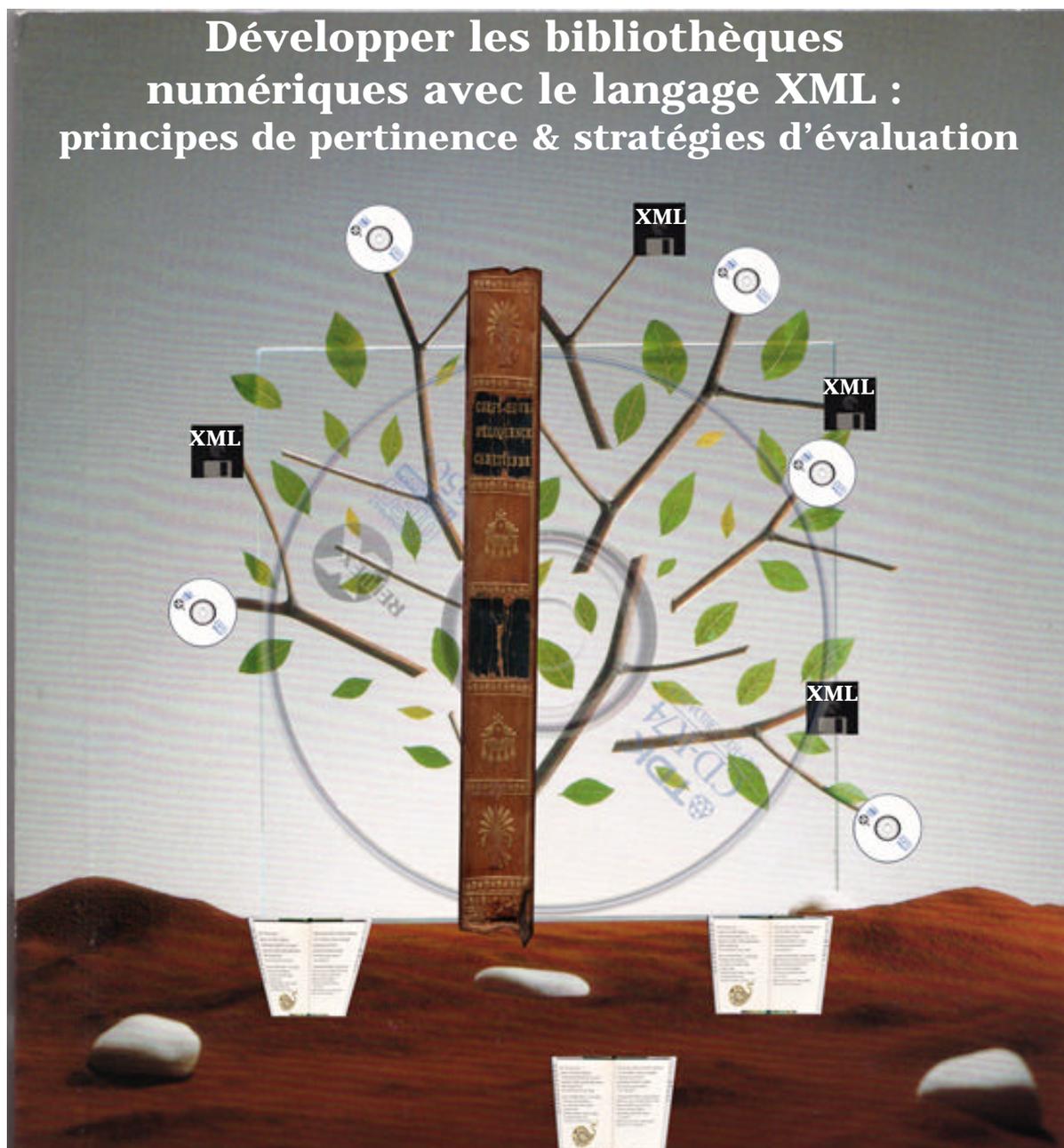


Université de Paris X – Nanterre
200, avenue de la République
92001 Nanterre Cedex

UFR Littérature, Langues et Philosophie (LLPhi)
DEA Sciences de l'Information et de la Communication



Mémoire réalisé par Lise SAVOURAT

Sous la direction de messieurs
François ROLE
Stéphane CHAUDIRON

Session de Septembre 2002

Développer les bibliothèques numériques avec le langage XML :

principes de pertinence et stratégies d'évaluation

Université de Paris X – Nanterre
DEA Sciences de l'Information et de la Communication

Développer les bibliothèques numériques avec le langage XML :

