'entrée précédente, les termes du titre de cette partie ne doivent pas être lus dans une acception trop académique. Depuis fort longtemps, le signe fait l'objet de très nombreux travaux savants. Même si certains d'entre eux nous serviront, nous ne cherchons pas ici à discuter le concept. Notre objectif est simplement de regrouper et présenter les recherches qui prennent le document d'abord comme un objet signifiant.

## **63. Spécialités concernées**

64. Cette catégorie concerne des disciplines sensiblement différentes de la précédente, certaines se présentant comme une avancée historique par rapport à celle-là, comme si, passant de la forme au signe, on se rapprochait du coeur d'un problème.

65. Ainsi, du côté des savoir-faire professionnels, on passe de la bibliothéconomie à la documentation puis aux professionnels de l'information, qui, plutôt que gérer des objets, fournissent des réponses aux questions du lecteur. Ou encore la GED devient le Knowledge Management (KM) qui au delà d'un système de gestion de stocks de fichiers permet de repérer directement les connaissances utiles pour une organisation. Et surtout, le Web gagne un adjectif en s'appelant Web sémantique, voulant ainsi signifier qu'une meilleure utilisation des capacités des machines interconnectées pourrait autoriser un traitement du contenu des fichiers en ligne en vue de l'organisation de services plus proches des demandes cognitives des internautes.
66. D'un point de vue académique, cette catégorie réunit d'abord ceux qui travaillent sur le texte (pris au sens large) donc des linguistes ou des sémioticiens toutes écoles confondues aussi bien ceux qui font de l'analyse de discours, de la linguistique de corpus, de la sémantique que ceux qui construisent des outils de traitement automatique de la langue en vue de traduction ou de recherche automatiques d'informations. Un rapprochement est en cours avec une seconde catégorie d'informaticiens issus plutôt des travaux sur l'intelligence artificielle, qui, à partir d'une tentative de modéliser le raisonnement, cherchent à construire des outils capables, eux aussi, de répondre aux questions sur la base d'une recherche dans des fichiers. Dans le même temps, on passe de la notion d'information à celle de connaissance qui a l'avantage sur la première d'intégrer le raisonnement. Une discipline nouvelle baptisée "ingénierie des connaissances" est ainsi en émergence.
67. Très vite, il est apparu qu'un travail d'information sur l'information était utile, parfois indispensable. Du catalogage à l'indexation, puis des thésaurus aux ontologies, les "métadonnées" sont devenues un outil et un objet de recherche essentiel.
68. Ici, comme pour l'entrée précédente, l'explosion du Web a modifié la donne en changeant l'échelle des ressources disponibles. Ainsi, une autre façon de lire la tentative de construction d'un Web sémantique serait la rencontre entre les architectes du Web et les chercheurs de cette dimension.

## 69. Evolution

70. La définition du document traditionnel selon cette dimension pourrait être symbolisée par l'équation : *Document = contenu + sens*.
71. Ici, la forme est accessoire, l'important est le contenu qui est porteur de sens. Autrement dit, la notion de support, qui nous a occupé dans la dimension précédente n'a plus qu'une pertinence secondaire, du moment que le contenu (la trace) est préservé.
72. Trois idées forces nous semblent fonder cette dimension, dans un triangle classique en sémantique. La première concerne la création des documents, la seconde leur interprétation et la troisième les signes qui les constituent.
73. "Penser c'est classer", en réalisant des documents nous isolons et rangeons des énoncés pour nous aider à penser le monde. La mise en document est une façon de construire, ou de traduire, notre compréhension sociale. Ainsi la notion de genre textuel ou celle de collection est fondamentale. En effet, les documents se regroupent par grandes catégories dont les différents items ont une homologie et une relation entre eux. Ce classement varie selon les situations et les époques. Il peut être très formalisé, comme simplement implicite. Il peut faire référence à des actions très précises et organisées (papiers d'identité, formulaires, contrats, etc.) ou à de simples attentions, impressions, sensations (médias, fictions, etc.). Il marque notre représentation sociale du monde. Il 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.

