✪ LA NORME DE BALISAGE DES DOCUMENTS SGML ET SON APPLICATION DANS LES DOMAINES DE L'EDITION ELECTRONIQUE ET DES BIBLIOTHEQUES

Par Yahia BAKELLI
Chef de projet au Département DST
CERIST (centre de recherche sur l'information scientifique et technique)

ABSTRACT

Cet article est une synthèse bibliographique sur la norme de structuration logique des documents SGML (Standard Generalized Markup Language). Nous introduisons cette synthèse par une présentation des normes ISO 8879: 1986 et ISO 12083: 1994 et une explication du principe de base de ce format. A l'origine, cette norme a été conçue pour répondre aux besoins des éditeurs et des organismes devant manipuler des documents techniques. Il s'agit essentiellement des publications électroniques sur supports multimédias et sur réseaux (internet, notamment). Les périodiques électroniques, la documentation technique et les livres électroniques constituent les formes privilégiées de cette nouvelle technologie. Cette norme a, par la suite, été élargie au domaine des bibliothèques grâce à ses avantages en terme de manipulation des documents et de la possibilité d'intégrer, à la fois, l'information primaire et l'information secondaire dans une même structure. Les diverses formes d'exploitation de cette norme par les professionnels de l'information sont exposées. Un intérêt particulier est accordé aux volets catalogage et diffusion des publications électroniques. Un des facteurs ayant favorisé le développement de ce format réside dans le fait qu'il a été supporté par des outils informatiques (éditeurs,

compilateurs et visualiseurs) conçus par des grands constructeurs (Jouve, Adobe, Softquad, etc.). Il a également pour avantage le fait qu'il a été adopté par de grands organismes américains (DoDefense, AAP, LC, OCLC), européens (l'Office des publications de la CEE, l'Office européen des brevets), et internationaux (Reed Elsevier, Springer Verlag, spatiale, l'ISO et la World Intellectual Property Organization, etc). Cependant ce format a ses inconvénients. Sont alors listées les avantages et les inconvénients de SGML, selon les avis d'experts. L'article se termine par un exposé des tendances et des pespectives de ce formats, notamment par rapport au dilemme SGML/HTML et le nouveau venu XML.

Mots clés: SGML/ édition électronique/ metadata/ HTML/ XML /Internet / Structuration logique / documents électroniques.

1 - Définition du format SGML (Standard Generalized Markup Language):

La norme SGML (Standard Generalized Markup Language) a une assez longue histoire: il faut remonter jusqu'en 1960 à Charles Goldfarb et son équipe qui, chez IBM, créa une méthode de description de documents appelée GML (Generalized Markup Language). Le concept a par la suite mûri pendant une vingtaine d'années pour aboutir à une norme adoptée par l'ISO sous le numéro ISO 8879 [25].

L'ISO 8879: 1986 - Traitement de l'information: systèmes Bureautiue - langages normalisé de balisage généralisé (SGML) ou la norme française NF EN 28879-1990: Systèmes Bureautique - Langage normalisé de balisage généralisé (SGML) est une norme de codification de la structure logique des documents révisables. En association avec toute une famille de normes, SGML permet l'échange de documents électroniques associés avec leur structure logique, leur format, les informations sur la structure physique. Sont ainsi autorisés l'affichage ou l'impression, de manière indépendante des logiciels de traitement, des ordinateurs ou de leurs périphériques et indépendamment des applications qui effectuent l'échange.

Un ensemble de normes internationales complète SGML, de manière à couvrir la totalité des traitements de la chaîne professionnelle de production des documents. Ce sont: SDIF (SGML Document Interchange Format), DSSSL (Document Style Symantics and Specifications Language), SPDL (Standard Page Description Language), Glyph Fonts Definition and Identification, HYTIME

(Hypertext Based Structuring Language) et en fin la norme HTML (Hypertext Markup Language). Cette dernière est la plus connue des dérivés SGML. Elle es utilisé pour le codage de la structure des documents manipulés par les serveurs www sur Internet. C'est une application de SGML avec une DTD particulière.

L'ISO 12083: 1994 - préparation et balisage des manuscrits électroniques

Elle est issue du projet « Electronic Manuscript Project » conduit de 1983 à 1986 aux USA. C'est une application de l'ISO 8879 fixantt le codage de la structure logique des monographes, des publications en série, des articles de périodiques et des formules mathématiques. Elle s'adresse initialement aux maisons d'éditions et à leurs auteurs. Mais elle concerne également les intermédiaires de la diffusion des publications qui font des traitements à valeur ajoutée. Elle est compatible avec le Braille et prend en compte les évolutions vers l'application de la norme HyTime.

Le principe de base de SGML:

L'idée à la base de SGML est de permettre une représentation logique d'un document et des différentes composantes de son contenu, plutôt qqu'une description de l'aspect qu'il doit présenter une fois imprimé. SGML permet de marquer chaque élément du texte d'après son rôle dans l'ensemble du document; si un mot représente un nom propre, il sera désigné « nom propre » et non signalé comme devant être écrit en italique ou en caractères gras, si une phrase représente un titre de niveau n, les balises indiqueront simplement cette particularité, au lieu de préciser quelque chose comme « fonte Roman 14 points gras », etc. [25]

SGML est donc une méthode universelle de balisage de la structure logique d'un document. La structure d'une classe de documents est déclarée, en langage SGML, dans une définition de type de document (DTD). Des logiciels « parseurs » permettent de contrôler que le balisage d'un document particulier appartenant à la classe considérée (appelé « document instance ») est conforme à la DTD qui définit sa structure.

Les principales DTD Standards sont celles de:

- L'ISO 12083 pour les ouvrages, publications en série et les articles de périodiques.

- La DTD MAJOUR (Modular Applications for Journals) de Springer Verlag sur l'en-tête des articles relevant des domaines de STM (sciences, techniques et medecine), adoptée par l'association des éditeurs européens.

- La DTD ATA pour les documents techniques de l'aéronautique civile.

- La DTD 12008 pour la documentation technique de l'industrie automobile.

- L'EAD (Encoded Archival Description) est l'une des plus fameuses DTS SGML. Elle est destinée à la structuration et au repérage des documents du type « archives ».

Le langage SGML permet de coder la structure des documents ayant des contenus composites: texte, graphique, image mais aussi des structures telles que celles des formules mathématiques, des formules chimiques et des tableaux. SGML permet également la notation des liens hypertexte.

Le document SGML est révisable, il peut être corrigé, mais à jour en utilisant un traitement de texte SGML. Le document SGML est donc potentiellement réutilisable sur un système différent de celui qui a servi à sa création, soit à l'aide d'outils SGML, soit dans des formats plus simples tels que les traitements de texte.

L'autre avantage de SGML c'est qu'il permet des sorties sur plusieurs supports: papier, télématique, Internet, CD-ROM et même en braille.

2 - Les domaines d'application de la norme SGML

Les domaines d'utilisation de ce format concernent principalement les secteurs de l'édition, à savoir:

- La Bureautique;
- L'édition générale;
- L'édition d'encyclopédies;
- La documentation technique et les manuels de l'industrie;
- La réalisation de livres électroniques sur CD-ROM;
- Les publications électroniques sur Internet.