principes de pertinence et stratégies d'évaluation

Lise SAVOURAT

REMERCIEMENTS

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE Merci SYSTEM"merci.dtd">
<Memoire DEA>
  <Entree_merci>
    <Nom>Role</Nom>
    <Prenom>Francois</Prenom>
    <Texte>pour son rôle dans la dynamique de développement
documentaire du langage informatique</Texte>
  </Entree_merci>
  <Entree_merci>
    <Nom>Chaudiron</Nom>
    <Prenom>Stéphane</Prenom>
    <Texte>pour ses informations et ses conseils, tant
sur le plan méthodologique que bibliographique</Texte>
  </Entree_merci>
  <Entree_merci>
    <Nom>Bianchi</Nom>
    <Prenom>Marie-Pierre</Prenom>
    <Texte>pour ses qualités de graphiste autant que pour
celles d'auditrice</Texte>
  </Entree_merci>
  <Entree_merci>
    <Nom>Roger</Nom>
    <Prenom>Ariane</Prenom>
    <Texte>parce que partager les mêmes réalités
rédactionnelles motivent forcément</Texte>
  </Entree_merci>
  <Entree_merci>
    <Nom>Proches, amis et collègues du CNT</Nom>
    <Texte>pour leur participation, consciente ou non,
au partage de connaissances sur le langage XML</Texte>
  </Entree_merci>
</Memoire DEA>
```

**A tous, dans une tournure certes plus académique :
ma sincère considération !**

SOMMAIRE

<u>INTRODUCTION</u>	10
1 <u>REALITE DU LANGAGE XML DANS LES BIBLIOTHEQUES NUMERIQUES : CADRE GLOBAL DE DEVELOPPEMENT ET CAS D'ETUDES SPECIFIQUES</u>	18
1.1 Petite typologie des bibliothèques numériques courantes	18
1.1.1 Les bibliothèques numériques patrimoniales	19
1.1.1.1 Principales caractéristiques de l'existant	19
1.1.1.2 La diffusion du patrimoine culturel	20
1.1.1.2.1 <i>Gallica</i>	20
1.1.1.2.2 <i>Guide de l'Internet culturel</i>	21
1.1.1.2.3 <i>Artistes et Œuvres</i>	22
1.1.1.3 La diffusion du patrimoine scientifique et technique	23
1.1.1.3.1 <i>Article@INIST</i>	23
1.1.1.4 La diffusion du patrimoine juridique	24
1.1.1.4.1 <i>Naiade</i>	24
1.1.1.4.2 <i>Légifrance</i>	25
1.1.2 Les bibliothèques numériques universitaires	26
1.1.2.1 Principales caractéristiques de l'existant	26
1.1.2.2 La diffusion de la recherche universitaire	27
1.1.2.2.1 <i>CodeX</i>	27
1.1.2.2.2 <i>Pelleas</i>	28

1.1.3	Les bibliothèques numériques de presse	28
1.1.3.1	Principales caractéristiques de l'existant	28
1.1.3.2	La diffusion de l'information périodique	30
1.1.3.2.1	<i>Le Point</i>	30
1.1.3.2.2	<i>Les Echos</i>	31
1.2	Contraintes générales de développement en XML	32
1.2.1	S'affranchir d'un contexte ambivalent	32
1.2.1.1	Un nombre limité d'applications	32
1.2.1.2	Une motivation croissante des professionnels	33
1.2.2	Dégager des axes prioritaires	35
1.2.2.1	L'information des utilisateurs réels et potentiels	35
1.2.2.2	La formation des utilisateurs professionnels et finaux	36
2	<u>REALISATION D'UNE GRILLE D'ANALYSE POUR L'EVALUATION DU XML APPLIQUE AUX BIBLIOTHEQUES NUMERIQUES</u>	39
2.1	Données essentielles de l'évaluation	40
2.1.1	Méthodologie	40
2.1.1.1	Qui ?	40
2.1.1.1.1	La source	41
2.1.1.1.2	L'environnement	41

2.1.1.2	Quoi ?	42
2.1.1.2.1	La plate-forme informatique	42
2.1.1.2.2	Le contenu documentaire	42
2.1.1.3	Comment ?	43
2.1.1.3.1	Le fonctionnement de la BN (approche développeurs)	43
2.1.1.3.2	L'exploitation de la BN (approche utilisateurs)	43
2.1.1.4	Pourquoi ?	44
2.1.1.4.1	Les besoins	44
2.1.1.4.2	Les publics	45
2.1.2	Cas d'applications	45
2.1.2.1	Le cas des bibliothèques numériques patrimoniales	46
2.1.2.2	Le cas des bibliothèques numériques universitaires	58
2.1.2.3	Le cas des bibliothèques numériques de presse	62
2.2	Observations générales	66
2.2.1	Le cadre logique : quelles forces dynamiques ?	68
2.2.1.1	L'activité du secteur professionnel	68
2.2.1.2	La complémentarité des métiers	69
2.2.2	Le contenu intellectuel : quels terrains propices ?	70
2.2.2.1	Des documents « pré-structurés »	70
2.2.2.2	Des documents « complexes »	70

3	<u>RESULTATS DE L'EVALUATION ET DEFINITION D'INDICES DE PERTINENCE DU XML</u>	72
3.1	De l'efficacité du langage informatique à la performance du dispositif technique	72
3.1.1	Le potentiel efficace du XML	72
3.1.1.1	Un langage normalisé	73
3.1.1.2	Un langage ouvert	74
3.1.2	Le principe de performance	74
3.2	De la valeur de l'information à la qualité de la documentation	75
3.2.1	La valeur informative	75
3.2.1.1	Disponibilité de l'information	76
3.2.1.2	Fiabilité de l'information	77
3.2.1.3	Adaptabilité de l'information	78
3.2.2	Le principe de qualité	78
3.3	Vers la pertinence du langage XML pour les bibliothèques numériques	79
3.3.1	Le principe de pertinence	80
3.3.2	Les indices de pertinence	83
3.3.2.1	La notion de contextualisation	83
3.3.2.2	La notion de réception	84
3.3.2.3	La notion d'usage	85

<u>CONCLUSION</u>	87
<u>BIBLIOGRAPHIE GENERALE</u>	90
<u>ANNEXE 1 : Revue de presse : « quelle actualité pour les bibliothèques numériques intégrant XML ? »</u>	95
<u>ANNEXE 2 : Rencontres professionnelles : « quelle (autre) actualité pour les bibliothèques numériques intégrant XML ? »</u>	103
<u>ANNEXE 3 : Grille d'analyse : modèle descriptif</u>	109
<u>ANNEXE 4 : Glossaire récapitulatif : terminologie liée à XML</u>	111
<u>ANNEXE 5 : Guide de sites Web : ressources électroniques sur le XML et les bibliothèques numériques</u>	113

INTRODUCTION

Les avancées technologiques depuis plus de vingt ans ont permis de redéfinir, de recomposer le monde des bibliothèques. Sans pour autant remplacer les bibliothèques physiques, espaces matérialisés de connaissances, les bibliothèques numériques, situées dans le cyberspace, sont venues enrichir la sphère informationnelle. Les livres imprimés ont été les piliers de l'éducation et de la transmission des savoirs. Aujourd'hui, les documents électroniques impliquent de nouveaux modes d'apprentissage et de transmission, où les langages informatiques, tels que le XML, ont leur « mot à dire ».

Cadres pratiques

Principales caractéristiques de la bibliothèque numérique

« Bibliothèque numérique », « bibliothèque électronique », « bibliothèque virtuelle », « bibliothèque universelle »... : ces termes sont-ils synonymes ? Ou bien désignent-ils des réalités très différentes pour une même instance, la bibliothèque ? En ce cas, qu'est-ce qui singularise chacun de ces termes ? Comment les définir ? Comment appréhender chaque réalité revêtue par une bibliothèque « sans murs » ? Il semble encore bien délicat aujourd'hui de résoudre durablement ces interrogations, d'une part parce qu'elles renvoient à des notions encore trop récentes, et d'autre part parce que ces mêmes notions sont évolutives, à l'image du support électronique dont elles dépendent. C'est pourquoi nous avons choisi de n'utiliser que le terme de « bibliothèque numérique », notamment parce que le qualificatif « numérique » sous-entend une matérialité, nuance absente des qualificatifs « virtuelle » ou « universelle ».

Que contient une bibliothèque numérique ? Comment s'organise t'elle ? Quelles fonctions remplit-elle ? Quelles contraintes génère t'elle ? Il est indispensable de résoudre chacune de ces interrogations, afin de justifier ultérieurement les exemples que nous avons choisis pour illustrer l'application réelle du XML et des standards associés dans les bibliothèques numériques.

Contenus d'une bibliothèque numérique

Nous pourrions dire simplement qu'une bibliothèque numérique est une collection de documents numériques normalisés et transmissibles. Il convient alors de définir le terme de « document numérique ». En tant que document, ce peut être tout élément organisé d'information, par exemple texte, carte, illustration, dessin, photographie, enregistrement sonore ou audiovisuel... Il n'entretient qu'un rapport indirect avec un document papier, lequel n'est qu'un véhicule transitoire s'il y a numérisation (modèle de départ pour transcription électronique de l'information). Un document numérique peut même parfaitement exister indépendamment d'un document papier, s'il n'y a pas eu numérisation et que le document a été créé directement sur le support informatique. Dans tous les cas, pour être autonome, un document numérique doit être structuré ; en ce sens, il doit contenir les éléments indispensables à son identification, à son administration, à son échange.

Fonctions d'une bibliothèque numérique

En devenant numérique, la bibliothèque doit veiller à conserver ses mêmes fonctions fondamentales. La première chose est qu'elle doit proposer une sélection de ressources documentaires indispensables à la transmission des connaissances. Ces ressources sont peut-être stockées à l'autre bout du monde, mais elles doivent demeurer accessibles depuis n'importe où, via un poste informatique. Cela implique donc une forte normalisation, depuis les unités de base (normes de transcription des caractères pour l'écrit -ex : ISO 10646-, des couleurs pour l'image...), jusqu'aux éléments d'organisation en vastes documents (normes de catalogage, systèmes de numérotation de documents -ex : DOI, URN...- scripts des documents interactifs, et évidemment codage de l'information). Mais au-delà de la gestion et de la conservation d'un fonds documentaire représentatif, une bibliothèque numérique assure aussi la rencontre entre l'écrit et les lecteurs, entre les concepteurs et les lecteurs. Une bibliothèque numérique peut donc être donc tout à la fois un lieu de connaissance, un lieu de recherche, un lieu de lecture et d'écriture, un lieu de discussion, un lieu de partage...

Réalisations et projets de bibliothèques numériques

En près de dix ans d'existence, de nombreuses bibliothèques numériques ont vu le jour. Elles sont de diverses envergures (locale, régionale, nationale, internationale), de différents statuts (public ou privé), mais se côtoient toutes indifféremment sur la Toile. Toutes ces réalisations actuelles constituent une capitalisation exceptionnelle tant de connaissances que

d'expériences, à la fois en méthodes de conduite de projet, en contextes de mise en œuvre, mais aussi en choix technologiques, parmi lesquels celui du langage XML. Nous avons recensé un certain nombre d'applications XML pour les bibliothèques numériques. Elles constituent un référentiel varié, que nous présenterons puis que nous soumettrons à une grille d'analyse, dans le but, si possible, d'en déduire des indices de pertinence valables pour la réalisation d'une bibliothèque numérique à l'aide du XML.