74. Le second point important est l'interprétation. 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, voir opposés, selon l'époque et la situation sociale ou individuelle de l'interprétant. D'une certaine façon, ce dernier re-crée le document chaque fois qu'il l'isole et en prend connaissance. Le lecteur doit ici être entendu dans un sens général, il peut s'agir aussi bien d'une personne physique que d'un groupe de personnes. Ainsi un document, pour la dimension qui nous occupe ici, est pris dans une double relation, relation au monde documentaire (idée précédente) et relation au monde naturel (interprétation). Ces relations se réalisent grâce à un "horizon d'attente", un ensemble de signes familiers qui construisent 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.
75. La troisième idée force concerne les signes eux-mêmes. Tout objet est potentiellement un signe et pourrait-être un "document". Une discussion, désormais classique, a montré par exemple qu'une antilope dans un zoo (donc dans un système social de classement) était un document. Mais la très grande majorité des documents sont construits à partir du langage, écrit majoritairement ou parlé. Le zoo, lui-même, est construit autour d'un discours, l'antilope est en quelque sorte "documentée". On peut faire la même remarque au sujet des documents audiovisuels qui sont toujours accompagnés d'un grand nombre de textes dès le moment de leur fabrication jusqu'à celui de leur exploitation. La structure de la langue écrite depuis la lettre d'alphabet jusqu'au discours organise donc la plupart des documents. Ceux-ci sont fait de morceaux discrets, plus ou moins isolables et réagencables, analysables, soumis à des règles de syntaxe et de style. Cette utilisation de la langue naturelle confère au document une très grande plasticité.
76. 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 à l'invention 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 de la triade précédente : il était possible en effet de construire à partir des textes des documents (ou des images, ou des objets eux-même) un langage abrégé permettant de les classer pour les retrouver à la demande.
77. La construction de ces méta-langages pose de nombreux problèmes. Elle suppose d'abord une normalisation, un certain nombre de règles communes sur lesquelles les différents protagonistes font consensus. Le consensus n'est pas suffisant, il faut y ajouter la motivation. Chaque personne qui participe à l'effort commun doit en récolter un avantage clair, faute de quoi il est peu probable que la construction collective soit effective. Ces langages enfin balancent continuellement entre l'universel et le contingent. L'hésitation est souvent mal comprise. Il ne s'agit pas d'une faiblesse conceptuelle ou d'une incapacité à choisir. C'est, au contraire, une dynamique fondatrice du mouvement documentaire, dialectique entre général qui classe et particulier qui réfère.
78. Un autre élément à souligner est l'insistance des documentalistes à se différencier des bibliothécaires par le service qu'ils rendent : la recherche d'information. Les premiers s'attachent à analyser le contenu des documents pour présenter à leurs usagers les réponses qu'ils attendent et non simplement le ou les documents qui les contiendraient éventuellement. Ainsi, les documentalistes participent à l'interprétation des documents disponibles en

reconstruisant, en quelque sorte, pour le lecteur un document, ou un dossier documentaire, adapté à son besoin.

