

# Document numérique et normalisation : questions autour de la fragmentation et de la dispersion de l'information

par ROUISSI Soufiane « [Soufiane.Rouissi@u-bordeaux3.fr](mailto:Soufiane.Rouissi@u-bordeaux3.fr) »  
CEMIC-GRESIC - Université de Bordeaux

Le document numérique peut être désormais produit et diffusé de plus en plus facilement dans le contexte du Web. Les possibilités de production autonome (dans les environnements de type blog, wiki ou système de gestion de contenu...), les possibilités de réutilisation et d'assemblage de fragments documentaires ou de services numériques reposant sur les principes du web 2.0 amplifient la situation d'explosion documentaire numérique. Partant d'une réflexion théorique sur le document numérique, sur ses caractéristiques particulières, nous étudions en quoi une démarche normative peut-elle aider à réduire les problèmes potentiels nés de la fragmentation et de la dispersion de l'information.

*Mots-clés* : document numérique, normalisation, fragmentation de l'information

Electronic documents can be produced and published more and more easily in the web context. The possibilities of autonomous production with blog, wiki or content management system, as well as the possibilities of re-use digital fragments amplify the situation of documentary explosion. In this paper we study how a normative initiative can help us to reduce the potential problems arisen from the fragmentation and from the dispersal of the information.

*Keywords* : electronic document, standard, fragmentation of information

## Introduction

Dans une société de l'information encore en construction, chacun est conduit à assurer sa présence à l'autre à travers une connectivité accrue. Chaque individu, chaque organisation se doit de développer sa capacité de dialogue et ses potentialités d'échange de ressources documentaires. Le contexte technologique (évolutif par définition) dans lequel la télévision, le téléphone et l'informatique convergent, est fortement marqué par le passage au mode numérique. Ce passage est pour nous une source de questionnements notamment du point de vue de l'élargissement des accès aux ressources documentaires numériques et du développement des environnements de production documentaires numériques. Notre travail de recherche repose sur la notion de « document », notion pourtant ancienne et revisitée du fait de son passage à un mode numérique. En effet, nombreuses sont les questions autour de l'émergence et du développement du document numérique. Au-delà de considérations uniquement techniques sur les formats d'édition, nous nous intéressons aux questions liées aux usages potentiels dans un environnement supporté par les technologies de l'information et de la communication (TIC) en particulier le Web. Notre travail sur le document numérique, « objet fédérateur de recherche en sciences de l'information » (Chartron, Guyot, Lafouge, 2002), s'attache à étudier les conditions qui permettraient de répondre en partie aux risques liés à la fragmentation et à la dispersion de l'information. Partant d'une réflexion théorique autour du document et notamment son passage au mode

numérique, nous présentons ses nouvelles caractéristiques de façon à mettre en évidence les problèmes potentiels auxquels nous devons faire face désormais.