SGML peut également être adopté lors de la conversion

rétrospective des documents imprimés, par le biais de la reconnaissance optique des caractères. Avec ce format SGML, cette opération de conversion ne se limite plus à une simple récupération de chaîne de caractères, mais à une reconsitution automatique de la structure logique du document numérisé. Cette technique repose sur le principe que l'apparence physique d'un document peut être exploitée afin de déduire sa composition logique. Une telle reconstitution signifie que le document en question peut être facilement traité au sens documentaire du terme.

Grâce au principe de structuration simultanée du document et de sa représentation bibliographique (métadonnée) le caractérisant d'une part et aux utilitaires le supportant d'autre part, l'usage du format SGML se trouve étendu aux activités relevant du traitement, stockage et diffusion de l'information et de la documentation:

- L'alimentation des bases de données;
- La construction des catalogues de bibliothèques;
- L'échange de données bibliographiques;
- L'indexation et la recherche documentaire;
- La diffusion électronique des documents.

Plusieurs organismes ont investi et ont adopté le format SGML en se mettant sur des DID standard. Citons les cas de:

- L'Association of American Publishers, le Department of Defense (dans le cadre du projet CALS: Computer Aided Acquisition and Logistic Support) aux USA.

- Des éditeurs internationaux: Reed Elsevier, Springer Verlag, Kleuwer, ...

- L'EDF, l'Aérospatiale, Syndicat National de l'Edition et du Cercle de la librairie, en France.

- Institutions régionales et internationales telles: l'Office des publications de la CEE, l'Office européen des brevets, l'ISO et la World Intellectual Property Organization, etc.

- Des bibliothèques et institutions documentaires comme l'OCLC, la LC, l'INIST, la BIEF, et plusieurs bibliothèques américaines et ce dans le cadre de projets de GED, d'échanges d'informations bibliographiques et de fourniture électronique de documents.

2.1 - L'édition électronique

La norme SGML a été initialement conçue pour une production performante des documents textuels plus particulièrement la saisie en vue d'une meilleure impression.

Les principaux types de documents concernés par l'encodage SGML sont les:

- Les publications scientifiques (périodiques, rapports, ...);
- La documentation technique et juridique;
- Les encyclopédies et les livres électroniques, sur CD-ROM;
- Les publications sur INTERNET.

2.1.1 - Les périodiques électroniques

Les nouvelles technologies de l'information transforment, depuis la fin des années 80, la structure de la chaîne éditoriale des périodiques scientifiques et techniques.

Du point de vue technique, les changements se sont opérées au niveau des modes de:

a) Rédaction des manuscrits d'articles par les auteurs;

b) Réception et collecte des articles par l'éditeur;

c) Lecture et sélection des articles;

d) Constitution et livraison du fascicule;

e) Consultation et exploitation des articles par les lecteurs.

L'introduction de SGML en tant que mode de structuration et de rédaction des articles influe sur le mode de réalisation des autres phases.

Moyennant des utilitaires dédiés SGML, un auteur peut faire la saise de son manuscrit de la même façon qu'avec un logiciel de traitement de texte classique (ex: Word6, WordPerfect, ...) ou d'un outil plus élaboré (ex: TEX pour les formules mathématiques);

Un éditeur peut recevoir le manuscrit selon un mode traditionnel (version papier) ou en version électronique, s'il dispose d'équipements adaptés. Selon la nouvelle configuration, l'éditeur est supposé avoir adopté un format logique unique permettant de maintenir une base de données de tous les articles. En plus de la lecture et de la correction des contenus, taches qu'incombe au comité de lecture, l'éditeur vérifie

que la composition de l'article est conforme aux règles de présentation propres à la revue (communément appelées « recommandations aux auteurs »). Ces dernières font justement l'objet de DTD SGML. Dans le passage qui suit nous reprenons la chronologie de l'exploitation de ce format à des fins éditoriales (présentation des articles de périodiques) telle que retracée par C. Lupovic[18]: « Les éditeurs ont d'abord utilisé des codifications de structuration qui leurs sont propres (souvent en relation avec leurs imprimeurs et les matériels informatiques et logiciels utilisés). Puis l'époque de la normalisation est arrivée, pour répondre à un besoin d'échanges. La normalisation des différents formats correspondant aux étapes de la chaîne de production appartient à la famille des normes associées à SGML. Parallèlement, une discussion entre les grands éditeurs (publiant dans les mêmes domaines) a été initiée de manière à inclure dans le format logique les règles de rédaction des articles scientifiques et à s'accorder sur un format SGML commun et traduisant ces règles.

Cette normalisation du format de l'article STM a tout d'abord fait partie de l'Electronic Manuscript Projet, conduit de 1983 à 1986 entre divers partenaires de la chaîne de production et de distribution, tels que les éditeurs représentés par l'American Association of Publishers (AAP), les bibliothèques, avec la participation de la LC et de la NLM et des banques de données avec Chemical Abstracts et Medline. Cette démarche subventionnée par le CLR (Council on Library Ressources) avait pour objectif de mesurer l'utilisation possible de SGML dans la chaîne allant de l'auteur à l'utilisateur, en passant par l'éditeur, et par les intermédiaires de la diffusion telles que les banques de données ou les bibliothèques. Une norme américaine a finalement été élaborée et introduite comme base de la norme internationale ISO 12083: Préparation et balisage des manuscrits électroniques, publiée en 1994.

Les éditeurs européens du secteur STM ayant travaillé à l'application du standard AAP ont distingué deux parties dans la structure logique d'un article:

- L'en-tête comprenant le titre, les auteurs et leurs affiliations, la notion de congrès, les mots clés, le résumé, c'est-à-dire toute l'information descriptive et signalétique de l'article;

- Le corps de l'article, c'est-à-dire toute la partie rédactionnelle.

Leurs travaux ont débouché sur Majour (Modular Application for Journals): une structure logique (l'en-tête de l'article) commune écrite en SGML et publiée en 1991 par l'éditeur Springer Verlag. Cette

structuration met les éditeurs en situation de fournir sous forme électronique la description signalétique de leur articles telle qu'on la retrouve dans les bases de données. C'est une forme de catalogage à la source, dans un format pivot pouvant être balancé en formats: Marc, CCF, Medline, Current Contents, etc. Les discussions relatives au corps de l'articles, c'est-à-dire une écriture SGML des règles de rédaction n'ont pas abouti. Ainsi, chaque éditeur convertit séparément son savoir-faire en SGML.

Aujourd'hui tous les éditeurs internationaux des revues STM sont en train d'appliquer la famille des normes SGML à leur chaîne de production et effectuent même, dans certains cas, une conversion rétrospective des revues qui ne sont pas encore produites en SGML ».

Deux exemples:

1 - TULIP

TULIP (The University Licensing Programe), 1992-1995, est un programme de test de diffusion de journaux sous forme électronique entre l'éditeur Elsevier Science BV/Pergamon et 15 collèges et universités des USA, dont 9 sites pilotes et 6 sites observateurs.