79. 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.
80. 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. Ainsi, comme précédemment, nous pourrions représenter la transformation par celle de l'équation première : *Document = contenu + sens* devient avec le numérique *Document = données structurées + connaissance*.
81. Le traitement automatique de la langue dépasse largement la problématique documentaire. Néanmoins à l'inverse, le traitement du document est nécessairement concerné par les avancées et les difficultés des outils de traitement de la langue dès lors que l'on s'attaque au plein texte. 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 morpho-syntaxiques. 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 méta-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 zéloteurs. Les meilleurs outils ont du intégrer une part de travail humain, se présentant plus comme des outils d'aide que comme des outils automatiques.
82. Néanmoins, leur efficacité est spectaculaire dans leur application au Web sous forme de moteurs. Il est frappant de retrouver dans ces derniers, peut être parce qu'ils s'adressent maintenant aux très grands nombres (de documents, comme d'internautes) les vieilles questions bibliothéconomiques : lois bibliométriques (Zipf), collections (copies caches), indexation et mots clés (métadonnées), citations (liens), prêts (hits).. évidemment largement renouvelées par la puissance de calcul, utilisant les apports du traitement automatique de la langue, mais résultant souvent plus d'un bricolage empirique de méthodes que d'une analyse scientifique trop rigoureuse.
83. 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. L'ambition est aussi plus large. La rencontre des promoteurs du Web avec les chercheurs de l'ingénierie des connaissances, dont l'objectif est la modélisation du raisonnement, débouche sur la volonté d'aller au delà d'un traitement de la langue pour intégrer la production du sens. Parmi les outils pour avancer dans cette direction, on trouve principalement "les ontologies" qui se présentent comme des sortes de thésaurus dont les termes, reliés ensemble, intègrent les modalités logiques de leurs significations et de leurs combinaisons. Il s'agit bien, une nouvelle fois, d'un méta-langage qui permettrait notamment d'interpréter les documents mis en ligne sur le Web.
84. Néanmoins l'augmentation du nombre de documents accessibles dans une forme ne comprenant pas de méta-donnée 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, on trouve deux dynamiques parallèles. 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.

85. Quand on observe l'évolution et les avancées des recherches dans cette dimension, il ressort un aspect cyclique : les changements de support, d'échelle ou d'outil obligent à reposer des questions anciennes. La construction controversée d'un méta-langage émerge à chaque étape. Ainsi les tenants de l'étape précédente ont l'impression que les nouveaux venus redécouvrent des problèmes anciens tandis que ces derniers pensent que le saut réalisé oblige à remettre à plat l'ensemble des problèmes. Il n'est pas vraiment étonnant que la construction d'un méta-langage soit cyclique. Chaque changement de support ou d'échelle nécessite de reconstruire sa structure. En même temps, ses fondements ne sont pas pour autant remis en cause, ils sont (ou devraient être) simplement mieux connus et plus solides.

## **86. Définition 2**

87. Selon cette deuxième dimension, nous pourrions donc présenter la définition du document de la façon suivante, sans prétendre toujours qu'elle englobe la totalité de la notion :
88. *Un document est un ensemble de données dont la combinaison est porteuse de sens.*

## **89. Questions**

90. En se rapprochant de la communication humaine, les chercheurs de cette entrée ont largement augmenté la complexité des problèmes à traiter. Bien des questions 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 sont en rupture avec les langues indo-européennes. Par ailleurs, la frontière entre l'automatisation et le travail intellectuel humain reste mouvante.
91. Mais pour notre propos, notons surtout que, pour les chercheurs qui privilégient cette entrée, le document paraît souvent une notion secondaire, seul le texte, le contenu, compte vraiment. Pourtant le contenu n'a de valeur, nous l'avons vu en introduction de cette approche, que par rapport à un contexte. Le document n'est-il pas justement une des constructions de ce contexte en positionnant les informations qu'il contient par rapport à celles contenues dans d'autres documents et en permettant au lecteur d'avoir une indication de la valeur du contenu par le statut du document ? Autrement dit en portant trop exclusivement son attention sur le traitement du texte, ne sous-estime-t-on pas la valeur sémantique de la mise en document ?
92. Il se pose alors une série de questions à partir de la triade notée en introduction de cette entrée :
93. Est-il possible de traiter du sens d'un document sans relation à l'ensemble auquel il se réfère (collection, catégorie, renvois, bibliographie, etc.) ? Autrement dit, comment, au delà d'un travail de laboratoire sur des corpus fermés, intégrer l'analyse du document comme "tête de réseau" ?
94. Comment valider une information, sinon par l'authenticité du document qui la contient ?