# Du document traditionnel au document numérique : vers un document « programmable » ?

## La notion de document numérique

Etymologiquement, le mot « document » venant de *docere* qui signifie instruire, se trouve défini dans la plupart des dictionnaires de la manière suivante : « *chose écrite qui peut servir à renseigner, à prouver* » ou encore « *écrit ou objet servant de témoignage ou de preuve, constituant un élément d'information* ». La preuve ou le témoignage renvoient à la notion de stabilité de cet objet dans le temps en tant que marque ou signe laissé. Cette idée de pérennité se retrouve dans la norme ISO – TC46<sup>[1]</sup> pour laquelle le document est un « *ensemble formé par un support et une information, généralement enregistrée de façon permanente, et tel qu'il puisse être lu par l'homme ou la machine* ». Dans cette définition, la question du stockage est mise en évidence par la dissociation du document à travers un support et une information tout en tenant compte de la contrainte de lisibilité. Dans un article intitulé « *What is a document ?* » Michael Buckland (Buckland, 1998) présente une synthèse des approches proposées sur le document. Il décrit en particulier l'approche de Paul Otlet pour qui les objets archéologiques ou les œuvres d'art peuvent être assimilés à des documents. Cette notion est reprise par Jean Meyriat qui considère que « *les objets rassemblés dans un musée sont des documents* » (Meyriat, 1978). Le besoin d'étudier un objet peut faire de celui-ci un document. Michael Buckland reprend les propos de Suzanne Briet (Buckland 1995) : pour elle, une antilope peut être considérée comme un document si elle est capturée et placée dans un zoo pour y devenir un objet d'étude. Ainsi un objet réel peut devenir un document s'il reste possible de garantir sa localisation pour la durée de son existence et s'il y a intention d'analyse et d'étude sur cet objet. Partant d'une approche qui consiste à considérer qu'une information sur un support constitue un document, celui-ci devient une trace matérielle de l'information portée et joue alors le rôle de preuve. Le support est ainsi un élément constitutif du document, approche que nous retrouvons avec Bruno Bachimont pour qui un document est « *tout d'abord un objet matériel exprimant un contenu. Il est indissociable d'un support matériel (un écran, une feuille de papier, etc.), support d'inscription où un contenu est exprimé. Ce contenu s'exprime en une forme interprétable (lettres, formes iconiques, etc.) pour un lecteur* » (Bachimont, 1998). La structure physique est liée au mode de stockage et de diffusion, la structure logique quant à elle concerne l'organisation des éléments qui le composent et leur type d'accès. Ces considérations font apparaître rapidement la question de la structuration du document et de sa pérennité. Si le document est indissociable du support matériel alors il faut pouvoir assurer des conditions satisfaisantes d'accès et de consultation de manière stable dans le temps. Cela suppose d'assurer son enregistrement, son stockage, sa gestion, son accessibilité sur des supports fiables et permanents. Le papier reste encore le support traditionnel de référence pour la lecture, mais les coûts de fabrication, d'envoi et de stockage face à l'explosion documentaire sont des problèmes potentiels. De plus, les

éléments tels que les images animées, le son ou les liens entre documents ne sont pas pris en compte. Un document apparaît donc comme une chose matérielle, stockable, définie par un support et un contenu. Traditionnellement, les documents peuvent se présenter sous la forme de livres (ouvrages, encyclopédies, dictionnaires, revues, bibliographies, thèses et mémoires), films, photographies (papier, diapositives), musique, sons, sculptures, peinture. Comme nous l'avons vu avec Suzanne Briet par exemple, nous reconnaissons que la notion de document peut être étendue à tout objet qui peut être étudié. Désormais, cette notion élargie est revisitée du fait du développement des technologies du numérique : nous parlons alors de document numérique. Le réseau thématique pluridisciplinaire 33 du département Sciences et Technologies de l'Information et de la Communication (STIC) du Centre National de la recherche Scientifique (CNRS)[\[2\]](#) a proposé dans un travail collectif de définir la notion de document 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). Trois définitions sont alors proposées et reprises ici. Selon la première, liée à la forme, « *un document numérique est un ensemble de données organisées selon une structure stable associée à des règles de mise en forme permettant une lisibilité partagée entre son concepteur et ses lecteurs* », la seconde, liée au signe, voit dans le document numérique « *un texte dont les éléments sont potentiellement analysables par un système de connaissance en vue de son exploitation par un lecteur compétent* ». La troisième définition, liée à la relation, y voit « *la trace de relations sociales reconstruite par les dispositifs informatiques* » (Pédauque, 2003). Notre point d'entrée, pour reprendre la terminologie utilisée dans le texte collectif de Roger T. Pédauque, se base sur la forme.