L'expérience porte sur 42 titres de périodiques auxquels les partenaires sont abonnés en version papier. L'abonnement est également fourni sous forme électronique, deux fois par semaine, avec la possibilité de définir un profil sur le type de document fourni qui peut comporter:

- La description bibliographique de chaque article correspondant aux informations de l'en-tête de l'article (auteurs, affiliations, titres, mots clés des auteurs, résumés, bibliographies) dans un format texte présenté;
- L'image en Fac-similé de la totalité de l'article;
- Le fichier ASCII de la totalité de l'article;

La version électronique est diffusée par Information Ing. sur réseau Internet. De plus, un article peut être commandé isolément avec un profil différent de celui de l'abonnement.

Pour créer ces documents électroniques l'éditeur opère une conversion rétrospective de la forme papier des journaux concernés en attendant d'avoir migré pour chacun d'eux vers une production fondée sur SGML comme format pivot. Les articles sont numérisés et les images sont traitées par un OCR non corrigé pour obtenir le fichier ASCII.

Le projet est accessible sur Internet via l'url: **Erreur! Source du renvoi introuvable.**

2 - APPLIED PHYSICS LETTERS ONLINE (APL):

APL est une revue hebdomadaire éditée par l'American Institue of Physics (ISSN 1077-3118). Elle est diffusée par OCLC, sous forme d'une base de données accessible en réseau (Internet, OCLC, Compusery) et ce depuis Janvier 1995.

L'interface utilisée pour la consultation de la base et la lecture des articles est nommée GUIDON. Développée par OCLC, cette interface est basée sur le format SGML. Elle permet de faire une recherche d'articles par mots clés ou en précisant le champ d'interrogation. Guidon offre également la possibilité de naviguer à l'intérieur de l'article et ses différentes parties (figures, tableaux, notes, bibliographies, etc.) en mode hypermedia. Le téléchargement des articles sur micro, peuvent être fait en trois formats possibles: ASCII, SGML ou en format spécifique à Guidon. Un savoir faire éditorial et un savoir faire documentaire sont ainsi combinés pour satisfaire au mieux les requêtes des utilisateurs.

2.1.2 - La documentation technique et les livres électroniques

Concevoir une base de données sur la base de SGML, permet une meilleure productivité au niveau de la saisie et de la mise à jour et surtout la possibilité de réaliser des produits dérivés à caractère multimédia. C'est ce qui explique le fait que SGML soit de plus en plus utilisé dans le domaine des ouvrages juridiques et de la documentation technique.

Quelques exemples français:

- Le CD-REEF: Un CD-ROM édité par le CSTB (Centre Scientifique et Technique du Bâtiment) et diffusant des règles techniques dans le domaine de la construction et du génie civil.

- Le CD-ROM ROMARIN: Registre International des Marques géré par l'OMPI (Organisation Mondiale de la Propriété Intellectuelle). L'interface de consultation et de recherche conçu pour ce CD a été développé par Jouve SI, à partir de son logiciel GTI. Ce dernier est composé de deux modules « GTI PubMakder » et « GTIPubUser ». Le premier permet la préparation des données structurées SGML. Le second permet leur consultation.

Le CD-ROM de la documentation technique de la société Aérospatiale et celui du CFMI (3ème constructeur mondial des moteurs d'avions).

- Le CD-ROM des normes AFNOR contenant les bases de données normatives et thématiques de l'AFNOR. Ces CD fonctionnent à l'aide d'une application standard qui permet de paramétrer divers produits,utilisant les formats de données existants à l'AFNOR (SGML et images).

- La collection juridique électronique NAVIS, des Editions Francis Lefebvre: Navis Fiscal, Navis Social, Navis Immobilier, Navis Sociétés. Ces titres de CD sont développés sur la base de la plateforme GTI livre électronique de Jouve SI.

2.1.3 - La publication sur INTERNET

Divers types de publications sont présents sur le réseau Internet: articles, revues, journaux, prépublications, rapports, proceedings, thèses et même ouvrages. Ces publications sont hébergées sur le réseau sous forme de pages dites Web. La conception de ces pages se fait grâce au langage HTML. Ce dernier n'est qu'une DTD spécifique du SGML. L'investissement opéré par de grands constructeurs informatiques (MICROSOFT, IBM, SUN, Netscape Communications...) sur ce format HTML, l'importance de l'offre en matière d'éditeurs (Front Page, Hotmetal, WebMaster, ...) et fureteurs HTML (Nestcape, Internet Explorer) et leur gratuité ont amené à une situation où le HTML a tendance à gagner plus de terrain que le SGML. On assiste actuellement à de larges débats et controverses, au sein de la communauté des experts de l'édition électronique, sur le devenir de ce format SGML et de la nécessité de sa banalisation et de sa généralisation, notamment auprès des auteurs et des sociétés savantes. Ce débatse trouve amplifié davantage avec la venue, à partir de Novembre 1996, du XML (Extensible Markup Language) et ce à l'occasion la Conférence SGML 96 de Boston. Cet XML soutenu par le W3 Consortium se donne comme ambition de concilier entre la convivialité et la simplicité du HTML d'une part et la flexibilité du SGML d'autre part.

Cela dit, « le marché des logiciels SGML se développe et des outils de plus en plus puissants sont disponibles. Avec le développement du World Wide Web, des solutions de coexistence de HTML et de SGML sont développées. Une première solution consiste à transformer des instances SGML en fichiers HTML. Une seconde

solution consiste à stocker les instances SGML sur le serveur et en permettre la visualisation avec un SGML viewer. A titre d'exemple, les documents constituant le site WWW du Réseau canadien des technologies (RCT) sont rédigés par un balisage et marquage SGML. Les marques, qui décrivent et structurent des zones du texte, étant en SGML, les fureteurs les plus répandus du WWW les ignorent tout simplement puisqu'ils ne les reconnaissent pas. Par contre, le logiciel de recherche OpenText les reconnaît et les prend en considération lorsqu'il répond aux requêtes structurées de recherche effectuées par l'usager du site WWW du RCT (On peut effectuer une recherche sur les sites du RCT à partir de la page http:// data. ctn. nrc. ca/recherch. html./[7]

2.2 - Introduction de SGML dans les bibliothèues:

Les possibilités de manipulation de textes codés SGML ont permis l'extension de l'utilisation de ce format à des fins bibliographiques. Les fonctions principales de la chaîne documentaire sont, pratiquement, toutes impliquées par ce nouveau mode d'organisation des données. De telles possibilités ont même amené certains éditeurs

La BIEF (Banque Information sur les Etats Francophones) entretient et publie sur disque compact la base de données ENSEMBLE: grand catalogue bibliographique collectif signalétique multilatéral. La base de données compte 39168 notices dont environ 10 000 notices de thèses acquises de University Microfilms International et tirées de la base de données Dissertation Abstracts International. Les données qui constituent cette base de donnéess ont recueillies par les Centres serveurs BIEF sous le logiciel CDS-ISIS et envoyées au bureau central de la BIEF qui s'occupe de les intégrer dans un fichier central. Une fois assemblées, les données sont exportables en format CCF/BIEF, un format qui tient à la fois du CCF et du MARC. Quelues années après le lancement de ce service, la BIEF décide d'étudier la possibilité de balancer les notices d'ENSEMBLE vers le format SGML et ce afin de s'adapter avec deux nouvelles données:

* Le disue compact avec son énorme capacité permet de livrer la base de données en entier aux usagers. Se pose alors le problème du choix d'un logiciel de consultation, de présentation et d'exploitation des données;

* L'Internet qui permet une consultation en direct par des usagers du monde entier.