Nous avons donc sélectionné un certain nombre d'applications répondant à ce premier élément de définition d'une bibliothèque numérique : collection normalisée et transmissible de documents numériques réunis du fait de leurs natures (ex : œuvres littéraires, textes de lois, mémoires de recherche...) ou du fait de leurs thématiques (ex : information culturelle, information scientifique et technique, information juridique...). Avec **GALLICA**, la Bibliothèque nationale de France donne à lire des textes numérisés, pour la plupart issus du domaine public. Cas très particulier de bibliothèque numérique, le **GUIDE DE L'INTERNET CULTUREL** permet l'accès à des bases de données, à des sites d'organismes, à des textes fondamentaux, entre autres ressources en ligne pour la diffusion du patrimoine culturel français. Autre réalisation du secteur culturel, **ARTISTES ET ŒUVRES** du Centre national de la Danse repose sur un corpus complexe de documents d'origines et de supports divers, mais autour d'un même thème : celui de la pratique artistique, en particulier de la danse. **ARTICLE@INIST** propose des ressources bibliographiques -monographies et revues- appartenant au champ de l'information scientifique et technique. **NAÏADE** est destiné à fournir un soutien informatif juridique (textes, articles de lois, fiches de procédure...) aux personnes relevant de l'Etablissement national des invalides de la Marine. Plus globalement, **LEGIFRANCE** s'inscrit dans une dynamique entamée voici plus de 15 ans par l'Etat, pour favoriser le développement de l'informatique juridique française dont les premières réalisations remontent tout de même à la fin des années 1960. Dans un tout autre domaine, **CODEX**, développé par l'université de Paris X, est une bibliothèque en ligne de mémoires universitaires numérisés. Elaboré par un consortium d'universités et de centres de recherche, **PELLEAS** est un système de gestion et d'accès numérique à des ressources documentaires faisant partie de l'environnement des chercheurs, des enseignants et des étudiants. Enfin, deux versions électroniques de presse périodique, **LE POINT** et **LES ECHOS**, constituent des bibliothèques numériques aux contours singuliers par rapport à toutes celles citées auparavant.

Apparition du langage XML et des standards associés

Le **XML** (*eXtensible Markup Language*) est un langage constitué de balises, ou tags. Mis au point par le XML Working Group sous la direction du W3C (*World Wide Web Consortium*)¹ dès 1996, il est destiné à la structuration de documents électroniques. En février 1998, les spécifications XML 1.0 sont devenues des recommandations, c'est-à-dire que le XML a été officiellement défini et reconnu comme standard. Par ailleurs, le XML dérive, non pas du HTML (*HyperText Markup Language*), mais du SGML (*Standard Generalized Markup Language*), défini en 1986 par le standard ISO 8879.

Le langage XML « n'enregistre » que la structure logique d'un document, c'est-à-dire l'organisation hiérarchique des données, sous la forme d'une arborescence. Au-delà de la structure logique enregistrée par XML, ce sont d'autres standards qui gèrent notamment :

- la représentation valide d'un document : DTD ou Schéma
- la présentation d'un document : XSL (XSLT) ou CSS
- la gestion des liens : XLink et XPointer

La **DTD** (*Document Type Definition*) est antérieure à XML ; c'est SGML qui a introduit cette notion de modèle générique de données. Avec SGML, une DTD est surtout une grammaire (pour décoder correctement le système de balisage et commencer à appréhender de façon non ambiguë l'information qu'il contient), mais aussi une syntaxe (pour instaurer des contraintes à respecter dans le document). Mais avec XML, la syntaxe est suffisamment claire pour pouvoir être analysée sans risque d'ambiguïté et donc la DTD va au-delà de son statut de grammaire, en s'intéressant à la fois à la cohérence syntaxique -parsabilité- et à la conformité à un modèle générique -validité-. Depuis peu d'autres formalismes, plus puissants et plus flexibles que la DTD, les **Schémas** (ou spécification « XML Schéma »), sont en cours de définition au sein du W3C. Ils commencent seulement à être connus, d'autant qu'ils n'ont réellement été adoptés qu'en mai 2001. Ils tendent à remplacer la DTD dans ses prérogatives, d'autant qu'ils présentent cet autre avantage : ils sont directement écrits dans une syntaxe XML !

Au même titre que la DTD avec SGML, les feuilles de style **CSS** (*Cascading StyleSheet*) préexistent à XML. Elles ont été initialement définies pour HTML, mais ne proposent que des fonctions limitées de génération de contenu (mise en forme) et aucune fonction de transformation de données. En d'autres termes, les CSS constituent des présentations assez

¹ Documents relatifs au XML et aux standards associés et spécifications téléchargeables sur <http://www.w3c.org/XML>

simples qui ne modifient pas la séquence de données. D'où l'apparition d'autres feuilles de styles, plus riches et plus ouvertes pour la présentation externe ou le formatage des données, les **XSL** (*eXtensible Stylesheet Language*). Elles intègrent quant à elles la notion de transformation de données, via **XSLT** (*eXtensible Stylesheet Language Transformation*). Ce standard permettant de transformer un document XML en autre chose -document HTML, WML ou texte-, est passé sous forme de recommandation du W3C en novembre 1999.

Enfin, les spécifications complémentaires **XLink** (*XML Linking Language*) et **XPointer** permettent la représentation de liens internes et externes (hypertextes, url, ancrés...) à un document XML. XLink est devenu une recommandation du W3C en juin 2001 ; elle utilise la syntaxe XML pour créer et décrire des liens dans un document XML. Dans le cas où la cible du lien est elle-même située dans un document XML, XPointer, autre récente recommandation du W3C, permet d'identifier le plus précisément possible puis d'atteindre la cible désignée.

Nous avons constitué en annexe un glossaire récapitulatif des principaux termes utilisés lorsqu'il est question du XML. Il sera aisé de s'y référer chaque fois qu'une abréviation ou un terme technique employé ci-après, méritera quelque bref éclaircissement. Mais il est évident que nous n'avons malheureusement pas pu décliner toute la terminologie adjacente au XML...

Cadres théoriques

Qualité(s) du langage XML et des standards associés : la question de l'évaluation

Dans le contexte actuel de multiplication des possibilités informatiques, il apparaît d'autant plus nécessaire de légitimer l'existence de chaque nouvelle solution offerte. Il existe dans cette perspective quelques façons de procéder, parmi lesquelles la démarche évaluative. L'évaluation est un processus qui vise à examiner aussi systématiquement et objectivement que possible un projet ou un programme en cours ou achevé, en particulier sa conception, son exécution et ses résultats. Cela devra permettre de déterminer l'efficacité du projet, son impact, sa viabilité, mais aussi tout ce qui fait la pertinence de la solution informatique choisie pour le projet. L'évaluation doit ainsi aider à poser la légitimité d'une solution informatique, dans notre cas le langage XML, mais elle doit également permettre de guider les décideurs, dans notre cas les concepteurs des bibliothèques numériques.

Usages et usagers de la bibliothèque numérique : la question de la réception

La sociologie des usages est aussi au cœur de notre propos. De nombreux travaux d'analyse autour de cette problématique ont mis en évidence l'importance des significations d'usages projetées et construites sur le dispositif technique proposé. Autrement dit, nous ne pouvons proposer une évaluation de la technologie XML associée aux bibliothèques numériques, sans prendre en considération la dimension sociale d'un tel processus, en particulier le rôle des usagers des dispositifs techniques nés de cette association. Comment les usagers des bibliothèques numériques « réceptionnent-ils » (appréhendent-ils) l'apport du langage XML ? Ou, de quelle(s) façon(s) la technologie XML s'insère t'elle dans les pratiques des usagers des bibliothèques numériques ?

Problématique

Alors que les bibliothèques numériques se développent maintenant depuis plus de dix ans, elles ont progressivement intégré des technologies nouvelles. Plusieurs technologies coexistent désormais en ligne. Devant la diversification de l'offre, le choix d'une technologie par rapport à une autre, est forcément plus complexe. Il est donc d'autant plus nécessaire de justifier de l'intérêt d'une technologie dans un cadre précis, par rapport à des besoins spécifiques et selon des perspectives variées.

Dans notre cas, la technologie XML est apparue sur la Toile en 1998 ; elle venait alors particulièrement concurrencer le SGML, jugé trop complexe, mais aussi le HTML, considéré comme moins rigoureux. Notre propos est donc aujourd'hui de justifier de l'intérêt du XML dans le cadre précis de la bibliothèque numérique, laquelle nourrit des besoins spécifiques et des ambitions variées. Cette démarche relève de l'évaluation de la pertinence du XML, non seulement par rapport à d'autres technologies, mais aussi et surtout selon des contraintes liées au cadre de développement, aux besoins exprimés et aux ambitions nourries. D'où l'entremêlement de considérations techniques et sociologiques qui existe déjà dans la définition-même de la pertinence. En effet, la notion de pertinence repose sur les aspects pragmatiques et rhétoriques du langage XML qui assurent la communication entre l'homme et la machine, via la bibliothèque numérique. En d'autres termes, l'ensemble des fonctionnalités techniques du XML doit faciliter la relation de l'utilisateur à la bibliothèque numérique. L'articulation entre des fonctionnalités techniques et des perceptions sociologiques s'opère

autour d'une propriété : la pertinence. Un élément se doit d'être pertinent à la fois par rapport au domaine concerné -comment le langage XML peut-il être utile à la bibliothèque numérique ?-, et relativement à sa présentation -comment le langage XML peut-il être utilisable par l'utilisateur de la bibliothèque numérique ?-.

Comment évaluer la pertinence du XML et des standards associés dans le cadre particulier des bibliothèques numériques ? Quels indices de pertinence pouvons-nous déduire des résultats de l'évaluation ? Notre problématique s'oriente donc vers un calcul et une modélisation de la pertinence. A partir du calcul de la pertinence associée au langage XML dans les bibliothèques numériques, il sera en effet possible de modéliser les résultats obtenus sous forme d'indices de pertinence. Pour ce faire, nous procéderons selon trois grands axes d'observation et de réflexion, repris par le plan de rédaction :

- 1 – Quelle est la situation actuelle du XML et des standards associés, en particulier dans le développement des bibliothèques numériques (taux de représentation...) ? Dans quels contextes de conception technique et d'usage social s'inscrit ce développement ? Le cadre global de l'existant devra permettre de situer notre démarche d'évaluation du XML par rapport à des besoins et à des contraintes liés à la médiation de l'information et de la documentation. Il conviendra ensuite de passer d'un état global de l'existant à une analyse spécifique d'applications XML représentatives de l'existant des bibliothèques numériques.
- 2 – Comment pouvons-nous modéliser l'analyse du XML et des standards associés appliqués à des bibliothèques numériques ? Qu'en déduire de l'intérêt du langage XML dans le cadre des bibliothèques numériques ?