### Les caractéristiques du document numérique

Pour le document, le passage à l'état numérique fait apparaître de nouvelles caractéristiques. Nous retiendrons principalement les caractéristiques suivantes : le stockage (la capacité du document numérique à être stocké), la plasticité (le document numérique peut prendre plusieurs formes) et l'association c'est-à-dire la possibilité de relier et d'associer les documents entre eux comme l'avait envisagé très tôt Vannevar Bush avec le « *Memex* » (Bush, 1945). Pour nous, les deux dernières caractéristiques (plasticité et association) peuvent être fusionnées : les possibilités d'association du document numérique découlent de sa plasticité, dit autrement un document numérique peut se présenter sous la forme de plusieurs fragments qui seraient associés. En ce qui nous concerne, nous ajoutons à ces caractéristiques la « programmabilité » du document conséquence des possibilités de génération automatisée ou semi-automatisée. Michael Buckland dans un article intitulé « *what is a digital document ?* » (Buckland, 1998), illustre son approche du document numérique sur la table des logarithmes. Il y a quelques dizaines d'années nous aurions eu recours à une table des valeurs au format papier pour faire nos calculs, cette même table de valeurs constituant un document du fait même de son existence. Il est assez facilement imaginable de penser que la version numérique de cette table serait aujourd'hui représentée sous la forme d'une page web interactive et reposant sur un formulaire et un script implémentant un algorithme de calcul. L'objectif du document (fournir les valeurs de la table) reste identique et le dispositif mis en place (l'algorithme), est une sorte de document dynamique, de document « programmable » ou de document « programmé ».

Dans le contexte du Web, cela se traduit par la possibilité de créer des documents à la volée. Le résultat de l'exécution d'un programme informatique est un document : la page web obtenue à partir d'une ou plusieurs ressources (le programme, la base de données, les éventuelles feuilles de style...). Deux exemples de langages de scripts peuvent être cités ici : PHP[3] et ASP[4]. L'intérêt de ces deux solutions s'explique par l'aspect réellement dynamique apporté au contenu fourni sur le Web. L'exécution se fait du côté du serveur, ce qui accélère les traitements. Ces technologies permettent de dépasser les limites du HTML[5] qui reste statique et permet uniquement de gérer la présentation d'un document. En partant de ce principe, de nombreux sites sur le Web ont basé leur fonctionnement sur la personnalisation, la reconnaissance d'un utilisateur devenant un facteur clé de sa fidélisation. Les sites de type portail offrent ce type de personnalisation. A travers des choix faits par un utilisateur ou à partir d'un enregistrement de ses habitudes le système s'adapte. Les systèmes de personnalisation se sont largement développés depuis le début des années 2000. « *My Yahoo !* »[6] par exemple ou plus récemment « *iGoogle* »[7] proposent une interface où la forme (couleurs, typographies, utilisations d'illustrations graphiques ou non) et le contenu (rédactionnels, liens hypertextuels) correspondent à un profil référencé. Si nous associons, à la production de briques d'information du document en unités, la gestion du profil de l'utilisateur, il est alors envisageable, selon des règles à définir, de produire un document personnalisé. Ces possibilités se multiplient avec l'avènement du « Web 2.0 » dont nous reprenons ici les principes essentiels énoncés par Tim O'Reilly dans son article « *What is Web 2.0 ?* » (O'Reilly, 2005).

- Le web est envisagé en tant que plate-forme et le travail se fait en mode collectif : ceci se traduit par un enrichissement continu par les utilisateurs eux-mêmes
- La puissance se situe au niveau des données : la génération automatisée des documents numériques s'appuie sur des bases de données
- La fin des cycles de « release » : la mise à jour se fait en temps réel, les utilisateurs deviennent des co-constructeurs, des co-développeurs
- Des modèles de programmation légers : la simplicité favorise le déploiement des services et leur appropriation par les acteurs
- Le logiciel se libère du PC : de plus en plus de nouveaux appareils peuvent utiliser la plate-forme constituée par le Web
- Enrichir les interfaces utilisateur : des interfaces de plus en plus familières et riches favorisant la personnalisation.

Ces nouvelles caractéristiques du document dans le contexte numérique (stockage, plasticité et programmabilité) ouvrent de nouvelles possibilités pour les utilisateurs qui deviennent des producteurs ou des assembleurs faisant appel à divers contenus ou services accessibles sur le Web.