95. Dans quelle mesure peut-on isoler un élément signifiant dans un ensemble qui a une unité de sens, le document comme un tout ? Cette unité n'a-t-elle pas souvent un poids décisif dans la signification des éléments qui la compose ? Comment prendre en compte la signification globale, l'unité sémantique, d'un document en appréhendant que ses parties ?
96. Ces questions ont sans doute des réponses différentes selon les types de documents auxquels on les applique. Mais faute, pour le moment, d'une avancée réelle sur la typologie, elles restent, nous semble-t-il, largement ouvertes.

### **97. Document comme relation**

98. Renouvelons pour la dernière fois notre précaution de vocabulaire : le terme "relation" doit être pris ici dans un sens large. Il regroupe toutes les approches qui analysent le document comme un phénomène social, un élément tangible d'une communication entre des personnes humaines.

### **99. Spécialités concernées**

100. Sans doute tous les chercheurs précédents pourraient être inclus dans cette catégorie, puisque tous s'intéressent à une activité sociale, néanmoins nous classerons ici ceux dont l'entrée est d'abord sociale avant d'être instrumentale.
101. Du côté des savoir-faire traditionnels aussi les métiers déjà cités se retrouvent ici, mais nous insisterons sur les archivistes, dont la mission première est de garder trace de l'activité humaine en conservant des documents produits au fil de l'eau par celle-ci, et les éditeurs, dont le métier est de favoriser la construction et de rendre public des documents intéressants une collectivité.
102. Les disciplines des sciences humaines et sociales qui s'intéressent aux échanges sont potentiellement concernées par cette dimension. De fait, des sociologues, des économistes, des juristes, des historiens quelques psychologues, un nombre certain de philosophes, et bien entendu des chercheurs en sciences de la communication, en sciences politiques et en sciences de la gestion s'intéressent, directement ou indirectement, aux documents à partir de leur entrée disciplinaire.
103. Le numérique a renouvelé l'intérêt de nombre de chercheurs de ces disciplines tant sur le phénomène global que sur des situations particulières. Ainsi, sans qu'il soit toujours assumé, il y a un rapport entre la réflexion sur le document et l'intérêt nouveau pour les communautés d'intérêt, les collaboratifs, le travail en réseau, la mémoire et le patrimoine, la propriété intellectuelle, etc.
104. Mais dans cette entrée le décalage entre les informaticiens et les autres chercheurs est plus grand. Bien peu des spécialistes en sciences sociales et humaines ont une connaissance réelle de l'informatique. Inversement, les informaticiens ont une compréhension souvent rapide des problématiques sociales. Ce fossé conduit parfois à des enthousiasmes fascinés ou à l'inverse à des rejets radicaux, de la part aussi bien des tenants des sciences sociales et humaines que de ceux des sciences de l'ingénieur.

### **105. Evolution**

106. Pour être en harmonie avec les entrées précédentes, proposons la troisième et dernière définition sous forme d'une équation : *Document = sens + légitimité*. Cette présentation est sans doute moins pertinente que les précédentes. Nous entrons ici dans le domaine social où la modélisation est plus délicate.