Le fichier SGML qui contient l'échantillon de ENSEMBLE a été installé directement sur le logiciel sans qu'aucune modification ne soit nécessaire.

Selon J.R. Dupuy et C. Schnéegans (de Tecsult Eduplus), cette expérience illustre bien l'usage assez général qui peut être fait à partir de document en format SGML par rapport au format MARC [25].

c) Le Catalogue Collectif de la Belgique (CCB)

Un projet d'échanges bibliographiques entre 40 bibliothèques universitaires et spécialisées. Il comprent 4 000 000 de notices sur deux CD-ROM. Ce projet a permis à toutes les universités belges d'échanger leurs données en format SGML(1).

d) Le Catalogue Antilope, également en format SGML, est une banque de données de périodiques totalisant 75 000 titres avec environ 230 000 localisations dans environ 80 bibliothèques(2).

2.2.2 - Le catalogage et la description des documents électroniques: les notions de TEI, Metadata et Dublin Core

La métainformation est une valeur ajoutée à l'information, elle en permet la compilation et le repérage. Les bibliothèques produisent depuis toujours ce type d'information. L'Internet, pour sa part, commence à peine et la tâche n'est pas mince: on évalue à plus de 50 000 000 le nombre de pages sur l'Internet qui constituent autant de documents. Indexer cette masse d'information pour en faciliter l'accès aux usagers nécessite avant tout un travail de préparation de l'information qui pour le moment est encore fait après publication. L'ajout de métainformation estt un travail inévitable. Et c'est dans ce sens que les mondes de l'Internet et des bibliothèques auront beaucoup à apprendre l'un de l'autre.

La norme SGML est facilement utilisable pour la description des données bibliographiques parce que ces données contiennent du texte et ont une structure logique (règles de catalogage) qui se prête à la description par une DTD.

---

(1) - La DTD du CCB est accessible à l'adresse: ftp: //lib.ua.ac.be/pub/ccb/ccb. dtd
(2) - La DTD du catalogue Antilope est accessible à l'adresse ftp suivante: ftp: //lib.ua.ac.be/pub/antilope/atp/atp.dtd

Cela représente une des originalités des formats de structuration logique des documents, SGML en particulier, à savoir le fait que la notion de distinction entre un document primaire et sa forme secondaire (signalement bibliographique) est rompue. Ceci permet aux éditeurs d'étendre le principe du CIP (Cataloguing in publications) aux documents électroniques.

Claude Scheneegans et Robert Dupuy considèrent [25] que les bibliothèques ayant opté pour le format MARC avaient un « avenir flou », ce qui n'est pas le cas pour les systèmes de bibliothèques futures qui auront à gérer des formats SGML. Pouvoir manipuler dans un même format un document et sa méta-information (notice bibliographique en SGML) constitue certainement un atout très appréciable pour les bibliothèques devant fournir des documents électroniques.

Un projet d'envergure international dénommé TEI (Text Encoding Initiatives) a été institué spécialement pour définir les règles de construction de l'en-tête bibliographiques, attaché à chaque document électronique. Une première version de directives BFD (Bibliographic File Description) a été publiée en 1994 (http: //etext. virginia. edu/TEI.html). Cette BFD est composée de cinq sous zones:

- Titre et mention de responsabilité,
- Edition,
- Collation (taille du document en bits...),
- Adresse,
- Zone des notes.

L'en-tête TEI comporte également des renseignements sur l'encodage du texte, les caractéristiques non bibliographiques et les mises à jours effectuées sur le texte en question.

Cet en-tête constitue, pour le document électronique, la page de titre d'un livre imprimé. Son intérêt concerne les auteurs des documents numériques (devant fournir des informations sur leurs textes), les bibliothécaires (devant assurer le catalogage des documents numériques) et les chercheurs (devant consulter et utiliser ces textes).

Malgré ces avantages, les directives relatives aux en-têtes TEI n'ont pas encore le statut de norme. Selon Panchysyne et Bouthillier [23] « Le format MARC assure toujours l'échange d'information entre les catalogues en ligne (au moyen de la norme Z39.50). Tant que les catalogues n'auront pas d'outils de navigation pour le Web et ne

seront pas compatibles avec le langage SGML, ce qui élargirait leurs potentiels, MARC restera la norme de choix ». Cette situation a poussé la Library of Congress à entreprendre depuis 1996, le projet SGML-MARC ayant pour but de développer une version SGML du format MARC. Le test d'une première version MARC-DTD, mené en février 1997, s'avère positif.

De son côté, la bibliothèque de l'Université de Californie, Berkeley, a mis au point une DTD SGML du format USMARC. Cette DTD est diffusée gratuitement sur Internet depuis Avril 1994 (ftp: //library. berkeley. edu/pub/sgml/marcdtd/.

La réussite d'un tel projet démontre selon Sally McCallum (de la Bibliothèque du Congrès)[13], la souplesse et la puissance de ce format SGML. Il est important de retenir le fait que ce format puisse englober les formats de type MARC tout comme le HTML.

Par ailleurs, la recherche de normalisation d'une métadonnée conçue comme un en-tête représente le coeur d'un autre projet d'OCLC, le Dublin Core. A l'issu de la première conférence sur les métadonnées (Metadata Workshop) de Mars 1995, une liste de 13 éléments devant permettre aux auteurs de décrire eux-même leurs documents (et ceux sans prérequis en matière de catalogage) a été retenue. Cette liste, baptisée Dublin Core, a subit certaines modifications et ajouts à travers plusieurs rencontres. Depuis juin 1997, cette liste comprend 15 éléments.

2.2.3 - L'indexation et la recherche documentaire:

L'usage du SGML dans ce domaine sert à faciliter les taches de production des bulletins signalétiques et l'indexation grâce au repérage sans équivoque de mots clés, dans un document. Dès la fin de la saisie d'un document SGML, on a la possibilité de générer, de façon automatiue, des index et une liste de références bibliographiques. Sur ce point on cite souvent le cas de l'Association for Computing Machine (ACM). Cet organisme qui publie plus de 30 000 pages/an procède depuis quelques années à l'archivage électronique de ses articles grâce à la technique SGML. Ce qui facilite l'édition de ses index et bibliographies.

2.2.4 - La diffusion électronique des documents:

SGML est aussi à la base de projets de bibliothèques électroniques pour le stockage et la diffusion de documents. A ce titre, une série de projets internationaux doivent être mentionnés:

a) - ELSA (Electronic Library SGML Application):

Le projet ELSA 1994-1995 est un projet de recherche et de développement subventionné par le CEE. Direction XIII, dans le cadre du programme Bibliothèques. Il rassemble trois partenaires européens: la Bibliothèque De Montfort University (Leicester UK), l'éditeur Elsevier Science BV (pays bas) et la société Jouve système d'information (France). C'est en quelque sorte une extension européenne du programme TULIP.