## **La normalisation au service de la production et de la diffusion des documents numériques**

## Les normes et les standards pour le document numérique

Les normes *de jure* (normes formelles, officielles) et les standards *de facto* (normes officieuses) se complètent. Le réseau Internet s'est basé sur des standards de fait qui sont devenus des normes. Un standard est un ensemble de recommandations émanant d'un groupe représentatif d'utilisateurs comme l'IETF (*Internet Engineering Task Force*), le W3C (*World Wide Web Consortium*) ou encore le DCMI (Dublin Core Metadata Initiative)[8]. Une norme est un ensemble de règles de conformité, édictées par un organisme de normalisation au niveau national, européen ou international. Pour la France, l'AFNOR est l'organisme chargé des questions liées à la normalisation, au niveau européen c'est le Comité Européen de Normalisation (CEN[9]) alors qu'au niveau mondial l'ISO[10] joue ce rôle. Une norme est le plus souvent le résultat d'un compromis technique et/ou commercial entre les différents acteurs pour un domaine donné et apparaît comme « *un document déterminant des spécifications techniques de biens, de services ou de processus qui ont vocation à être accessibles au public, résultant d'un choix collectif entre les parties intéressées à sa création, et servant de base pour la solution de problèmes répétitifs* » (Lelong, Mallard, 2000). En reprenant la typologie des normes énoncées par Benoît Lelong et Alexandre Mallard nous pouvons distinguer :

- des normes de définition (terminologies, unités de mesure, métrologie),
- des normes d'interchangeabilité (pièces détachées),
- des normes de compatibilité et d'interopérabilité,
- des normes de qualité (seuils de qualité, de performance).

Parmi ces différentes fonctions pouvant être associées, nous soulignons particulièrement la notion d'interopérabilité. Pour les auteurs, « *les fonctions de compatibilité et d'interopérabilité renvoient à la nécessité de garantir l'interfonctionnement des objets techniques dans les systèmes qu'ils composent, ou dans les environnements avec lesquels ils sont amenés à interagir* » (Lelong et Mallard, 2000). Dans le domaine du document numérique, cette question d'interopérabilité est essentielle : le document doit être lisible par n'importe quel type de plate-forme aussi bien à un niveau matériel qu'à un niveau logiciel.

Notre objectif n'est pas de présenter une liste exhaustive des normes disponibles mais il consiste plutôt à indiquer celles qui, selon nous, s'inscrivent directement dans la production du document numérique. Nous présentons ici celles qui participent à la normalisation du codage des caractères et à la structuration du contenu. Le codage des caractères, au niveau le plus élémentaire, est né de la nécessité de communiquer des données entre systèmes informatiques. Différentes normes ont été mises au point, en partant d'ASCII[11] et de ses dérivés, jusqu'à Unicode. Par extension les formats de fichiers qui stockent ces suites de caractères doivent également être normalisés (un fichier produit à l'aide d'un outil logiciel doit pouvoir être lu dans d'autres configurations). Chaque langue possède des lettres et des signes de ponctuation. Pour pouvoir traiter ces multiples caractères, des tables (de caractères) ont été constituées, attribuant un nombre (code) à chaque caractère. Il est facile d'imaginer que la multiplicité des langues, des systèmes informatiques, rend cette gestion de caractères complexe. Cette complexité est surmontée par Unicode dans un contexte où les échanges se sont multipliés entre les postes individuels et les serveurs répartis sur une

échelle mondiale. Les navigateurs web doivent pouvoir afficher les caractères diffusés par les serveurs distants. Cela est possible, car en effet, « *Unicode spécifie un numéro unique pour chaque caractère, quelle que soit la plate-forme, quel que soit le logiciel, quelle que soit la langue* » [12]. La lecture d'un document électronique dépend du format du document et nécessite l'outil logiciel ad hoc. Dans le contexte Internet est apparu le type MIME (*Multipurpose Internet Mail Extensions*) [13], standard proposé initialement par les laboratoires Bell Communications en 1991, et qui a pour but comme son nom l'indique d'étendre les possibilités du courrier électronique. Ces extensions prévoient l'insertion d'éléments de type image, son, vidéo dans un courrier. Le type MIME est également utilisé dans la communication de pages web simples ou complexes en indiquant au navigateur la nature de l'objet (format) pour permettre la consultation de ce dernier. Un contenu décrit selon le type *video/msvideo* signifie que le fichier fourni est de type vidéo et le navigateur devra en fonction de cette information « passer la main » à l'outil de visualisation configuré sur la machine de l'internaute. Ceci permet de réduire le risque logique lié à la « lisibilité » de la production. En effet, selon les versions de systèmes d'exploitation voire dans la série des versions d'un même produit, des incompatibilités apparaissent (abandon des formats par exemple). Cela signifie qu'il faudrait sauvegarder le document lui-même bien évidemment mais aussi tout l'environnement logiciel qui a permis sa production. Ceci est pris en compte à la Bibliothèque Nationale de France, mais la multiplication des environnements et des versions des logiciels est source de nombreux problèmes. Dans le contexte du Web, la production du document numérique s'appuie principalement sur les langages de marquage qui doivent nous permettre de nous affranchir du problème de « lisibilité » que nous venons d'évoquer.