Nous proposerons une grille d'analyse des bibliothèques numériques, qui mettra en valeur leurs caractéristiques, voire même leurs spécificités du fait de l'intégration du XML et des standards associés. Notre intention est en effet de définir ce qu'apporte réellement le langage informatique au dispositif numérique, et ce qui fonde la pertinence du XML pour les bibliothèques numériques.
- 3 – Quels sont les aspects techniques, par exemple liés à la syntaxe XML, qui font la valeur ajoutée d'une bibliothèque numérique ? Quels sont les aspects sociologiques, notamment inhérents aux rôles des utilisateurs professionnels et finaux, qui font la qualité d'une bibliothèque numérique ? Quoi ou qui peut définir la pertinence du XML pour les bibliothèques numériques ? Quel est alors l'intérêt de bâtir des stratégies d'évaluation pour l'application du langage informatique ?

Une chose est sûre, la pertinence n'est pas un concept universel. Le langage XML n'est pas non plus une technologie universelle, mais une technologie qui s'adresse à un ensemble de personnes -depuis des concepteurs à des utilisateurs-, partenaires d'un échange, par exemple via des bibliothèques numériques. De l'association de ces acteurs, de la conjonction de tous ces aspects découlera la définition d'indices de pertinence, principal enjeu de notre démarche d'évaluation et peut-être nouvel argument pour une appropriation du XML et des standards associés par les utilisateurs potentiels des bibliothèques numériques.

Démarche de réflexion et plan de rédaction

Notre évaluation de la technologie XML appliquée à la bibliothèque numérique s'inscrit donc dans un double contexte, à la fois technique et sociologique. La dimension technique qui nous préoccupe est celle des qualités du langage informatique (XML) associées à l'élément documentaire (la bibliothèque numérique). Quant à la dimension sociologique, c'est celle des usagers, utilisateurs professionnels et finaux du langage XML et de la bibliothèque numérique. Ces deux dimensions ne sauraient évoluer indépendamment l'une de l'autre. Elles sont complémentaires et doivent permettre de nous situer dans un cadre de référence socio-technique. Pour évaluer efficacement la technologie XML appliquée à la bibliothèque numérique, il faudra donc considérer successivement :

- les avantages effectifs du XML par rapport à un autre langage (principalement HTML ou même SGML), voire par rapport à toute autre solution informatique (par exemple une base de données relationnelle) envisageable pour le développement des bibliothèques numériques ;
- les résultats réels obtenus par les utilisateurs des bibliothèques numériques développées avec le langage XML et les standards associés.

C'est pourquoi nous avons élaboré un plan en trois parties :

- 1 – Réalité du langage XML et des bibliothèques numériques : cadre global de développement et cas d'études spécifiques
- 2 – Réalisation d'une grille d'analyse pour l'évaluation du XML appliqué aux bibliothèques numériques
- 3 – Résultats de l'évaluation et définition d'indices de pertinence du XML

1 REALITE DU LANGAGE XML DANS LES BIBLIOTHEQUES NUMERIQUES : **CADRE GLOBAL DE DEVELOPPEMENT ET CAS D'ETUDES SPECIFIQUES**

Nous avons choisi de commencer par l'élaboration d'une typologie des bibliothèques numériques françaises. Elle devra non seulement nous permettre de dresser un panorama général des applications existantes ou en devenir, mais aussi d'évaluer la présence du XML et des standards associés dans ces applications. Il sera par exemple intéressant de nous demander :

- si XML et les standards associés sont particulièrement utilisés dans une catégorie donnée de bibliothèque numérique ;
- dans quels cas l'intégration du XML et des standards associés peut constituer une valeur ajoutée (sous quelles contraintes et avec quelles perspectives) ;
- dans quels autres cas l'intégration du XML et des standards associés peut sembler encore inutile.

1.1 Petite typologie des bibliothèques numériques courantes

En terme de réalisations concrètes, les bibliothèques numériques revêtent de très nombreux aspects. Pour les physiciens, elles représentent généralement d'importants réservoirs de documents non structurés, non indexés, non validés. Pour les chercheurs du secteur médical, les bibliothèques numériques sont certes plus organisées que celles des physiciens, mais se résument souvent à des accès à des périodiques électroniques ultra spécialisés et hautement contrôlés. Pour les chercheurs en sciences humaines, les bibliothèques numériques sont plutôt des réservoirs de documents littéraires pour lesquels la normalisation est indispensable étant donné que les documents font l'objet de nombreuses éditions, et qu'ils doivent pouvoir être manipulés, annotés, comparés, échangés. Enfin, pour la lecture publique, les bibliothèques numériques constituent des collections de documents numérisés, textes anciens, manuscrits...

Une telle variété de facettes rend d'autant plus nécessaire l'élaboration d'une typologie des bibliothèques numériques, de façon à les classer selon certains critères évidents. Le premier critère de distinction que nous avons choisi, est celui de la fonction, ou prérogative. Quelle

fonction principale est attachée à telle ou telle bibliothèque numérique ? Nous avons distingué trois grandes fonctions :

- diffusion du patrimoine
- diffusion de la recherche universitaire
- diffusion de l'information périodique

Dans le cas de la diffusion du patrimoine, il nous a semblé nécessaire de définir un second critère de distinction, celui de la nature du patrimoine. Quelle est la nature principale de l'information contenue dans telle ou telle bibliothèque numérique ? Nous avons choisi de classer les bibliothèques numériques patrimoniales selon qu'elles contiennent :

- un patrimoine culturel : par exemple, relevant de la littérature, de l'art, de l'archéologie, de la musique...
- un patrimoine scientifique et technique
- un patrimoine juridique

Il est bien évident que nous ne pouvons nous attacher qu'aux types les plus courants de bibliothèques numériques ; notre souci est moins d'être exhaustif, que de dégager des modèles d'organisation et de diffusion de l'information, destinés à l'évaluation. Il convient de remarquer à cet égard que les bibliothèques numériques intégrant XML et citées à titre d'exemple dans la typologie, ne font dans une première partie que l'objet d'une présentation relativement succincte : bref historique, contenu informationnel et interrogations soulevées. Ce n'est que dans un deuxième développement que nous les soumettons à évaluation, notre intention étant de déduire du contexte général de développement des bibliothèques numériques, des principes d'évaluation ultérieure de celles-ci.

1.1.1 Les bibliothèques numériques patrimoniales

1.1.1.1 Principales caractéristiques de l'existant

Toute instance publique ou privée, détentrice d'un patrimoine, c'est-à-dire d'un ensemble ordonné de connaissances, peut décider de constituer une bibliothèque numérique patrimoniale. En d'autres termes, une bibliothèque numérique patrimoniale fait état d'une collection électronique de documents, textes et images, constituant le bien d'une instance telle qu'une bibliothèque physique, une administration, une institution... Par exemple, des documents d'archives, des textes de lois, des textes d'auteurs peuvent constituer une collection patrimoniale. Celle-ci peut être rendue publique, autrement dit devenir le bien de

tous. Elle présente de grands intérêts, soit du fait du caractère esthétique de ses documents (iconographies), soit grâce au contenu intellectuel de ceux-ci (écrits).

De fait, les publics des bibliothèques patrimoniales, d'autant plus si elles sont numériques donc accessibles n'importe où via l'informatique, sont nombreux et variés : historiens, historiens d'art, chercheurs sur les textes et philologues, chercheurs sur les images, scientifiques, juristes... Pour toute bibliothèque numérique patrimoniale, multiplier les angles de vision et de présentation des documents est donc particulièrement nécessaire pour restituer leur véritable richesse. C'est là l'une de leurs premières contraintes : s'adapter à chaque profil d'utilisateur, personnaliser le rapport de l'utilisateur avec l'outil numérique.

1.1.1.2 La diffusion du patrimoine culturel

Si une majorité de bibliothèques numériques patrimoniales est aujourd'hui constituée de collections électroniques de textes d'auteurs, c'est évidemment du fait de l'extrême richesse du patrimoine littéraire français, ou même francophone, depuis le Moyen-Age jusqu'à nos jours.

1.1.1.2.1 Gallica



Figure 1 : page d'accueil <http://gallica.bnf.fr>

Les créations numériques de la BNF (Bibliothèque nationale de France), sont particulièrement représentatives du potentiel actuel du XML et des standards associés. En effet, la BNF a entrepris depuis quelques temps la conception d'applications XML sous différentes formes : bibliographies, guides de ressources Web... En effet, des applications telles que la « Bibliographie nationale française » et les « Signets », ont permis à la BNF d'expérimenter le langage XML et les standards associés sur de « petites » bases de données, avant d'entreprendre l'intégration du XML pour sa grande bibliothèque numérique, plus connue sous le nom de « Gallica ». L'idée d'une « très grande bibliothèque d'un genre entièrement nouveau », apparue en juillet 1988 dans un discours du président de la République François Mitterrand, n'a jamais été aussi prégnante, à l'heure où Gallica s'associe les atouts techniques du XML et des standards associés, en particulier les qualités de structuration et de personnalisation de l'information.

1.1.1.2.2 Guide de l'Internet culturel

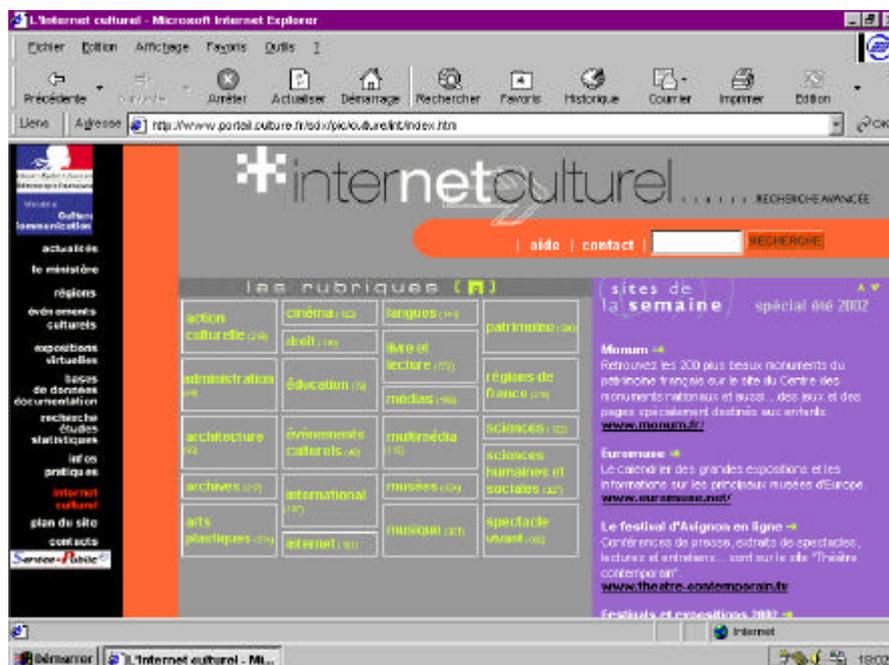
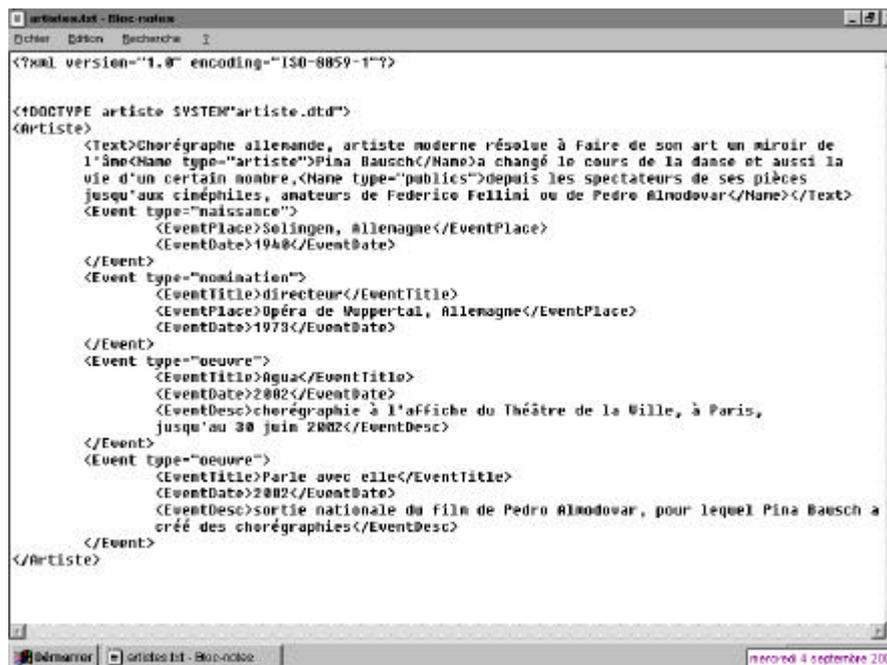


Figure 2 : page d'accueil <http://www.portail.culture.fr/sdx/pic/culture/int/index.htm>

A l'image de la BNF avec Gallica, le Ministère de la Culture et de la Communication a mis en place un large processus de développement de bibliothèques numériques, à l'aide du XML et des standards associés. Au cours de l'année 2001, le Ministère, plus particulièrement la

Mission de la Recherche et de la Technologie (MRT), a ainsi développé plusieurs applications documentaires intégrant XML, parmi lesquelles les inventaires de villes (La Rochelle, Lyon...). En mars 2001 naît le «Guide de l'Internet culturel», service phare de la MRT sur le Web. Il fédère plusieurs ressources numériques d'information culturelle. A ce titre, l'Internet culturel est une réalisation informatique patrimoniale d'un autre genre que toutes les autres applications citées en exemples dans ce mémoire. En effet, il ne constitue pas réellement une bibliothèque numérique en tant que collection organisée de documents numérisés. Mais il s'agit surtout d'un portail vers plusieurs bibliothèques numériques issues de la numérisation du patrimoine culturel français. C'est pourquoi, nous avons malgré tout choisi à ce titre de l'intégrer à la typologie des bibliothèques numériques.

1.1.1.2.3 Artistes et Œuvres