107. Un document donne un statut à une information, à un signe. C'est une preuve qui fait foi d'un état des choses. C'est une annonce qui prévient d'un évènement. C'est un discours dont la signature le rattache à un auteur, etc. Ce statut ou cette légitimité s'acquiert sous deux conditions : le sens doit dépasser la communication intime (entre quelques personnes privées) et il doit s'affranchir de l'éphémère (dépasser le moment de son énonciation). Ces conditions impliquent que si tout signe peut être un document, un signe particulier (même répondant aux deux dimensions précédemment traitées) ne l'est pas nécessairement. Par exemple, un journal intime n'est pas un document, sauf si quelqu'un prend l'initiative de le rendre public ou au moins de le communiquer au delà du cercle restreint des familiers de son auteur. Ou encore, une émission de radio ou télévision en direct n'est pas un document, sauf si quelqu'un l'enregistre pour une utilisation sociale future. De la même manière, le statut de document n'est pas acquis pour l'éternité, il se donne et il peut aussi se perdre dans l'oubli collectif ou encore se retrouver.
108. Ainsi les documents sont utilisés pour la régulation des sociétés humaines en assurant une communication et une pérennisation de connaissances. Leur multiplication est donc liée à l'évolution de ces sociétés autour de trois dynamiques étroitement corrélées, que nous ne ferons que suggérer ici, tant leur étude dépasse les limites de ce texte.
109. L'organisation politique et sociale tout d'abord 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. Ainsi, citons les principales sources de l'activité documentaire sans prétendre à l'exhaustivité :
- En France, le passage de l'Ancien régime à la République, puis celui de l'Etat-gendarme à l'Etat-providence, et enfin aujourd'hui l'intégration de l'Etat à des ensembles plus vastes, comme l'Europe, ou la mondialisation, n'a pas été sans conséquence sur la production des documents, leur rôle et leur nombre. Il suffit, par comparaison, d'évoquer l'importance du document dans l'histoire parallèle de l'administration de la Chine pour percevoir combien celui-ci est à la fois fondamental et pourtant bien spécifique à chaque civilisation.
  - L'industrialisation avec tous les savoirs et normalisations techniques, organisationnels, transactionnels et comptables qui l'ont accompagnée, a "produit" un nombre considérable de documents. C'est peut-être le facteur principal de l'explosion documentaire notée plus haut.
  - Les progrès scientifiques et ceux de l'éducation ont considérablement augmenté le nombre des producteurs et des consommateurs de documents, pour le fonctionnement interne de la science et, plus encore, pour la popularisation des innombrables savoirs partiels qui l'accompagne.
  - Les échanges commerciaux ou non commerciaux, qui ont explosé avec le développement des transports, celui des télécommunications, et l'ouverture des frontières utilisent nombre de documents pour se "fluidifier".
  - Le développement du temps libre, l'allongement de la durée de vie, l'accroissement de "l'espace public" sont encore des facteurs essentiels du développement de la culture et d'un de ses principaux vecteur : le document.