Un texte n'est pas qu'une suite de caractères. Un document est pourvu d'une structure logique liée à l'organisation des éléments qui le composent. Par exemple, une revue scientifique est organisée en articles eux-mêmes structurés en chapitres ou paragraphes. Les technologies du XML permettent de coder cette structure logique pour une communication de ces documents à travers le réseau Internet et les systèmes de traitement de l'information. Parmi les possibilités de marquage nous citons les trois procédés les plus connus (dans l'ordre de leur apparition) : SGML, HTML et XML. Tout d'abord, à l'origine, apparaît SGML (*Standard Generalized Mark-up Language*) norme internationale de description de la structure et du contenu de différents types de documents électroniques depuis 1986 (norme ISO 8879). Sa complexité n'a pas permis de développer aisément des produits accessibles. Quant à HTML (*HyperText Mark-up Language*), mis au point au début des années 1990, il s'agit d'un procédé centré sur la présentation d'un document adapté au Web mais qui se base sur une collection de balises relativement réduite. XML (*eXtensible Mark-up Language*) [14], standard du W3C (*World Wide Web Consortium*) depuis 1998, allie la richesse sémantique du SGML (les balises ont un sens) et les possibilités de présentation offertes par le HTML. XML, format public, accessible à tous, donne également la possibilité d'échanger, de communiquer des documents entre des systèmes hétérogènes. Avec XML, il est possible de séparer le contenu d'un document de sa présentation et de définir son propre langage pour décrire ce contenu. Le document ne contient aucune information concernant l'affichage, c'est sa feuille de style qui définira la présentation en fonction du support (écran d'ordinateur, écran téléphone mobile, imprimante...). Les feuilles de style CSS (*Cascading Style Sheets*) et les feuilles de style XSL (*eXtensible Style sheet Language*) utilisables avec XML offrent plus de possibilité qu'avec HTML. Le XSL est composé de deux

langages dont XSL *Transform* (XSLT), langage qui permet de transformer les documents XML en d'autres types de documents ou en d'autres documents XML. Ces feuilles de styles dépassent le cadre de la présentation en permettant par exemple de filtrer les contenus (utilisation de conditions) donc de personnaliser le document numérique.

### Les normes peuvent-elles être une réponse à la fragmentation et à la dispersion de l'information ?

Les caractéristiques du document numérique, la situation d'explosion documentaire numérique nous invitent à une réflexion sur une régulation normative au service de la gestion des contenus (production, indexation, conservation...). Les besoins en termes de structuration de données, de modalités de codification semblent évidents pour assurer les conditions de l'échange mais on ne peut se limiter à envisager la question de la fragmentation uniquement sur le plan technique. Il est vrai que le passage du cadre « traditionnel » au cadre numérique a modifié la donne. D'un côté des normes qui peuvent faciliter l'échange. D'un autre côté nous pouvons observer des utilisateurs faisant preuve de plus en plus d'autonomie dans leurs pratiques de production et de diffusion de contenus informationnels. Nous avons vu précédemment les caractéristiques du Web dit « 2.0 » dont nous pouvons retenir qu'il rend les interfaces bien plus accessibles aux utilisateurs. Du statut de lecteur, l'internaute est devenu producteur mais nous pourrions dire que désormais il ressemble de plus en plus à un « assembleur ». Il reste producteur de fragments documentaires numériques (une brève sur son blog par exemple) mais fait appel à d'autres ressources (afin d'assembler sa page personnelle par exemple). Les technologies combinées sous l'appellation Ajax et s'appuyant sur la présentation (XHTML et feuilles de style CSS), l'échange de données avec XML (et les feuilles de style et de transformation XSLT), la récupération des données provenant d'autres serveurs, la gestion du document avec DOM (Document Object Model) et Javascript permettent techniquement ce « travail d'assemblage ». Leur particularité essentielle est d'avoir simplifié leur mise en place et leur réutilisation de plus en plus de manière transparente pour l'utilisateur. Tout se passe dans le back-office et celui-ci se trouve de plus en plus en arrière plan. Le Web 2.0 ne s'occupe pas d'interopérabilité mais bel et bien de réutilisabilité facilitée, l'objectif des producteurs de services étant la récupération de données en les centralisant de plus en plus. Cette centralisation est masquée par cette diffusion / dispersion sur un nombre sans cesse croissant de sites.