```

artistes.dat - Bloc-notes
D:\citer  Edition  Recherche  I

<?xml version="1.0" encoding="ISO-8859-1"?>

<!DOCTYPE artiste SYSTEM"artiste.dtd">
<Artiste>
  <Text>Chorégraphe allemande, artiste moderne résolue à faire de son art un miroir de
  l'âme<Name type="artiste">Pina Bausch</Name>a changé le cours de la danse et aussi la
  vie d'un certain nombre,<Name type="public">depuis les spectateurs de ses pièces
  jusqu'aux cinéphiles, amateurs de Federico Fellini ou de Pedro Almodovar</Name></Text>
  <Event type="naissance">
    <EventPlace>Selingen, Allemagne</EventPlace>
    <EventDate>1948</EventDate>
  </Event>
  <Event type="nomination">
    <EventTitle>directeur</EventTitle>
    <EventPlace>Opéra de Wuppertal, Allemagne</EventPlace>
    <EventDate>1973</EventDate>
  </Event>
  <Event type="oeuvre">
    <EventTitle>Agua</EventTitle>
    <EventDate>2002</EventDate>
    <EventDesc>chorégraphie à l'affiche du Théâtre de la Ville, à Paris,
    jusqu'au 30 juin 2002</EventDesc>
  </Event>
  <Event type="oeuvre">
    <EventTitle>Parle avec elle</EventTitle>
    <EventDate>2002</EventDate>
    <EventDesc>sortie nationale du film de Pedro Almodovar, pour lequel Pina Bausch a
    créé des chorégraphies</EventDesc>
  </Event>
</Artiste>

```

Figure 3 : un exemple de document XML pour "Artistes et Œuvres"

Le Centre national de la Danse (CND) est un établissement public créé en 1998, sous la tutelle du Ministère de la Culture et de la Communication. La dynamique de création numérique entreprise par le Ministère a donc aussi marqué l'activité du CND. En 2000, le CND a donc décidé la création d'une bibliothèque numérique spécialisée en danse, et reposant sur le langage XML. « Artistes et Œuvres » est actuellement en phase de production test ; les

documents XML sont validés par la DTD créée spécifiquement pour le fonds documentaire. D'après le responsable du projet, la possibilité de « structurer l'information de façon pérenne et nette, sans faire fi des nuances essentielles de l'univers auquel on se réfère », a grandement motivé le choix du XML et des standards associés pour le développement de la bibliothèque numérique. Ainsi, le développement de la bibliothèque numérique du CND a notamment demandé l'élaboration d'une hiérarchie simple (« structurer l'information de façon nette »), ainsi que la définition de balises spécifiques au champ d'application de l'information (« nuances essentielles de l'univers auquel on se réfère »).

1.1.1.3 La diffusion du patrimoine scientifique et technique

1.1.1.3.1 Article@INIST

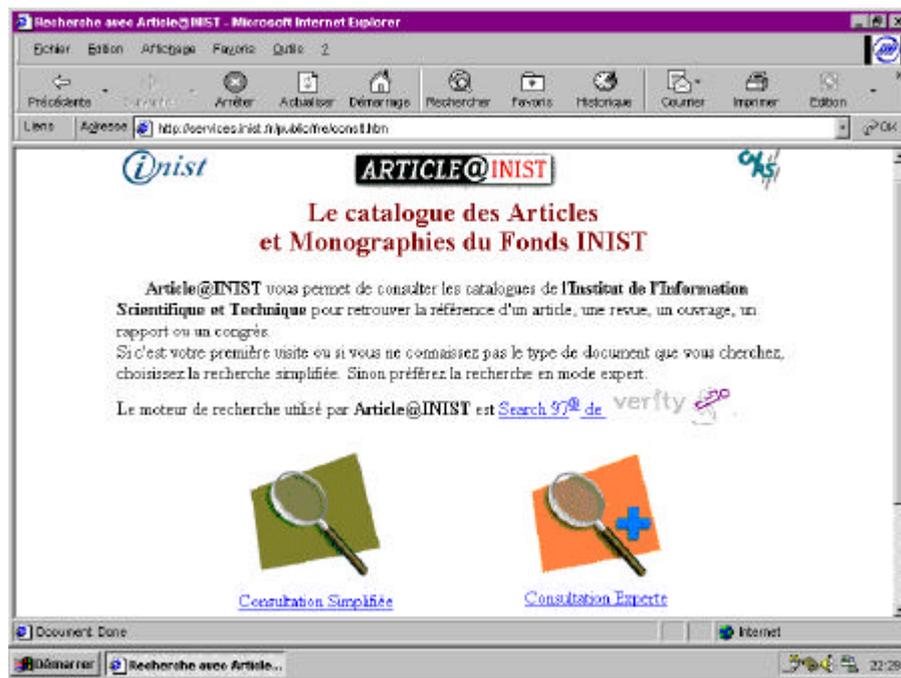


Figure 4 : page d'accueil <http://services.inist.fr/public/fr/conslt.htm>

L'INstitut de l'Information Scientifique et Technique (INIST) est une unité de service du Centre National de la Recherche Scientifique et technique (CNRS), en charge de la collecte, de l'analyse et de la diffusion des résultats de la recherche scientifique internationale. L'INIST dispose d'un fonds documentaire spécialisé en sciences, technologie, médecine, sciences humaine et sociale, qu'elle met à disposition via plusieurs outils :

- ArticleSciences : moteur de recherche et de commande d'articles scientifiques
- ConnectSciences : portail de l'information scientifique et technique
- Thesa : base de données des thèses en cours dans les grandes écoles
- Services@INIST, comprenant Form@INIST, Compte@INIST et surtout Article@INIST qui est une bibliothèque numérique d'articles et de monographies scientifiques et techniques

« Article@INIST » est une bibliothèque numérique pluridisciplinaire permettant non seulement la recherche de documents -monographies et revues- scientifiques et techniques, mais aussi leur commande en ligne. Etant donné la taille du fonds (plus de 5 millions de références en ligne !), la bibliothèque numérique se doit d'être évolutive. C'est pourquoi elle intègre depuis peu la technologie XML.

1.1.1.4 La diffusion du patrimoine juridique

1.1.1.4.1 Naiade