Les documents codés en SGML proviendront de la chaîne en cours de codification chez l'éditeur, et suivront la DTD d'Elsevier pour l'en-tête et pour le corps de l'article.

Le projet est accessible sur Internet à l'url suivant: http://www.elsa. dum. ac.uk

b)) CAPCAS:

C'est un service commercial de fourniture d'information bibliographique sur les articles de journaux scientifiques réalisés par l'éditeur Elsevier Science BV. Ce service qui entre en phase de commercialisation concerne 400 titres de l'éditeur. Par ailleurs, celui-ci est en train de modifier la chaîne de production de 1100 de ses titres pour étendre le service, par l'introduction du codage en SGML, au moins pour l'en-tête de l'article auteurs et leurs affiliations, titres, mots clés, résumé, bibliographie).

Les données sont mises à disposition deux mois et demi avant l'apparition de la version papier, avec un résumé non finalisé, et font l'objet d'une licence d'utilisation électronique. La diffusion est effectuée en SGML, selon la version Elsevier de la DTD d'en-tête des articles de journaux scientifiques.

Le public visé est celui des organismes d'information secondaire: Bibliothèques, bases de données dans le monde entier. C'est une sorte de CIP (Cataloguing in Publication ou catalogage à la publication). L'expérimentation commerciale en cours à Tilburg University, Pays bas, doit permettre de connaître le comportement des utilisateurs, et de déterminer la tarification et les licences d'utilisation.

Le projet est accessible sur Internet via l'url: http://www-east.elsevier.com/ees/ capcas.htm

c) - Right Pages:

Right Pages est un projet mis en oeuvre par AT&T et dans un premier temps de l'éditeur Springer Verlag, d'autres éditeurs étant

contactés pour l'extension du service. L'objectif est de construire un réseau de serveurs de journaux électroniques relayant un serveur central. Le système s'appuie sur la production des journaux et sur leurs stockage dans le format SGML. Les abonnés sont avertis des nouveautés par messagerie électronique et reçoivent le signalement des articles avec le résumé.

Le projet est accessible sur Internet via l'url: http://www.ckm.ucsf.edu/ Projets/ RedSage

d) - CORE:

Le projet CORE (Chemistry On line Retrieval Experiment) réunit 5 partenaires des USA couvrant eux aussi la totalité de la chaîne: La bibliothèque Mann de l'Université Cornell, Bell Communication Reserach, the American Chemical Society ACS, Chemical Abstracts Services et le réseau OCLC (Online computer Library Center). Le projet concerne 20 périodiques de chimie depuis 1980 (1974 pour certains), soit l'équivalent d'environ 500 000 pages.

3 - Les outils et les grands constructeurs:

3.1 - De quelques outils:

Les outils permettant de manipuler SGML existent sur le marché. On peut choisir entre des outils très professionnels ou des outils plus orientés vers l'utilisateur final offrant une ergonomie de type WSIWYG (whaht you see is what you get). Une série de constructeurs s'imposent dans le marché des outils SGML.

Un package SGML complet comprend un éditeur et un visualiseur. En effet, pour produire un document balisé SGML, un éditeur (ou imprimeur ou autre...) a besoin:

a) D'un éditeur SGML qui lui permettra de saisir le texte, d'intégrer les balises de structure nécessaires, de créer des liens logiques entre ces balises et d'en valider la conformité par rapport au modèle construit (à l'aide d'un parseur intégré). Il doit éventuellement permettre la retroconversion de documents antérieurs en format SGML.

Cette opération s'exécute selon trois procédés possibles:

- Nouvelle saisie des textes;
- Lecture optique et technique d'OCR;

- Capture et récupération par transcodage et conversion, en utilisant des utilitaires spécifiques (Ex: Dynatag, Omnimark, Balise...).

Un éditeur doit être plus ou moins convivial et facile à apprendre. Il doit permettre à l'opérateur de continuer son travail sans que le document soit validé. Le choix de ces éditeurs doit tenir compte des compétences des opérateurs, de l'assistance technique possible et de la productivité souhaitée.

Il existe, sur le marché, plusieurs outils pour l'entrée de documents en SGML. Les plus répandus de ces « Authoring Tools » (outils pour l'auteur) sont:

- ArborText d'Adept *Editor
- SoftQuad Author/Editor
- InContext
- FrameMaker+SGML d'Adobe
- Near & Far Author de Microstar
- Interleaf 6 SGML
- GTI Publisher de Jouve SI.

Il existe également des adjuvants pour le traitement de documents SGML par Word Perfet ou Microsoft Word comme SGML TagPerfect (voir http: //www.jsp. fi/delta/deltatgp.htm) ou SGML TagWizard (voir: **Erreur! Source du renvoi introuvable).**

Par ailleurs, l'offre logicielle en matière d'entrée SGML s'étend également aux convertisseurs SGML. Citons le cas de:

- OmniMark.
- Avalanche FastTag
- perl

Un site Web est consacré à ce sujet d'entrée des documents en SGML. Il est accessible au: <http://www.sil.org/sgml/commSW.html>

b) D'un visualiseur rendant l'information lisible et consultable par un lecteur donné. Trois modes de lecture sont possibles: l'écran, le CD-ROM et Internet. On distinguera entre des visualiseurs simples pour le premier cas et visualiseurs de plus en plus sophistiqués pour les deux derniers.

Parmi les visualisseurs (ou viewers) les plus connus nous citons:

- Dynatext/Dynaweb de Ebt/Inso
- Folioviews d'Opentext
- Insight into information d'Enigma

Netscape Navigator, Internet Explorer et les autres Web Browsers peuvent en réalité être considérés comme des utilitaires SGML, dans la mesure ou le HTML n'est qu'une version simplifiée (et dédiée au documents web) de SGML.

Il y a aussi des utilitaires SGML destinés spécialement pour les environnements de types bases de données. Citons à ce titre, le logiciel Edibase: logiciel de consultation de bases de données en SGML (voir **Erreur! Source de renvoi introuvable**. Mais aussi:

- Xyvision PDM
- XSoft Astoria
- Documentum
- WorkGroup Technology CMS

3.2 - De quelques constructeurs et de leurs produits:

3.2.1 - Les produits de JOUVE Système d'Information:

La norme SGML est omniprésente dans l'ensemble des prestations Jouve S.I:

* Ecriture et optimisation de DTD.

* Saisie SGML de document (saisie clavier).

* Saisie électronique (OCR/ICR) de documents avec structuration SGML.

* Rétroconverson: structuration SGML de documents électroniques existants.

Jouve réalise des applications pour des éditeurs qui consistent à partir d'ouvrages papier qui doivent être réimprimés dans un format différent. Il faut donc convertir ces ouvrages sous forme électronique pour pouvoir les recomposer.

Jouve développe et commercialise GTI-Publisher. Un produit permettant de créer des livres électroniques multimédias simplement à partir des données SGML. Il est composé de deux parties:

a - GTI PubMaker qui permet à l'éditeur de préparer les données SGML qui seront stockées sur le CD-ROM.

b - GTI PubUser permet au lecteur de visualiser et de naviguer dans le livre électronique.