Mais cette redondance de services se fait sans aucun respect des normes, les possibilités technologiques et les pratiques individuelles amplifient le problème de la fragmentation et de la dispersion de l'information. Nombreux sont ceux qui désormais utilisent plusieurs comptes sur différents services web (webmail, pages personnelles, blogs, wikis, bureau virtuel...), utilisent plusieurs machines (avec plusieurs adresses électroniques)...

Les questions de recherche suivantes restent d'actualité :

- Comment assurer l'identification et la localisation d'un document numérique (comment retrouver un document numérique dans un ensemble de plus en plus vaste ?)



- Comment limiter la prolifération et la volatilité du document numérique (plusieurs versions d'un même document situées sur des adresses différentes, altération des versions, disparition de la ressource ...)
- Quel est le devenir des nouveaux modes de diffusion du document numérique (quelles sont les sources à retenir ? avec quelle validité ? avec quelle fiabilité ?)

Dans ce contexte numérique à évolution constant, les normes peuvent-elles aider à apporter quelques éléments de réponse ? Rien ne semble acquis. Une norme fait l'objet d'un consensus entre tous les acteurs, le résultat qui en découle est global pour satisfaire les attentes de tous. De plus la participation de chacun est volontaire. Ces aspects marquent bien le caractère général et global des normes qui doivent être applicables pour tous. Mais c'est une des raisons pour lesquelles une norme apparaît souvent comme lourde et inadaptée lorsqu'il est envisagé de passer à son application. En effet, la construction d'une norme est une opération relativement longue et nécessitant de longs et nombreux échanges qui doivent s'inscrire dans un schéma précis. Ce travail long est également répété dans le temps car les normes sont revues périodiquement pour faire face à l'évolution (technologique, juridique, économique...). La règle générale prévoit de ne pas dépasser cinq ans avant révision. Nous le comprenons aisément ce processus ne semble pas adapté au contexte du web, contexte à évolution rapide et ce dans une accélération constante.

Deux exemples permettent d'illustrer nos propos. Le premier exemple est celui du Dublin Core, le second exemple repose sur les flux d'information RSS.

Tout d'abord le Dublin Core<sup>[15]</sup> créé en 1995, devenu Dublin Core Metadata Initiative (DCMI), puis norme ISO 15836 (adoptée en 2003) consiste en un effort de normalisation au niveau international pour la définition d'éléments bibliographiques. Ces éléments sont aussi appelés métadonnées<sup>[16]</sup> : données structurées sur d'autres données. Cet ensemble appelé DCMES<sup>[17]</sup> comprend quinze éléments (optionnels et pouvant être répétés) : titre, auteur, description, éditeur, autres collaborateurs, date, type de ressource, format, code d'identification de la ressource, source, langage, relation, portée, gestion des droits. La liste de base des éléments peut être utilisée pour des documents numériques en ligne ou pour tout autre type de document, voire tout objet. Par exemple, la page Web peut ainsi contenir elle-même sa « carte d'identité » (fournie dans l'en-tête du document HTML). Ceci pourrait permettre de faciliter son traitement par les outils d'indexation des moteurs de recherche. La norme Dublin Core, du fait de sa simplicité (nombre réduit d'éléments), est souvent prise comme base de travail pour construire d'autres normes de description de documents. Son adoption par exemple pour décrire les pages web reste encore bien réduite par rapport aux métadonnées descriptives (keywords, description...) déjà existantes.