```

naide.txt - Bloc-notes
Fichier  Édition  Recherche  ?

<?xml version="1.0" encoding="ISO-8859-1"?>

<!DOCTYPE circulaire
[
<ELEMENT zometitre {typecir, numcir?, datecir, objet, pjanmex}>
<ELEMENT typecir {codecir, compcir?}>
<ELEMENT codecir {NPCDATA}>
<ELEMENT compcir {NPCDATA}>
<ELEMENT numcir {NPCDATA}>
<ELEMENT datecir {%datef;}>
<ELEMENT objet {allinea+}>
<!-- pjanmex est affiché sous forme de lien (référence aux PJ ou annexes) -->
<ELEMENT pjanmex {%infobase;}*>
<!-- piècejointe : définit si PJ ou Annexe (0-PJ, N-Annexe) -->
<!-- lienpj : correspondance avec l'identifiant de l'annexe (attribut idannexe) -->
<ATTLIST pjanmex
    piècejointe {%choix;} #REQUIRED
    lienpj IDREF #IMPLIED
]>
]>

```

Figure 5 : un exemple de fichier DTD pour "Naiade"

En juin 2000, l'Établissement National des Invalides de la Marine (ENIM) décide la création d'une structure documentaire décentralisée, le CEDIF, dont la mission est de construire un

intranet juridique : « Naï ade ». Il est destiné à l'information des assurés de l'ENIM (qui gère le régime social des marins professionnels français), et constitue pour ce faire une véritable bibliothèque numérique de textes juridiques, de documents de jurisprudence, de fiches techniques...

1.1.1.4.2 Légifrance



Figure 6 : page d'accueil <http://www.legifrance.gouv.fr>

« Légifrance » est une initiative des pouvoirs publics annoncée par le Premier Ministre dans son discours d'Hourtin du 25 août 1997 et précisée depuis dans le Plan d'action gouvernemental « Préparer l'entrée de la France dans la société de l'information (PAGSI) ». Il est destiné à apporter au public français et étranger les informations juridiques françaises essentielles et en vigueur. Ouvert en février 1998, il s'est progressivement enrichi, tant du point de vue du contenu intellectuel (ajout de codes et de textes fondamentaux consolidés...), que de celui de l'ergonomie. La version actuellement accessible (la troisième à ce jour !) a vu le jour en janvier 2001.

1.1.2 Les bibliothèques numériques universitaires

1.1.2.1 Principales caractéristiques de l'existant

Ce type de bibliothèque numérique peut être considéré comme un sous-ensemble des bibliothèques numériques patrimoniales. En effet, les universités sont bien détentrices d'un patrimoine littéraire, mais d'une nature particulière ; il s'agit en effet exclusivement de travaux de recherche encadrés et effectués dans le cadre d'un enseignement spécifique. Cela suffit à justifier d'un type spécifique de bibliothèque numérique. La recherche universitaire constitue une ressource importante d'information scientifique et technique, que les bibliothèques numériques se chargent de faire parvenir au plus grand nombre. Les bibliothèques numériques universitaires servent non seulement à la transmission des connaissances, mais également à l'évaluation des chercheurs. Leur renommée peut aujourd'hui tenir pour beaucoup à la consultation de leurs travaux sur des bibliothèques numériques. Mais au-delà de la renommée du chercheur, c'est aussi le rayonnement et la reconnaissance des universités françaises qui sont en jeu. Pour tenir leur rang dans le contexte international, les universités françaises se doivent d'avoir une politique ambitieuse pour leur recherche. Le développement de bibliothèques numériques participe de cette politique ambitieuse ; les universités devenant ainsi de véritables gisements d'information au même titre que des bibliothèques ou des centres de documentation. Mais à la différence des bibliothèques numériques patrimoniales des bibliothèques, les bibliothèques numériques universitaires sont contraintes par la forme des documents. Il ne peut en effet s'agir que de mémoires de fin d'études, de thèses, de compte-rendus de travaux de laboratoire ou de rapports scientifiques. Les deux formes les plus courantes sont les mémoires et les thèses.

En outre, dans le milieu universitaire le développement de bibliothèques numériques repose sur l'association de différents acteurs. Outre les cellules de recherche, les départements d'informatique sont les partenaires naturels de ce développement. Administrations, associations professionnelles, sociétés commerciales et fondations viennent compléter le panorama des sources potentielles de soutien.

1.1.2.2 La diffusion de la recherche universitaire

1.1.2.2.1 CodeX

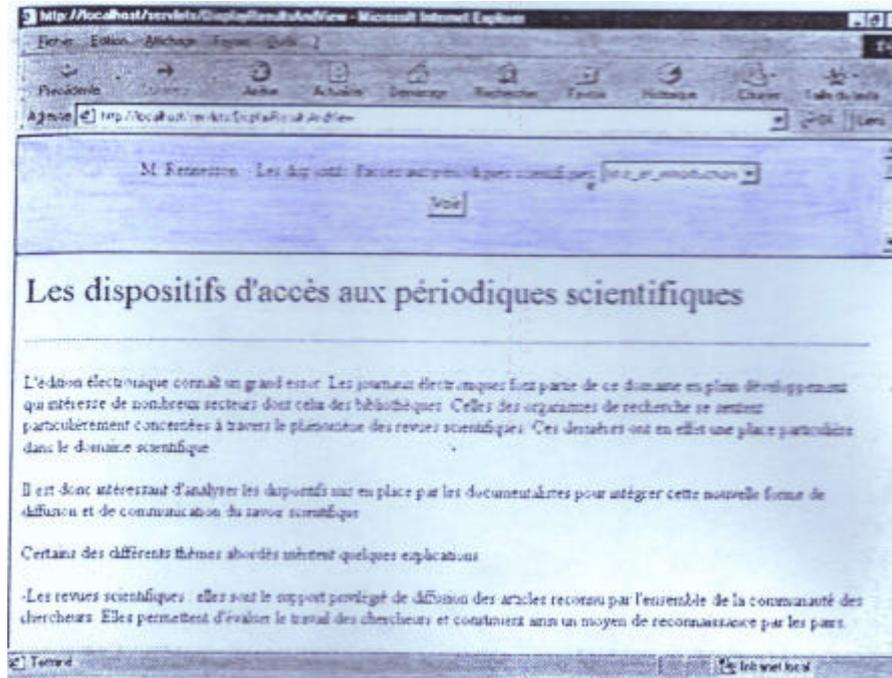


Figure 7 : une « vue » particulière (titre et introduction) d'un document XML

Parmi les bibliothèques numériques universitaires développées avec XML et les standards associés, le projet « CodeX » (Consultation et Organisation de Documents Electroniques à Paris-X)² est une plate-forme de diffusion de recherche, développée par l'université de Paris X-Nanterre. CodeX a pour but de concevoir un système destiné à rendre accessible une base de documents scientifiques sur le Web, selon des modes de consultation personnalisés. Les documents concernés sont des mémoires de maîtrise et de DEA en Sciences de l'Information et de la Documentation, qui ont été convertis en XML à partir de leurs formats bureautiques d'origine.

² CHAUDIRON Stéphane, IHADJADENE Madjid, ROLE François, « Exploitation de l'indexation par les auteurs pour le filtrage de documents structurés », in *Filtrage et résumé automatique de l'information sur les réseaux – Actes du 3^{ème} Colloque du Chapitre français de l'ISKO*, sous la direction de CHAUDIRON S. et FLUHR C., Nanterre 5-6 juillet 2001, Université de Paris X, p.39-47

CHAUDIRON Stéphane, IHADJADENE Madjid, ROLE François, « CodeX : un système pour la définition de vues multiples guidées par les usages », in *Document électronique dynamique - Actes du 3^{ème} Colloque international sur le document électronique (CIDE 2000)*, sous la direction de GAIO M. et TRUPIN E., Lyon 4-6 juillet 2000, Europa Productions, p.71-81

1.1.2.2.2 Pelleas

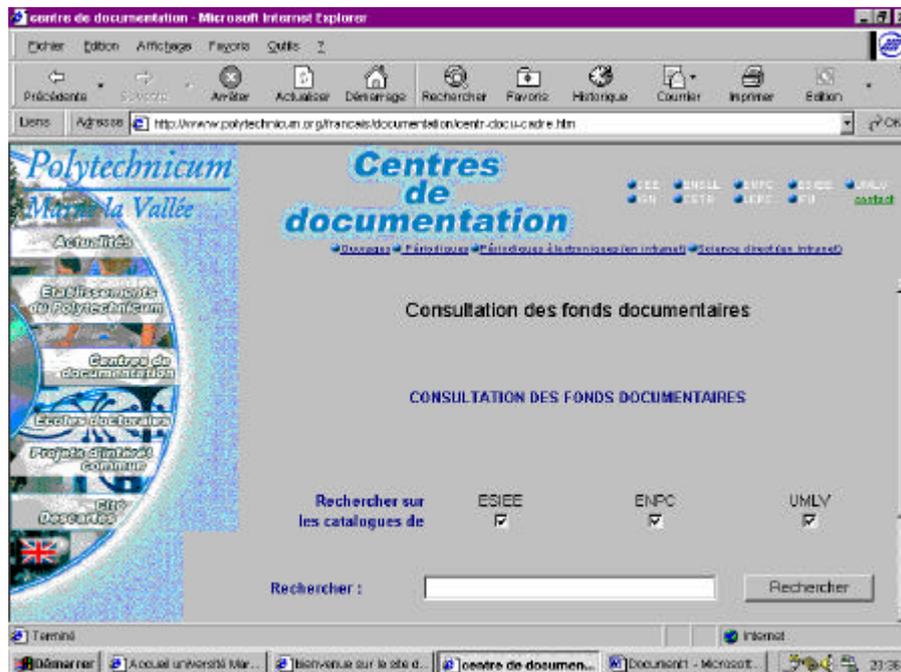


Figure 8 : page d'accueil <http://www.polytechnicum.org/francais/documentation/centr-docu-cadre.htm>

« Pelleas » est un projet élaboré par l'Université de Marne-la-Vallée (UMLV), associée en consortium avec d'autres universités, parmi lesquelles celle de Reims, et avec des centres de recherche. Créé en 1994, le Polytechnicum de Marne-la-Vallée regroupe ainsi 15 établissements d'enseignement supérieur et de recherche, à statuts divers, sur le site de la ville nouvelle de Marne-la-Vallée. Parmi les membres du réseau, on compte aussi bien l'INA (Institut national de l'Audiovisuel) que le CSTB (Conseil supérieur et technique du Bâtiment). Le Polytechnicum a notamment pour but de mettre en œuvre une bibliothèque numérique, intégrant tous types de ressources appartenant à l'environnement des étudiants, des enseignants et des chercheurs. Le lancement officiel de cette bibliothèque numérique est prévu pour septembre 2002.

1.1.3 Les bibliothèques numériques de presse

1.1.3.1 Principales caractéristiques de l'existant

Au même titre que les bibliothèques numériques universitaires, la presse électronique est en réalité un sous-ensemble de bibliothèques numériques patrimoniales. Pourrait-on en effet

autrement considérer la presse que comme un patrimoine littéraire ! Cependant, sa grande hétérogénéité et les nombreuses difficultés que son existence électronique supposent (notamment la question des droits juridiques), font que la presse électronique requièrent un traitement particulier. Les bibliothèques numériques de presse peuvent se présenter sous deux formes différentes : la première dédiée à un seul titre de périodique (version électronique d'un journal ou d'une revue), la seconde intègre plusieurs titres de périodiques, mais souvent dans un domaine donné (agrégateur de presse). Concernant la première forme de bibliothèque numérique de presse, la plupart des publications périodiques, revues et journaux, disposent aujourd'hui d'une version électronique, en plus de leur version papier. Certains titres n'existent même que dans un format électronique. Tous constituent de véritables bibliothèques numériques, associant une version intégrale de l'édition du jour, à des dossiers d'archives. Pour se distinguer les unes des autres, ces nouvelles bibliothèques numériques se doivent, plus que pour tout autre type, d'explorer, d'exploiter les voies de la diffusion multimédia en réseau. En ce sens, elles intègrent généralement des illustrations sonores ou vidéo, ou encore s'adjoignent les services d'applications informatiques permettant des animations et des modèles (applets JAVA, modélisations 3D...).

Ici les cas d'applications du XML et des standards associés sont rares ; sans doute en grande partie parce que la structuration de l'information périodique est la moins évidente, et aussi parce que les enjeux économiques liés à la diffusion de la presse, très contraignants, ne facilitent pas l'adoption de nouveaux modes d'organisation de l'information. Le cas du Point est donc relativement isolé !

1.1.3.2 La diffusion de l'information périodique

1.1.3.2.1 Le Point

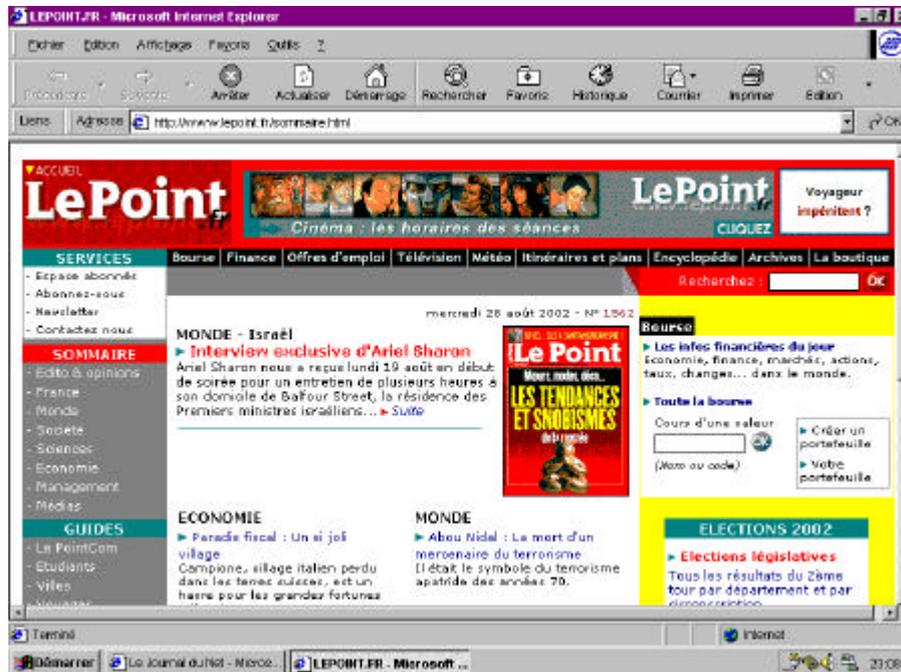
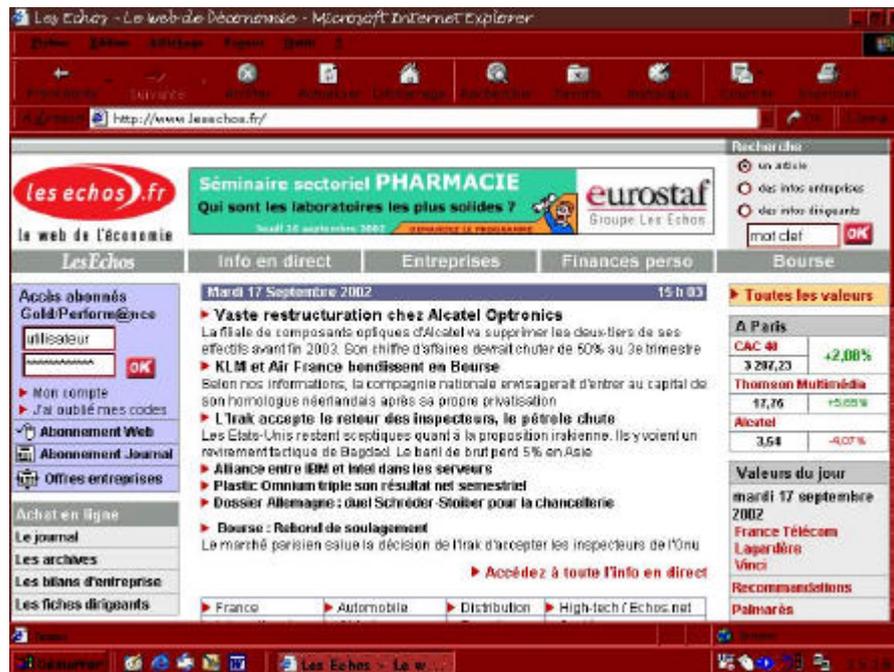


Figure 9 : page d'accueil <http://www.lepoint.fr>

En novembre 1999, Le Point, filiale de Artémis S.A., holding du groupe PPR (Pinault-Printemps-LaRedoute) décide de lancer deux sites Internet : « lepoint.fr », version électronique de l'hebdomadaire créé en 1972, et « lepoint-etudiants.com », service d'informations aux étudiants. La version électronique du journal Le Point est récemment passée de HTML à XML. L'objectif principal de la nouvelle version, en ligne depuis le 22 décembre 2001, est de s'assurer un meilleur contrôle du système de publication, depuis le traitement de l'information jusqu'à sa diffusion, afin notamment de s'assurer une plus grande indépendance (par exemple vis-à-vis des prestataires techniques extérieurs). Quant à améliorer la hiérarchisation de l'information, cela constitue un argument supplémentaire pour le choix du XML.

1.1.3.2.2 Les EchosFigure 10 : page d'accueil <http://www.lesechos.fr>

« Lesechos.fr », version électronique actuelle du quotidien économique, constitue l'aboutissement d'un chantier entrepris en 2000. Il portait sur l'optimisation de l'information proposée aux lecteurs du quotidien en ligne. Il importait alors « d'assurer une transversalité totale de l'information. » Le choix du XML, qui structure et indexe l'information, donne de la flexibilité dans le traitement et la diffusion de l'information. Les flux XML permettent d'envoyer des informations spécifiques à une société cliente suivant ses besoins en matière de veille.

1.2 Contraintes générales de développement en XML

Qu'exige aujourd'hui le choix du XML et des standards associés pour le développement de bibliothèques numériques ? Quels sont les tenants et les aboutissants d'une telle entreprise ? La réponse à ces questions devra nous permettre de déterminer un premier cadre de définition d'indices de pertinence. Cette partie s'apparente quelque peu à une marche à suivre afin de justifier du choix du XML pour le développement actuel des bibliothèques numériques. Elle n'a rien d'une loi absolue, mais vise à mieux démontrer que la pertinence d'une technologie, telle que XML, dépend avant tout d'indices contextuels. L'intégration du XML aux bibliothèques numériques s'ancre dans une actualité, plus exactement dans une réalité très particulière, dont nous devons d'abord connaître les caractéristiques pour mieux évaluer la pertinence du XML.

1.2.1 S'affranchir d'un contexte ambivalent

Comment expliquer que le XML, en développement depuis 1998, ne s'impose-t-il pas plus vite sur le Web ? Pourquoi le XML ne supplante-t-il pas logiquement le HTML, alors que l'on peut lire partout qu'il est plus rigoureux, mais aussi plus ouvert que lui ? Vanter les performances techniques du XML suffit-il à convaincre les professionnels de son intérêt, notamment dans le cadre très précis du développement de bibliothèques numériques ?

1.2.1.1 Un nombre limité d'applications