Dans le cadre de l'ingénierie des systèmes de gestion de documnts électroniques, Jouve intègre également les outils et bases de données SGML du marché.

3.2.2 - Les outils d'ADOBE:

a - Le logiciel Acrobat:

Permet de transcrire en PDF des documents créés avec d'autres éditeurs tels que des éditeurs SGML ou de traitement de texte. Il traduit en PDF des balises SGML en fonction d'une DTD donnée, ou des codages de traitement de texte en fonction d'une feuille de style donnée. Adobe travaille également à la conversion inverse permettant de restituer le document SGML ou le format de traitement de texte d'origine.

b - Adobe FrameMaker+SGML 5.1.1

Il est constitué de deux segments: le premier Framemaker permettant la composition et la mise en page, le second (+SGML) permettant le balisage logique du texte.

Version avancée de FrameBuilder, FrameMaker+SGML 5.1.1 se présente comme une solution intégrée SGML compatible à toutes les plates-formes, avec des améliorations, notamment de la puissance de l'option SGML import/export.

Ce logiciel permet de créer, d'éditer et de publier des documents SGML en utilisant un environnement convivial.

Par ailleurs, les logiciels bibliographiques disponibles actuellement sur le marché(3) ont tendance à tenir compte du succès du format SGML. Ainsi et selon J-Bourgault(4), il y'a « possibilité d'exporter en HTML (un cousin du SGML) les références bibliographiques gérées par ces logiciels, principalement ProCite. Toutefois attendus, la souplesse potentielle des fichiers de configuration qui génèrent la mise en page, il est possible de définir quelques choses qui se rapprochait fort des véritables codes SGML ».

---

(3) - Voir Réf. [2] (J-D Bourgault).

(4) - Email du 16/01/97 from JD.-Bourgault to **Erreur! Source du renvoi introuvable.**

4 - Avantages et inconvénients du SGML: avis d'experts

Selon C. Lupovici (JOUVE SI), il est important de préciser que le succès de SGML est dû « aux outils disponibles commercialement pour créer, manipuler et traiter des documents codés en SGML, ainsi que de son utilisation dans de grands programmes prestigieux (CALS, Office Europ, Brevets, EDF, ...). A l'opposé, ODA: Office Document Architecture-ISO 8613 (qui est une norme développée en Europe qui permet l'interactivité et la manipulation des documents en combinant la bureautique et l'industrie des télécommunications) souffre de la non disponibilité commerciale d'outils de création et de manipulation de documents ODA [19].

A. el Zaim et S. Tellier estiment que, depuis peu, le milieu des bibliothèques adopte le format SGML pour l'exploitation des données bibliographiques à la fois pour les catalogues collectifs, la publication électronique en ligne, les documents de commande électronique et la publication pour le WWW sur l'Internet. Cette adoption de SGML n'étonne pas considérant la liste impressionnante de ses avantages:

* Il intègre les formats MARC;
* Il intègre les formats non-MARC;
* Il permet la recherche en texte intégral;
* Il permet la navigation hypertextuelle;
* Il est bien adapté pour les données bibliographiques qui comprennent du texte et sont structurées;
* Il est indépendant de la plate-forme et du logiciel;
* Il permet de réutiliser les données.

Mais en se demandant: pourquoi cette adoption (de SGML de la part des bibliothèques) vient-elle si tardivement?, ils rétorquent:

* SGML est complexe et coûteux à développer.
* Un bon nombre de bibliothèques ont tardé à s'informatiser.
* Le milieu des éditeurs et celui des bibliothèques continue à fonctionner sur un mode traditionnel. Mais, l'avènement de la bibliothèque numérique est sur le point d'entraîner l'acquisition de documents électroniques codés en SGML directement des éditeurs.
* Le marché des logiciels SGML se développe et des outils de plus en plus puissants sont disponibles.
* Avec le développement du World Wide Web, des solutions de coexistence de HTML et de SGML sont développées. Une première solution consiste à transformer des instances SGML en fichiers

HTML. Une seconde solution consiste à stocker les instances SGML sur le serveur et en permettre la visualisation avec un SGML viewer [7].

Par ailleurs, la conférence internationale sur l'avenir des formats de communication (du 7-11 octobre 1996 à la Bibliothèque nationale du Canada-Ottawa) fait ressortir un certain nombre de contraintes liées à SGML:

- Les logiciels qui gèrent SGML sont chers et peu répandus;
- Les documents qui font l'objet d'une DTD sont encore rares;
- La complexité de fabrication d'une DTD;
- Les problèmes de préservation du droit d'auteur et du droit de reproduction des documents électroniques sont encore non résolus.

Tout en admettant la cherté des logiciels SGML, la deuxième critique est à prendre avec réserve puisque les DTD des publications scientifiques et de certains types de documents techniques sont déjà disponibles (voir paragraphe 1).

Pour certains experts HTML et SGML ne constituent pas de parfaites solutions pour l'échange d'informations documentaires structurées sur l'Internet ou dans les intranets. Et que l'alternative doit être cherchée du côté du XML (eXtensible Markup Language). Résultat des travaux du consortium W3C, le projet de standard XML a été annoncé pour la première fois en novembre 1996, lors de la conférence SGML96, à Boston. Ce nouveau format se veut une solution qui englobe à la fois la fléxibilité de SGML, et la simplicité de HTML. Et comme meilleur compromis entre la norme générique SGML, considérée par la communauté Internet comme trop lourde et complexe à gérer, et les limites de codage HTML. A ce titre, ce format XML est appelé à faciliter la conception, la consultation et l'échanges de documents aussi bien SGML qu'HTML, sur le Web.

Patrick Fenouil (PDG des éditions LAVOISIER-France) répond à la question quelle est la politique de LAVOISIER vis-à-vis de SGML? de la façon suivante(5):

« Nous n'utilisons pas SGML dans notre société mais seulement, comme des millions d'autres, la version dégradée de SGML qu'est HTML pour notre Web. Il y a deux raisons à ce fait: SGML est très coûteux à mettre en oeuvre surtout pour des tirages aussi réduits que

---

(5) - Email du 7 mai 1997, from Patric Fenouil **Erreur! Source du renvoi introuvable.** To: **Erreur! Source du renvoi introuvable.**

ceux de l'édition scientifique; de plus, les données scientifiques évoluent très vite rendant leur conservation inutile (il est souvent plus rapide pour un auteur de tout réécrire que de mettre à jour pour qu'ensuite quelqu'un reporte les modifications et qu'il soit obligé de contrôler, etc.). Mais j'insiste bien sur le fait qu'il s'agit d'ouvrages scientifiques comme ceux que nous publions, car il est bien évident que dans d'autres domaines plus proches du multimédia et destinés à un plus large public, SGML est certainement d'une grande utilité. Il est en revanche pas interdit de penser que l'utilisation de HTML se répandra dans l'édition (voir le CD-ROM Histoire au jour le jour du journal Le Monde conçu pour être consulté avec un logiciel comme Microsoft Explorer ou Netscape Navigator) comme elle se répand actuellement dans les Internet ».