110. Cette dynamique a fait l'objet d'éclairages particuliers, mais les quelques tentatives d'appréhensions générales, peut être parce qu'elles restent des travaux solitaires, nous semblent des essais plus spéculatifs que démonstratifs.
111. La seconde dynamique qui permet de fonder le statut du document est celle des technologies intellectuelles. L'invention des écritures tout d'abord, puis celle de l'imprimerie et ses nombreux perfectionnements, puis celle de l'enregistrement analogique de l'image et du son dans l'audio-visuel ou encore le transport à distance du signal par les télécommunications et, aujourd'hui d'intégration de l'informatique par le numérique place le document dans une histoire technique, dont il est un des acteurs mais dont bien des paramètres lui échappent et, par conséquent, qu'il subit aussi très largement.
112. Certains ont fait de cette dynamique la clef de son évolution, parlant pour aujourd'hui de période "post-gutenbergienne". Cette dénomination sous-entend qu'auparavant l'imprimerie, ou la feuille imprimée, était le facteur déterminant de l'organisation documentaire. Ainsi, le caractère d'objet des feuilles de papier imprimées, leur inaliénabilité, l'organisation de leur mise en page, sont présentées comme des limites qui contraignaient la communication par l'imprimé et sont dépassées par le numérique. Pourtant, l'évolution ne concerne pas que l'écrit imprimé sur une feuille de papier. Bien au contraire, elle peut aussi se lire comme l'addition et l'entrelacement progressif de différents médias, dont l'audiovisuel. Dans cette perspective, le numérique peut être aussi analysé non pas comme un aboutissement ou une rupture, mais simplement comme l'intégration d'un média supplémentaire à l'histoire générale du document.
113. La problématique de ce texte qui tente de mesurer les conséquences du numérique sur l'évolution de la notion de document pourrait laisser croire que nous privilégions cette évolution technique, réduisant le reste à des conséquences. Soulignons donc que nous ne cherchons pas à hiérarchiser les trois dynamiques sociales qui permettent de donner son statut au document, il nous suffit de les repérer.
114. La troisième et dernière dynamique donc est celle de l'économie des documents. La mise en document suppose un travail dont il faut bien trouver les moyens de la réalisation. Cette mise en document peut être analysée comme un acte de communication ordinaire avec d'un côté un (ou plusieurs) expéditeurs et, de l'autre, un (ou plusieurs) destinataires. Des métiers se sont spécialisés sur tel ou tel moment du processus 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.
115. Les premières avancées de l'analyse ne touchent qu'une partie de l'activité documentaire et sont encore loin de formaliser les contours d'une économie générale des documents. Il reste bien des incertitudes à lever par exemple sur les relations entre les systèmes documentaires et les systèmes d'organisation, l'évolution des différents métiers de la médiation ou encore l'économie des bibliothèques ou des archives.
116. Mais l'économie de plusieurs secteurs est bien connue parce qu'elle a fait l'objet d'analyses particulières. Citons la communication scientifique avec le rôle de la publication des articles, la révision par les pairs, les citations, les pré-publications, les archives ouvertes etc. Ou encore les médias grand public avec la gradation entre une édition, artisanale fondée sur la vente individuelle d'objets, la dialectique entre le fonds et le best-seller, les réseaux de distribution, et la radio-diffusion, plus industrielle organisée sur la captation à son domicile de l'attention du destinataire vendue à des annonceurs intéressés.
117. D'autres thèmes ont été approfondis, compte tenu des enjeux économiques qu'ils représentent, comme le droit de la propriété intellectuelle. Dans ce domaine, les traditions

différentes entre le droit d'auteur latin et le copyright anglo-saxon permettent de distinguer quelques propriétés du document.. Le premier privilégie l'attachement de l'auteur à son oeuvre tandis que le second met en avant la notion de publication, donnant la propriété intellectuelle à celui qui en prend l'initiative. D'une certaine façon, nous pourrions dire de le droit d'auteur est un droit de l'oeuvre, tandis que le copyright est un droit du document.

118. Notons parmi les réflexions des économistes sur l'information un dernier point concernant l'économie des documents, de première importance pour notre propos : le document est un bien dit "d'expérience", c'est à dire qu'il ne peut être essayé. Cette caractéristique confère une valeur essentielle à l'information qui entoure le document. Plus un document est connu, plus il sera lu et plus il sera lu, plus il sera connu. Grâce aux relations entre les lecteurs et celles entre les documents, il peut se développer un phénomène de résonance, qui prend différentes formes et différents noms selon les secteurs ou les spécialités. Les professionnels du marketing ou les stratèges des médias l'utilisent régulièrement en construisant des notoriétés qu'ils revendent sur d'autres supports. Dans la communication scientifique, le facteur d'impact, basé sur le comptage de citations dans les articles, relève du même processus (et conduit aux mêmes dérives..). Cette qualité est peut être l'explication de bien des caractéristiques de la distribution des documents : best-sellers pour l'édition, prime-time pour la radio-télévision, modes diverses, concentration et éclatement ou encore régularité quasi-parfaite des lois de la bibliométrie quand de grandes quantités de documents sont accessibles de manière égale par un grand nombre de demandeurs.
119. L'entrelacement de ces trois dynamiques permet de repérer quelques changements notables qui accompagnent le développement du document numérique dans les processus sociaux.
120. Le premier constat est celui de l'effacement, non mesuré précisément mais effectif depuis plusieurs dizaines d'années et toujours en cours, d'un nombre important de documents qui rendent compte de procédures. Formulaire, 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. Ce mouvement qui fut baptisé il y a quelques années "informatisation de la société" risque de s'accélérer encore avec les développements que nous avons soulignés dans les dimensions précédentes.
121. Cette mutation accroît considérablement les possibilités de contrôle par les facilités de croisement des informations. Dans le domaine social, la France s'est doté d'une protection légale, mais fragile compte tenu du développement des opérations numériques, par la loi "Informatique et liberté". Dans le domaine économique, certains analystes y ont vu l'émergence d'une nouvelle économie dont l'étude des déboires dépasse largement le cadre de cet article. Retenons, pour notre propos, l'idée d'un changement radical des structures socio-économiques. Comme l'ère industrielle a été marquée par interchangeabilité des parties, la société de l'information serait caractérisée par la possibilité de ré-utiliser l'information.
122. Dans cette perspective le document disparaîtrait pour laisser place à des processus permettant des affichages éphémères d'informations, éventuellement validées par des procédures certifiées. Nombre de documents peuvent s'apparenter à des procédures ou à une de leur étape, ainsi ces documents ne seraient finalement que des programmes (au sens informatique). Si l'on ajoute les progrès de la signature électronique, bien des transactions pourraient à l'avenir se réaliser sans le recours à un document formel, même électronique.
123. En suivant un raisonnement parallèle, mais dans une perspective plus 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, 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.