Si l'on examine aujourd'hui l'ensemble des applications XML et des standards associés, on constate qu'elles appartiennent majoritairement à trois secteurs professionnels : la documentation, l'édition et le commerce électronique. Pourtant, malgré une présence forte de ces secteurs sur Internet, les applications XML constituent à peine plus de 1% du réseau Internet.

De récentes conférences sur le XML³ ont montré que, malgré des avantages certains, il n'était pas forcément cette révolution qui promettait de reléguer ses ascendants (HTML, SGML) au rang de pièces de musée. En effet, HTML fait de la résistance, et demeure le principal outil de gestion et de partage des données sur la Toile. L'état de l'existant que nous avons précédemment dressé pour les bibliothèques numériques, nous a montré que le taux de représentation du XML par rapport à d'autres formats de développement, est encore restreint.

Plusieurs éléments plaident contre une généralisation rapide du XML. Tout d'abord, contrairement à HTML, XML n'est pas « prêt à l'emploi ». L'intégration des données XML, surtout pour des systèmes complexes d'information comme les bibliothèques numériques, est longue car très normalisée. Pour développer en XML, il ne suffit pas de récupérer des données. Il faut surtout les structurer, les mettre en relation (élaborer une hiérarchie, trouver des rapports d'association...). Ensuite, pour échanger ces données structurées, les professionnels doivent adapter leurs outils de communication actuels (compatibilité des plateformes d'échange...). Ce qui fait que le développement en XML requiert un investissement encore lourd, tant en charges de travail qu'en coûts purement financiers. La tendance générale concernant le développement en XML, est encore d'attendre pour voir : non seulement voir si l'investissement requis peut être allégé, mais aussi si les expériences entreprises avec XML se révèlent de réels succès. Faut-il donc « attendre et voir venir » ? Il est d'autant plus intéressant aujourd'hui d'analyser quelles motivations poussent certains secteurs professionnels à s'affranchir de la tendance générale pour gérer des bibliothèques numériques à l'aide du XML et des standards associés.

1.2.1.2 Une motivation croissante des professionnels

En 2001, nous avons réalisé une enquête intitulée « quelle place pour le XML dans la culture documentaire ? »⁴. Elle mettait en évidence les motivations des professionnels de l'information et de la documentation, concernant leur appropriation du langage XML. Il est intéressant de voir que l'espoir de normalisation, la liberté de création et la qualité

³ ADBS [Association des professionnels de l'information et de la documentation], *XML, pour quoi faire ?*, Paris-La Défense, 4 décembre 2001

Université Paris 13-UFR des Sciences de la Communication, *L'édition multi-supports, une révolution éditoriale ? Questions posées par l'adoption des standards de description des documents : l'exemple de XML*, Villetaneuse, 14 juin 2001

⁴ Enquête réalisée dans le cadre du mémoire de maîtrise « Pour une appropriation du langage XML par les professionnels de l'information et de la documentation », Université de Paris XIII-Villetaneuse, Septembre 2001

d'adaptation du XML sont les trois motivations principales, mises en exergue par les professionnels. Comment cela se traduit-il dans la pratique ? Ou comment passer d'une connaissance générale à une pratique particulière ?

L'un des atouts du XML est qu'il permet de créer des documents composites à partir de sources variées, de modéliser l'information par une structuration textuelle normalisée (balisage répondant à certaines règles) et spécifique (balisage adapté à la nature de l'information), enfin de décliner les présentations de l'information en fonction des usages et des publics. Toutes ces possibilités participent de l'espoir de normalisation et de la liberté de création souhaités par les professionnels. Par exemple, la possibilité de créer des affichages personnalisés de l'information est mise en valeur, notamment pour des bibliothèques numériques pluridisciplinaires, dont les publics sont en toute logique les plus variés. Quant à la possibilité de modéliser l'information par une structuration textuelle spécifique, elle est d'autant plus appréciée pour des fonds documentaires spécialisés, car cela permet de définir des balises précises et adaptées à la nature de l'information structurée. Cet atout est particulièrement prisé dans les bibliothèques numériques patrimoniales.

De plus, à partir de la syntaxe XML originelle, chaque discipline scientifique, corps de métier ou même entreprise, peut développer son propre formatage de données. Il est en effet possible d'adapter la syntaxe XML à un vocabulaire spécifique, qu'il se rapporte au secteur économique (ex : transactions financières...), scientifique (ex : formules mathématiques, molécules chimiques...), ou même artistique (ex : partitions de musique...). Témoignent de cette qualité d'adaptation du XML les exemples déjà éprouvés de MathML (Mathematical Markup Language), CML (Chemical Markup Language), AIML (Astronomical Instrument Markup Language) ou encore eb-XML (electronic business XML). Ainsi MathML permet l'affichage des notations mathématiques, l'échange de formules. CML donne des représentations graphiques de la structure moléculaire des composés chimiques. AIML gère notamment les coordonnées célestes des galaxies. Quant à eb-XML, il est développé pour le commerce électronique et s'attache aux places de marché... La qualité d'adaptation du XML est d'autant plus réelle que des secteurs aussi pointus que l'industrie agro-alimentaire ou la recherche génétique, ont eux aussi mis en œuvre des projets de langages spécifiques : CheeseML pour le traitement des données de l'agro-alimentaire, BSML (Biosequence ML) pour la représentation de séquences de code génétique...

Recenser et mettre en valeur les réalisations et les projets intégrant le XML et les standards associés permettent de montrer que l'espoir de normalisation, la liberté de création et la qualité d'adaptation ne sont pas que des formules abstraites ; elles peuvent avoir une représentation concrète, adaptée à un contexte et à des besoins professionnels spécifiques. Chaque expérience peut aussi prouver la faisabilité, voire la réussite de l'application du XML et des standards associés à une bibliothèque numérique. La multiplication d'exemples, donc « l'accumulation de preuves », contribueront sans doute à faire croître la motivation des professionnels.

S'affranchir d'un contexte ambivalent, c'est-à-dire partagé entre une théorie séduisante (le XML doit enrichir les pratiques documentaires !) et une pratique limitée (le nombre réel d'applications documentaires en XML est encore restreint !) : tel est aujourd'hui la première exigence du choix du XML et des standards associés pour les bibliothèques numériques.

1.2.2 Dégager des axes prioritaires

1.2.2.1 L'information des utilisateurs réels et potentiels

Avec le langage XML, les acteurs informatiques s'accordent enfin à miser sur un standard commun, qui s'adapte mieux à la réalité du Net et aux nombreuses applications qui sont apparues, parmi lesquelles les bibliothèques numériques. A l'évidence, la technologie XML suscite de l'intérêt ; depuis son apparition en 1998, les réalisations avec XML se sont multipliées, notamment dans le domaine documentaire. Toutefois, la technologie XML suscite autant d'intérêt que de perplexité, et ce malgré les réussites de nombreuses applications et les efforts du marché pour rendre l'offre XML plus accessible et plus diversifiée. Il apparaît de fait toujours nécessaire de convaincre les utilisateurs de la pertinence du XML.

De récents sondages ont montré que le XML demeurait encore difficile à appréhender, à définir même simplement. Dans un sondage du Journal du Net⁵ en 2001, près d'un quart des personnes interrogées sur la définition du XML, répondait : « c'est quoi le XML ? » (23,4%). S'ajoutent encore les sceptiques qui s'interrogent sur les finalités et même l'utilité du XML (7,7% des votants du sondage pensent que XML est une « fausse bonne idée » tandis que

⁵ DIZ José, « Le langage XML doit encore convaincre ! », *JNet Solutions*, 2 février 2001

6,3% le jugent « trop complexe »). Sans pour autant connaître les aspects très techniques de ce langage, il serait donc utile qu'aujourd'hui les utilisateurs sachent non seulement ce que peut représenter l'apport de ce langage plus qu'émergent, mais aussi dans quels cas il est soit pertinent, soit inapproprié.

En effet, la mise en œuvre du XML, notamment dans le développement des bibliothèques numériques, suppose un contexte approprié, où la connaissance des atouts du XML doit précéder l'analyse de faisabilité qui décidera du choix ou du rejet du XML, en fonction des besoins et des objectifs. Pour des documents similaires, traités dans des contextes proches, deux entreprises peuvent opter pour des solutions radicalement différentes ; par exemple choisir de rassembler les documents dans une base de données classique telle qu'Access, ou décider la création d'une bibliothèque numérique de documents codés en XML. Il convient donc plus que jamais d'observer (« enquêter ») pour trouver les indices susceptibles de nous orienter vers certaines solutions, de nous mener peut-être jusqu'à la pertinence du XML pour le développement des bibliothèques numériques.

1.2.2.2 La formation des utilisateurs professionnels et finaux

Tout utilisateur, qu'il soit utilisateur professionnel ou utilisateur final, a sans doute déjà été formé aux usages des bibliothèques numériques. Or le langage XML et les standards associés appliqués aux bibliothèques numériques, changent la donne et constituent une nouvelle variable de « l'apprentissage » des utilisateurs. Ils doivent avoir connaissance des performances des bibliothèques numériques liées à l'application du XML et des standards associés. Forcément avec XML, les bibliothèques numériques s'associent de nouvelles fonctionnalités, parfois plus complexes à appréhender. La formation des utilisateurs professionnels et finaux devient donc un défi essentiel des bibliothèques numériques, surtout qu'elles sont en évolution permanente. Pour autant, il est encore difficile de définir clairement les modalités d'un tel processus de formation : comment sont transmises les connaissances ?

La formation s'effectue à un double niveau, celui des professionnels (« administrateurs ») des bibliothèques numériques, et celui des utilisateurs de ces mêmes bibliothèques numériques. Dans un premier temps, la formation concerne donc les professionnels des bibliothèques numériques, en particulier les informaticiens et les bibliothécaires ou les documentalistes.

Pour ces deux catégories, les connaissances et les techniques liées au développement d'applications documentaires, telles que des bibliothèques numériques, ne cessent de se renouveler et laissent de côté les barrières entre disciplines. C'est d'autant plus vrai avec le XML dont la pratique ne doit pas être réservée aux informaticiens, d'autant que le langage est construit selon des préoccupations documentaires (préservation de l'information primaire, personnalisation de l'information diffusée...). La formation des informaticiens et celle des bibliothécaires ou des documentalistes sont évolutives et leurs connaissances respectives du langage XML et des standards associés sont complémentaires. Dans un second temps, ces mêmes professionnels peuvent transmettre leurs connaissances du XML aux utilisateurs, les former aux possibilités du langage affectées aux bibliothèques numériques, de façon à permettre aux utilisateurs d'appréhender globalement le dispositif technique et donc d'entreprendre une démarche autonome sur celui-ci.