Les éditions juridiques et techniques Lamy ont opté pour le SGML pour leurs informations de type documentaires, synthétiques, on nécessitant une actualisation permanente. Dans leur banque de données unique, l'information brute une fois rassemblée est structurée en SGML. Cette norme assez complexe, exige des investissements importants, notamment dans l'analyse structurée des documents, mais elle ouvre des horizons intéressants pour l'interrogation, la production et la diffusion de documents à la demande et sur mesure, sous forme électronique et sous forme papier [15].

5 - Tendances et perspectives:

Selon Lupovici, les éditeurs ont tendance à travailler de plus en plus à la création de l'information à l'aide de chaînes d'édition électroniques utilisant des normes telles que SGML qui les rendent indépendants des constructions informatiques et des logiciels... L'année 1992 est celle de la prise de conscience de l'importance pour les éditeurs des produits électroniques et du multimédia. La foire de Francfort 1992 a montré que le livre électronique commençait à s'étendre au secteur de littéraire et les organisateurs ont décidé de faire une place spéciale à l'édition électronique en 1993.

SGML a tendance à induire des changements des pratiques professionnelles des auteurs et des éditeurs. En effet, les éditeurs commencent à accepter l'idée de diffuser des produits électroniques destinés à d'autres intermédiaires de la chaîne de distribution, voire à l'utilisateur final lui-même. Ils sont conscients que cette diffusion, pour être adoptée par les destinataires, doit s'appuyer sur une offre comportant une masse critique de documents ainsi que sur une normalisation du codage de contenus et de structure des documents.

Il serait, certainement préférable que les bibliothécaires investissent dans la connaissance et la manipulation de SGML sur le document complet lui-même et sur les services qui peuvent lui être associés, plutôt que de peaufiner le catalogage en SGML [17].

En répondant à la questionn « qu'est-ce que c'est que les SGML/XML?? » posée sur la liste de discussion ebsi-I, de l'EBSI (Canada), le Professeur Jean-Claude Guédon écrit(6): « SGML permet de "baliser" un document, c'est-à-dire d'en donner une description précise à partir d'éléments pré-identifiés selon des règles que donne justement la norme SGML. Pour ce faire, on construit ce que l'on appelle une DTD (Description de type de document) en fonction des documents que l'on veut traiter et des besoins que l'on a). Il existe énormément de DTD, mais la plus célèbre est sans contexte HTML. Mais HTML n'offre que relativement peu de possibilités et tante de tout faire pour tout le monde, ce qui laisse tout le monde insatisfait. Le consortium chargé de veiller sur l'évolution harmonieuse et ouverte des normes de la Toile a donc décidé d'introduire une version simplifiée de SGML, XML qui permet de définir différentes DTD pour différents besoins. On peut donc s'attendre à la création de « canaux » différents sur Internet en fonction des besoins. Parallèlement, on peut s'attendre à une révision en profondeur des fureteurs et une convergence très nette entre fureteurs courants fondées en HTML et fureteurs plus spécialisés ouvrant la porte à diverses DTD de SGML (Panorama de Softquad, par exemple) ».

Selon François Chahuneau (directeur général de la filiale informatique éditoriale du groupe Berger-Levrault, et spécialiste de SGML depuis 1988) « Le groupe de travail sur XML, dédié au W3C (composé notamment d'experts SGML) a prévu trois étapes. La première concerne la publication de la version XML 1.0 adaptée aux applications Web. Celle-ci sera complétée lors d'une deuxième phase par un ensemble de spécifications des liens hypertextes d'applications SGML sur Internet. Ces méthodes d'association de liens plus complexes devraient être dévoilées lors de la sixième conférence du World Wide Web en avril 1997. Enfin, la troisième étape consistera à intégrer à XML les feuilles de style adaptées aux navigateurs du marché.

---

(6) - Email du 15 fev 1998, from: Guedon Jean-Claude **Erreur! Source du renvoi introuvable.** Subject: Re: Re: Projets SGML/XML (fwd) to: ebsi-I@CC. U Montréal. CA.

Le futeur standard XML apparaît comme le meilleur compromis entre la norme générique SGML, vieille de dix ans, et considérée par la communauté Internet comme trop lourde et complexe à gérer, et les limites de codage HTML.

Le succès de cette nouvelle spécification dépendra bien sûr de son adoption par les fournisseurs d'outils Web. Microsoft l'a compris et travaille aujourd'hui à l'intégration d'XML dans sa version Internet Explorer 4.0. De son côté, Nestcape n'a pour l'heure communiqué aucune intention de supporter le futur standard [13].

6 - Conclusion:

En guise de conclusion, il ressort l'importance accordée aussi bien par les éditeurs que par les professionnels de l'information à cette question de structuration et de représentation logique des documents et des métadonnées.

De par leurs besoins et des contraintes liées à leur métier, les éditeurs ont été les premiers à investir ce créneau, ce qu'il leur a même permis d'élargir leur domaine d'intervention à certains maillons de la chaîne documentaire. Les bibliothécaires et les professionnels de la documentation, notamment aux USA et au Canada, ont aussitôt pris conscience de l'importance du format SGML et ce sous l'effet de la multiplication des documents électroniques au sein de leurs collections et des problèmes techniques posés par leur catalogage et échanges. La pratique de SGML par ces deux principaux acteurs de la chaîne de l'information électronique a conduit vers une situation de mise en commun du savoir-faire de tout un chacun pour atteindre davantage un objectif commun, celui de satisfaire le lecteur et l'utilisateur de l'information. Les projets de diffusion électronique des revues TULIP, ELSA, CORE, CAPCAS, Right Pages, etc. illustrent bien cette nouvelle tendance. Nous remarquons cependant, que ces projets sont généralement initiés et pilotés par les grands éditeurs internationaux.

En matière d'utilisation de SGML dans le domaine de catalogage et de l'échange bibliographique, le problème ne se pose plus en terme de disparition de la fonction catalogage et du Marc ou non??, mais plutôt jusqu'où peut on partir dans le domaine des passerelles SGML/MARC?. Dans ce domaine, il est à noter le rôle fédérateur de la LC (Library of COngres) et de l'OCLC dans le domaine de la normalisation et de l'échange des informations bibliographiques codées SGML, HTML, MARC, etc. A retenir aussi le fait que, les

initiatives d'identification de métainformation sur l'Internet (MARC, TEI, Dublin Core, EAD, etc.) se rapprochent de plus en plus de SGML [7].

Servir comme langage d'échange d'informations bibliographiques constitue ainsi, une des exploitations des plus pertinentes du SGML. Des travaux sont en cours pour que cette fonction d'échange soit valable aussi bien entre les différents postes de la chaîne de production et de distribution des documents qu'entre les intervenants d'un même maillon de cette chaîne.

La prise en charge du format SGML par de grandes entreprises informatiques, à travers le développement d'utilitaires (éditeurs, viewers et convertisseurs SGML) d'une part et l'adoption de ce format par des organismes renommés d'autre part, constituent deux principaux arguments justifiant l'investissement sur un tel standard. La popularité du HTML (langage de programmation Internet) qui n'est qu'un cas particulier du SGML amènent paradoxalement les auteurs, éditeurs et documentalistes à s'intérroger (et parfois remettre en cause) sur l'utilité du format source SGML. Ainsi toute une polémique se trouve instaurée autour de la question SGML/HTML: quel choix faire?