Autre situation avec RSS acronyme traduit par « *Rich Site Summary* » ou « *Really Simple Syndication* »<sup>[18]</sup> qui, selon les approches, permet d'échanger du contenu entre sites web. Il est ainsi possible de lire et d'agrégier dans une page web des « news », des « podcats », des « vidéocasts » provenant de différentes sources. Nous sommes alors dans un système de type pull-push média alliant demandeurs et diffuseurs. Les articles récents dans un site contributif peuvent ainsi être accessibles directement depuis une autre page (devenue dynamique) du même site ou d'un autre site. Le concepteur d'une page peut décider d'incorporer des données distantes produites et mises à jour par d'autres. En retour, ce



concepteur, producteur de contenu peut à son tour alimenter de façon régulière un canal fourni sous la forme d'un flux XML. Basés sur le principe de RSS, les weblogs ont fait leur apparition. Le weblog est un site qui peut être collaboratif, conçu pour recevoir des « brèves » partageables avec d'autres[19]. Pour incorporer son contenu il suffit de pointer sur un document XML du site émetteur-producteur[20] et de l'incorporer en fonction d'une feuille de style à définir.

Mais ce format a fait l'objet de nombreuses querelles, plusieurs acteurs se disputant la gestion et la formalisation de celui-ci[21].

Pour ces deux exemples, nous le voyons, la situation de déploiement ou de généralisation n'est pas aisée. Les acteurs du Web 2.0 sont plus soucieux de la réutilisation que de répondre aux problématiques d'identification formelle et de localisation pérenne de l'information. Nous n'aborderons pas ici les évolutions liées au web sémantique qui permet de définir des liens entre des ressources considérant que nous pouvons y voir le signe de nombreuses évolutions en cours et à venir et qu'elles constituent de véritables pistes de recherche encore à explorer.

## Conclusion

Nous poursuivons l'objectif de contribuer à la réflexion en cours sur le document numérique en particulier sur les problèmes liés à sa mise en jeu. Il peut être reconnu que notre société est désormais fortement marquée par le recours quotidien à la plate-forme web sur laquelle la logique de simple lecture laisse la place à une logique de production. Cette production évolue et se réduit souvent en une simple réutilisation de fragments documentaires numériques par des jeux d'assemblage dans des interfaces de plus en plus accessibles et favorisant l'autonomie. La fragmentation et la dispersion de l'information découlent à la fois des caractéristiques du document numérique et des possibilités technologiques proposées à tous dans l'environnement web. Dès lors, la production et la diffusion de document numérique sous une forme normalisée peuvent-elles répondre (en partie au moins) aux problématiques de fragmentation et de dispersion évoquées ici ? Nous avons vu que le cycle relativement long de construction des normes, leur formalisme, leur approche globale ... ne semblent pas toujours adaptés au Web dont les évolutions sont rapides et dans un environnement fortement concurrentiel. L'accélération dans laquelle nous nous trouvons, les évolutions des applications amplifiées par une diffusion de plus en plus large des technologies du numérique dans notre société, nécessitent une prise de recul. Nous avons toutefois essayé d'apporter quelques éléments d'éclairage sur ces questionnements difficiles du fait de leur évolutivité et de leur actualité sans cesse renouvelées.

## Bibliographie

Bachimont B., 1999, *De l'hypertexte à l'hypotexte : les parcours de la mémoire documentaire*, In : *Mémoire de la technique et techniques de la mémoire*, sous la direction de Charles Lenay et Véronique Havelange, Technologies, Idéologies, Pratiques, ERES, pp. 195-225

Buckland M.K., 1995, "The centenary of "Madame Documentation", Suzanne Briet, 1884-1989", *Journal of the American Society for Information Science*, Vol. 46, n° 3, pp. 235-237.

Buckland M.K., 1997, "What is a document ?", *Journal of the American Society for Information Science*, Vol. 48, n° 9, pp. 804-809.