124. De plus, le numérique a permis de renforcer les "anciens" médias, édition et audiovisuel, en leur autorisant d'intéressants gains de productivité en favorisant les synergies et les diversifications. Ainsi, à partir d'une même base de données par exemple, un journal peut décliner les actualités à la fois sur des sorties papier, sur le Web, sur la radio, par SMS, par audiotel, etc. Ou encore chaque média peut faire valoir ses propres domaines d'excellence (notoriété pour la télévision ou la radio, interactivité pour le Web et le téléphone, appropriation pour l'édition) et par la résonance notée plus haut conduire à des profitabilités inédites. Ces changements récents restent à évaluer. Ils conduisent aussi à des investissements lourds dont les retours sont à moyen terme alors même que l'avenir reste incertain.
125. Le dernier point sur lequel nous voudrions insister est le changement d'échelle brutal que le Web a fait subir à trois modes de communication dont l'économie était auparavant marginale. Le Web a été construit à l'origine sur le modèle de la communication scientifique qui se réalisait à l'intérieur de communautés éclatées géographiquement, mais limitées en nombre, culturellement homogène et autorégulée. Dans ces communautés, le document joue un grand rôle au travers de trois pratiques : la communication de plusieurs à plusieurs (colloques, pre-print, etc.), les revues scientifiques (document validés et diffusés largement), les bibliothèques (mémoire accessible sur place). Le développement très rapide de l'activité scientifique rendait cette communication de plus en plus difficile. Le Web a été inventé pour résoudre ces difficultés. Il n'est pas douteux qu'il a très largement réussi, même si la réorganisation de la communication scientifique est encore loin d'être stabilisée.
126. Mais dans le même temps, l'ouverture du Web à l'ensemble de la société favorisée par l'implication des grands groupes informatiques et de télécommunication a décalqué les pratiques (ou au moins certaines d'entre elles) de communication du monde scientifique à l'ensemble de la culture. Il est frappant, par exemple, de constater l'insistance mise sur les communautés virtuelles géographiquement éclatées, les bibliothèques virtuelles favorisant l'accès de tous au savoir ou encore les sites spécialisés régulés par les "experts", légitimes ou auto-proclamés, d'un domaine. Cette importation d'un outil sur un autre terrain a fait surgir une dynamique communicationnelle touffue qui était bridée par l'organisation traditionnelle des médias et qui parfois la conteste brutalement par des pratiques de piratage. Mais, cette dynamique fait aussi l'impasse sur l'économie des contenus, comme si les industriels du contenant récupéraient à leur profit l'ensemble des dépenses de communication des ménages.
127. Dans cette perspective, nous avons donc repéré des mouvements forts, parfois divergents, souvent chaotiques. Il paraît prématuré à ce stade de modifier radicalement la définition de la notion que nous avons donnée.