Etant donné les arguments presque équilibrés de chacun, ce débat n'avait pas l'air de statuer ni de trancher sur tel ou tel format, jusqu'au mois de novembre 1996 lorsque le W3C annonce, à l'occasion de la conférence SGML96 de Boston, le nouveau format XML. Mais le SGML constitue, quoique l'on dise, un format générique source et de base pour le HTML et les autres formats, y compris le XML.

Techniquement le choix de tel ou tel format n'exclut pas le recours à une solution mixte. Mais quoi qu'il en soit, et selon les dernières rencontres spécialisées dans ce domaine(7), la réflexion sur une méthodologie de choix s'impose. En d'autres termes, le choix ne doit pas se faire inconsidérément mais sur la base d'une grille d'analyse et de comparaison des formats, en fonction d'un certain nombre de paramètres et de critères propres à une situation éditoriale et documentaire donnée.

---

(7) - Référence à la journée d'étude organisée par le Groupement français de l'industrie de l'information GFII, le 22 novembre 1996 au ministère de l'Education nationale, de l'enseignement supérieur et de la recherche, sur l'édition électronique. Voir compte rendu d'Annie Le Saux [15].

Dans ce sens, il y a lieu d'accorder une grande importance à quatre paramètres: le coût, le volume d'informations à traiter, l'utilisation que l'on veut en faire et le public cible.

Mars 1998

Bibliographie:

[1] - Barden Phil., La fourniture de documents en l'an 2000: naissance d'une nouvelle industrie. in: BBF: t. 41, n°1, 1996, pp. 42-46.

[2] - Bourgault Jean-Claude, Usage des ressources de l'Internet comme outils de support pour les logiciels bibliographiques. in Argus, vol. 24, n° 2, mai-août 1995, pp. 36-40.

[3] - Clarisse Burger, XML: un pas en avant pour l'édition sur le Web! in: Le Monde Informatique, n°711, 28 février 1997, htttp: //www.Im.fr/Imi/Imihome.html

[4] - Comberousse Martine, Les nouvelles technologies au service de la littérature grise. in BBF, t. 40, n° 2, 1995, pp. 51-53.

[5] - Coquard Guy, Conversion de documents papier vers SGML: la productivité par la technologie. in IDT 91, pp. 262-263.

[6] - Desrichard Yves, L'avenir des formats de communication: le point sur... in BBF, t. 42, n°1, 1997, pp. 76-78.

[7] - El Zaim Adel et Tellier Sylvie, Formats Internet et formats bibliographiques: mariage d'amour ou mariage de raison? in: L'avenir des formats de communication: actes: Bibliothèque nationale du Canada, Ottawa, 7-11 octobre 1996 organisée par la BIEF et la BNC, actes publiés en 1997, pp. 77-85.

[8] - Ghislaine Chartron, Nouvelles problématiques pour l'IST. in: Documentaliste, vol. 32, n° 6, 1995, pp. 289-295.

[9] - Ghislaine Chartron, La presse périodique scientifique sur les réseaux. in: Solaris-dossier n° 3: "médiations scientifiques et réseaux électroniques", (Sous la direction de Ghislaine Chartron) Solaris est une revue électronique éditée sur Internet par le GRESIC via le serveur de l'université de Caen Url: <http://www.info.unicaen.fr/bnum/jelec/Solaris>

[10] - INRIA, Le traitement électronique du document: cours INRIA, 3-7 octobre 1994, Aix en Provence, ouvrages coordonné par Jean-Claude Le Moal et Bernard Hidoine, ADBS Editions, Paris, 1994, 287 p. ISBN 2-901046-76-2

[11] - Jouve SI, Jouve s.i.Infos. Lettre d'information sur les activités de la société Jouve SI.

[12] - Keddari Djalila et Ghida Nassima, Système de génération de structure hypertexte pour les bases de données bibliographiques [s.n] Alger, 1994, mém.ing.: informatique: INI-Alger, 1994, 143 p.

[13] - L'avenir des formats de communication: actes: Bibliothèque nationale du Canada, Ottawa, 7-11 octobre 1996: conférence internationale organisée par la banque internationale d'information sur les états francophones de l'ACCT et la Bibliothèque nationale du Canada-Ottawa: BIEF, 1997, 229 p.

[14] - Le Crsonier Hervé, Les journaux scientifiques électroniques ou la communication de la science à l'heure du réseau mondial. in Solaris, dossier n° 3: "médiations scientifiques et réseaux électroniques" (sous la direction de Ghislaine Chartron).

[15] - Le saux Annie, Le document électronique: le point sur... in BBF, t.42, n°1, 1997, pp. 78-80.

[16] - Les nouvelles technologies dans les bibliothèques/dir. Michèle Rouhet; collab. Claude Aubrie, David Aymonin, Claude Bonelly et al.; préf. Christian Huitema, Paris, Electre-Ed. du Cercle de la librairie, 1996, (Bibliothèques) 386 p. ISBN 2-7654-0623-5.

[17] - Lupovici Catherine, Révolution électronique, in BBF, t.38, n° 5, 1993, pp. 22-31.

[18] Lupovici Catherine, Le périodique électronique, in BBF, t.40, n° 2, 1995, pp. 32-37.-

[19] - Lupovici Catherine, Les évolutions des techniques éditoriales: de l'édition électronique à l'apparition du livre électronique in: Une fenêtre sur ... l'information documentaire et technique (IDT): recueil de textes, Ottawa: Banque Internationale d'Information sur les Etats Francophones, 1995, pp. 168-172.

[20] - Lupovici Catherine, L'information secondaire du document primaire. in: Bulletin de l'ABF, n° 174, 1er trimestre 1997, pp..68-72.

[21] - Lupovici Christian, Les bibliothèques et le défi de l'édition électronique. in: BBF, t. 41, n°1, 1996, pp. 26-31.

[22] - Makhlouf Nawal et Boukettab Fatiha, Conception et réalisation d'un système de gestion d'hyperdocuments: HYMAS (Hyperdocument Management System) [s.n], Alger, 1994: mém. ing.: informatique: INI-Alger, 1994, 114 p.

[23] - Panchyshyn R.S. et Bouthillier F., Cataloguer le cyberespace: le défi des ressources électroniques. in: Documentation et bibliothèques, juillet-septembre 1997, pp. 137-147.

[24] - Role François, La norme SGML pour décrire la structure logique des documents. in: Documentaliste: sciences de l'information, vol. 8, n° 4-5, 1991, pp. 187-192.

[25] - Schéegans Claude et Robert Dupuy, De Marc à SGML: l'expérience de la BIEF. in: L'avenir des formats de communication: actes: Bibliothèque nationale du Canada, Ottawa, 7-11 octobre 1996: conférence internationale organisée par la banque internationale d'information sur les états francophones de l'ACCT et la Bibliothèque nationale du Canada-Ottawa: BIEF, 1997, pp. 145.155.