Buckland M.K., 1998. "What is a "digital document" ?", *Document numérique*, n° 2, pp.221-230.

Bush V., 1945, "As We May Think", *The Atlantic Monthly*, Vol. 176, n° 1, pp.101-108.

Chartron G., Guyot B., Lafouge T., et al., 2002, "Le document numérique : un objet fédérateur de recherche en sciences de l'information", *Revue Documentaliste - Sciences de l'information*, Vol. 39, n° 6, pp.298-305.

Lainé-Cruzel S., 2004, "Documents, ressources, données : les avatars de l'information numérique", *Revue I3 Information – Interaction – Intelligence*, 4(1), pp. 105-120.

Lelong B., Mallard A. (coords). 2000, *La fabrication des normes*, Paris, Hermès Science Publications.

Meyriat J., 1978, *De l'écrit à l'information : la notion de document et la méthodologie de l'analyse du document*, In : Inforcom 78, Société française des sciences de l'information et de la communication, premier congrès Compiègne. Paris : SFIC 1978, Vol. 1, pp. 23-32.

Meyriat J., 1981, *Document, documentation, documentologie*, In : L'écrit et le document, Schéma et schématisation, n°14, pp. 51-63.

O'Reilly T., *What Is Web 2.0 ?*, in : O'Reilly, [en ligne], septembre 2005 [consulté le 18 avril 2008]. Disponible sur :

<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

Pédauque R.T., 2003, "Document : forme, signe et médium, les re-formulations du numérique", [en ligne], juillet 2003 [consulté le 18 avril 2008]. Disponible sur :

[http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/05/11/index\\_fr.html](http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/05/11/index_fr.html)

---

[1] ISO International Organization for Standardization, TC 46 est le comité technique (Technical Committee) pour les normes en matière d'information et de documentation

[2] RTP 33 : documents et contenu. Création, indexation, navigation (RTP DOC), <http://rtp-doc.enssib.fr> (consulté le 21 avril 2008)

[3] PHP : Personal Home Page créé par Rasmus Lerdorf en 1994, devenu PHP Hypertext Preprocessor en 2000

[4] ASP : Active Server Pages (solution Microsoft), plus un procédé qu'un langage (langage utilisé : Vbscript)

[5] HTML, HyperText Markup Language

[6] Espace personnalisé sur Yahoo, <http://my.yahoo.com> (consulté le 21 avril 2008)

[7] iGoogle, depuis avril 2007, <http://www.google.fr/ig> (consulté le 21 avril 2008)

[8] Etude des normes, <http://www.educnet.education.fr/tech/normes/0402.htm> (consulté le 21 avril 2008)

[9] CEN, Comité Européen de Normalisation, <http://www.cenorm.be> (consulté le 21 avril 2008)

[10] ISO, International Standard Organization, <http://www.iso.ch> (consulté le 21 avril 2008)

[11] ASCII, American Standard Code for Information Interchange, <http://www.asciitable.com/> (consulté le 21 avril 2008)

[12] Unicode, <http://www.unicode.org/standard/translations/french.html> (consulté le 21 avril 2008)

[13] <http://www.iana.org/assignments/media-types/> (consulté le 21 avril 2008)

[14] Spécification XML-Schema, <http://www.w3.org/XML/Schema> (consulté le 21 avril 2008)

Spécifications XSL, <http://www.w3.org/TR/WD-xsl> (consulté le 21 avril 2008)

Spécifications XSLT, <http://www.w3.org/TR/WD-xslt> (consulté le 21 avril 2008)

[15] <http://dublincore.org> (consulté le 25 avril 2008)

[16] metadata, structured data about data, <http://dublincore.org/resources/faq/#whatismetadata> (consulté le 25 avril 2008)

[17] DCMES : Dublin Core Metadat Element Set

[18] (ou encore RDF Site Summary)

[19] Par exemple <http://google.blogspot.com> (consulté le 21 avril 2008)

[20] Exemple de contenu incorporable : <http://google.blogspot.com/index.xml> (consulté le 21 avril 2008)

[21] Les dates essentielles de RSS : <http://opikanoba.org/xml/040315> (consulté le 21 avril 2008)