### **128. Définition 3**

129. Nous pouvons, maintenant proposer une troisième définition conforme à cette dernière entrée : *Un document est la mémoire d'une relation assumée et socialisée.*

## 130. Questions

131. Pourtant nous relèverons trois séries de questions qui toutes concernent le statut de document et nous paraissent insuffisamment travaillées.
132. La première série concerne 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. "Record management", 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. 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, etc.
133. La seconde série de questions concerne la notion d'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.
134. 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. Ainsi 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. Ces phénomènes et leurs conséquences sont encore bien peu étudiés
135. Par ailleurs, l'intention des promoteurs du Web est de mettre à disposition de façon égale pour l'ensemble de la planète les sites ou les documents. Mais la diffusion des innovations est très inégalitaire. Le Web et le numérique n'échappent pas à la règle. Pire, il semble que ce soit le média le moins bien partagé entre les pays et entre les populations à l'intérieur de chaque pays.
136. La troisième série de questions concerne l'oubli du financement du contenu. La loi du moindre effort appliquée à l'accessibilité du Web fait que l'internaute préférera éviter tous les obstacles et barrières à sa navigation plutôt que les affronter. Ainsi, il contournera toute demande de financement directe. Dans la même dynamique, un mouvement militant fait de la gratuité une sorte de qualité naturelle du Web ou une accessibilité au savoir et à la culture libéré des contraintes commerciales.
137. Opportunisme et politique se conjuguent pour que, progressivement, l'économie du contenu sur le Web se configure comme un marché institutionnel, du "B2B". Est-on vraiment sûr que cette structure de financement garantit à moyen terme la diversité et la pluralité des documents mis en ligne ? Est-on sûr simplement qu'elle contient les ressources suffisantes, secteur par secteur, pour alimenter la production et gestion de documents ?

## 138. Conclusion

139. Nous avons organisé notre présentation du document selon trois dimensions : forme, signe, relation. Il reste à valider si, ainsi, nous avons pu toucher l'essentiel de la notion.
140. Nous en avons déduit trois définitions :
  - *Un document est une représentation organisée selon une structure stable.*

- *Un document est un ensemble de données dont la combinaison est porteuse de sens.*
- *Un document est la mémoire d'une relation assumée et socialisée.*

141. A supposer que ces propositions soient pertinentes, on pourrait envisager une synthèse en proposant un éclairage général, et donc une définition qui englobe les trois points de vue, un peu comme la palette entière du peintre peut se construire à partir de trois couleurs primaires. A partir de cette définition, nous pourrions ensuite mieux mesurer les conséquences actuelles et à venir du numérique.

142. Néanmoins, ce serait prématuré. Cet objectif suppose d'abord que les parties précédentes aient été bouclées. Le texte est soumis actuellement à la discussion collective en vue d'une version complétée et remaniée. Il est préférable donc d'attendre que le débat ait pu s'engager et la réflexion mûrir au sein du réseau.

Roger T. Pédaque  
CNRS

Si vous souhaitez participer à la rédaction de ce texte, il vous suffit d'envoyer vos remarques, en précisant le cas échéant le numéro du ou des paragraphes auxquels elles réfèrent, à [pedauque@enssib.fr](mailto:pedauque@enssib.fr). Votre message sera envoyé au forum Pédaque qui regroupe tous les contributeurs et vous serez inscrit à ce forum. Si vous êtes déjà abonné au forum, il vous suffit d'envoyer le message à la liste selon la procédure habituelle.

Ce forum est consacré uniquement à la discussion scientifique sur la notion de document. Seuls les messages s'inscrivant dans cette perspective seront pris en compte. Les simples demandes d'inscription, ou les annonces sans contribution seront rejetées. Le forum sera clos au 30 juin.