La réalité montre que, malgré un potentiel technique fort, le langage XML est encore minoritaire dans le développement des bibliothèques numériques. Cela s'explique par un contexte ambivalent, où les contraintes, notamment de technicité des savoirs et des savoir-faire exigés pour l'intégration du XML, sont trop lourdes et ont encore souvent raison de la motivation des professionnels. Autrement dit, le choix du XML pour les bibliothèques numériques, est partagé. A ce stade, comment pourrions-nous le dire pertinent ?

- Sur quoi repose la définition d'une pertinence du XML dans le développement des bibliothèques numériques ?
- Quel rapport cette pertinence entretient-elle avec le cadre social que nous venons juste de définir ?
- Comment évaluer la pertinence du XML dans les bibliothèques numériques, à partir des conclusions que nous venons d'observer sur le cadre actuel de développement ?

Notre intention est désormais de mettre en place une grille d'analyse du XML, que nous appliquerons à des bibliothèques numériques existantes, parmi celles que nous avons

mentionnées ci-avant. L'idée est de déterminer des indices de pertinence valables pour toute évaluation du XML dans les bibliothèques numériques.

2 REALISATION D'UNE GRILLE D'ANALYSE POUR L'EVALUATION DU XML APPLIQUE AUX BIBLIOTHEQUES NUMERIQUES

La multiplication des ressources numériques sur Internet rend d'autant plus nécessaire la mise en place d'un protocole d'appréciation, de remise en cause de la validité de l'information. En effet, contrairement à l'édition papier régulièrement validée par des comités de lecture par exemple, l'information disponible sur le réseau, notamment celle organisée en bibliothèque numérique, est plus complexe à appréhender. L'objectif de cette partie est de fournir un ensemble de critères qui peuvent être utilisés pour évaluer la pertinence du XML dans le développement des bibliothèques numériques. Alors que nous venons de dresser un panorama des bibliothèques numériques courantes, et en particulier de constituer un corpus d'applications XML, il semble désormais intéressant de soumettre ces applications à une évaluation, à la fois en termes de réalisation technique et de coût, mais aussi en termes d'observation et de quantification des usages. Dans cette perspective, nous élaborerons des stratégies d'évaluation prenant en compte les compétences techniques du XML, les particularités des bibliothèques numériques.

Il existe beaucoup de méthodes pour l'évaluation des sites Web, tous domaines confondus. Elles vont du questionnaire général à une liste détaillée de critères. Il y a pourtant peu de principes spécifiques à la technologie XML, en particulier appliquée à des bibliothèques numériques. C'est pourquoi nous sommes partis de certaines méthodes proposées pour l'évaluation de ressources électroniques, que nous avons tenté d'adapter à notre cadre particulier, afin de proposer une grille d'analyse spécifique. Nous avons pris pour référence les articles suivants :

- pour une interprétation de l'interface graphique : « L'écran efficace : une approche cognitive des objets graphiques », article de Xavier Casanova et Joëlle Cohen⁶
- pour une mesure de la qualité de l'information : « A guide to finding quality information on the Internet », ouvrage de Alison Cooke⁷

⁶ CASANOVA Xavier, COHEN Joëlle, « L'écran efficace : une approche cognitive des objets graphiques », *Documentaliste – Sciences de l'information*, 2001, volume 38, n°5-6, p.272-289

⁷ COOKE Alison, *A guide to finding quality information on the Internet : selection and evaluation strategies*, London, Library Association publishing, 2001

- pour une analyse du rapport dispositif-utilisateur : « Interfaces graphiques ergonomiques (conception et modélisation) : programmes et interfaces homme-machine », ouvrage de Jean-Bernard Crampes⁸
- pour un calcul mathématique de la pertinence : « Evaluation des systèmes d'information documentaire », article de Thierry Lafouge⁹

2.1 Données essentielles de l'évaluation

L'élaboration d'une grille d'analyse fonctionnelle d'une technologie telle que le XML, appliquée au cadre très particulier des bibliothèques numériques, repose sur une démarche méthodologique précise. Nous avons aussi élaboré un « mode d'emploi », dont les données principales sont reprises dans un modèle explicatif de la grille d'analyse, présenté en annexe.

2.1.1 Méthodologie

Nous avons choisi de répartir les interrogations selon qu'elles se rapportent aux bibliothèques numériques, ou à la technologie XML. L'association de ces interrogations fera naître une première esquisse d'évaluation. Il nous restera alors à organiser les informations dans une grille descriptive, combinant les réponses relatives aux questions concernant les bibliothèques numériques et à celles relevant de la technologie XML et des standards associés.

Plusieurs questions viennent d'emblée à l'esprit, dans le contexte général d'intégration des nouvelles technologies aux bibliothèques numériques. Nous les avons réparties en quatre grandes préoccupations : l'environnement (qui ?), le contenu (quoi ?), la réalisation (comment ?), la finalité (pourquoi ?). Afin de mieux remplir la grille, nous avons tenu préalablement à recenser et à expliciter les différentes questions, à l'instar d'un mode d'emploi.

2.1.1.1 Qui ?

⁸ CRAMPES Jean-Bernard, *Interfaces graphiques ergonomiques (conception et modélisation)*, Paris, Ellipses, [Technosup], 1997

⁹ LAFOUGE Thierry, «Evaluation des systèmes d'information documentaire », Université de Lyon 1– Projet Recodoc

2.1.1.1.1 La source

- *Qui est à l'origine de la bibliothèque numérique ?*

Il est indispensable de pouvoir identifier la source responsable d'une bibliothèque numérique, voire de chaque document présent dans l'application. L'affichage du nom et des références de l'auteur de l'information, voire même l'existence d'un logo, sont parmi les gages de crédit.

- *Quelles sont les compétences de la source ?*

L'application peut contenir des renseignements concernant les champs d'expertise de la source. Cela doit permettre aux utilisateurs de pouvoir facilement vérifier si les documents présents dans la bibliothèque numérique se situent bien dans les domaines de compétences de la source. Il faut pouvoir vérifier l'adéquation entre une bibliothèque numérique et les compétences de son (ses) auteur(s), car elle témoigne du potentiel de crédibilité du dispositif technique.

- *Quels sont les moyens de la source ?*

Il s'agit d'informations souvent difficiles à obtenir. Quant nous y parvenons, elles sont souvent révélatrices. Par exemple, les moyens financiers de la source sont gages d'indépendance. Plus ils sont importants, plus ils assurent à la source une certaine liberté, une marge de manœuvre non-négligeable. Mais il est également intéressant de connaître les moyens humains (nombre de personnes travaillant à l'alimentation ou à la gestion quotidienne de la bibliothèque numérique). Evidemment, plus la structure professionnelle est importante (ex : la BNF), plus les moyens humains peuvent être importants, et par conséquent plus le développement de la bibliothèque numérique se fera rapidement. A contrario, des structures moins importantes ne peuvent charger qu'un petit effectif (ex : au CND, 5 personnes maximum) du développement de la bibliothèque numérique.

2.1.1.1.2 L'environnement

- *Qui collabore avec la source ?*

Mais il y a aussi les moyens associés à la source, c'est-à-dire toute contribution extérieure telle que collaborations techniques, associations financières, références intellectuelles... A noter qu'il peut se révéler parfois difficile d'évaluer les références intellectuelles d'une source, même lorsqu'elles sont correctement révélées. Pourtant la connaissance des références

peut être significative, par exemple d'une orientation culturelle (idées, théories...). La question des collaborations techniques renvoie notamment à l'existence de prestataires informatiques, mais aussi de comités d'évaluation (comité scientifique, comité éditorial...). Quant aux associations financières, elles posent le problème potentiel d'un conflit d'intérêt à l'information fournie. S'il existe une source de financement public ou privé, celle-ci est-elle clairement mentionnée ? Quelles sont les conditions implicites d'une telle conjugaison d'intérêts ?

2.1.1.2 Quoi ?

2.1.1.2.1 La plate-forme informatique

- *Sur quelle architecture technique repose la bibliothèque numérique ?*

L'intérêt de connaître les différents supports informatiques choisis pour stocker, gérer, diffuser les documents de la bibliothèque numérique, est de constituer un véritable panel de solutions disponibles, notamment compatibles avec le langage XML. A l'égard de l'architecture, des questions surviennent également en ce qui concerne l'aisance de la navigation : la structure apparaît-elle logiquement aux utilisateurs ? Ceux-ci peuvent-ils aisément trouver les chemins pour avancer, revenir... ?

2.1.1.2.2 Le contenu documentaire

- *Quel(s) type(s) de documents contient la bibliothèque numérique ?*

Le typage des documents peut s'effectuer à plusieurs niveaux et impliquer des traitements différents selon que :

1/ s'agit-il de monographies, de périodiques, de vidéos, de sons ?..

2/ pour les documents littéraires, sont-ils numérisés en mode texte ? en mode image ? Par exemple, le mode texte permet d'effectuer des recherches fines à l'intérieur d'un contenu intellectuel, tandis que le mode image permet de traduire de la manière la plus fidèle possible l'ensemble des particularités de l'objet et de l'œuvre.

- *Quelle est la taille du fonds documentaire ?*

Le volume du fonds documentaire est parfois difficile à estimer. Pourtant, lorsque c'est possible, il donne des indications sur les besoins d'adaptabilité, d'extension du support technique.

2.1.1.3 Comment ?

2.1.1.3.1 Le fonctionnement de la BN (approche développeurs)

- *Comment l'information est-elle saisie ?*

Comment l'information est-elle organisée ? Y a-t-il une hiérarchie d'évidence pour l'information diffusée ? Autrement dit, l'interface reproduit-elle graphiquement, donc extérieurement, l'organisation interne logique de l'information ?

Les données sont-elles directement créées en XML ou traduites en XML ? En cas de création en XML, où les données sont-elles saisies (par exemple, dans un éditeur XML tel que XML-Spy ou XML Notepad) ? En cas de traduction en XML, d'où sont extraites les données (par exemple, depuis un logiciel documentaire) ?

- *Comment l'information est-elle traitée ?*

La bibliothèque numérique dispose t'elle de DTD ou de Schéma associés validant les documents XML ? Ou bien les documents XML sont-ils simplement bien formés, c'est-à-dire ni associés à une DTD ni à un Schéma ?

- *Comment l'information est-elle présentée ?*

La bibliothèque numérique dispose t'elle de feuilles de styles CSS ou XSL/XSLT assurant une, voire plusieurs présentations de l'information ? Quels sont les éléments visuels qui découlent de ce(s) choix de présentation de l'information, et concernant :

- le design général (charte graphique) ?
- la mise en page (partition de la surface en zones, hiérarchisation des zones) ?
- la typographie (tailles, polices et couleurs des caractères) ?
- l'iconographie (illustrations, icônes, symboles) ?

2.1.1.3.2 L'exploitation de la BN (approche utilisateurs)

- *Comment chercher et trouver l'information ?*

Quels modes de recherche documentaire la bibliothèque numérique intègre t'elle ? Avec quelles contraintes ? Dans quelles perspectives ? Logiquement la bibliothèque numérique dispose pour la recherche d'au moins un ou plusieurs catalogues, avec un nombre variable d'entrées (titre, auteur, date...). Mais au-delà du catalogue, dispose t'elle d'un moteur de recherche interne ? D'un index général ? D'une rubrique « quoi de neuf dans la bibliothèque » ? D'une page d'aide ? D'un plan du site ? Par ailleurs, la bibliothèque numérique a t'elle développé des modes de recherche qui lui sont spécifiques ?

Un moteur de recherche interne doit être capable de chercher le contenu précis par mot-clé ou en texte intégral le plus souvent, et doit récupérer seulement les éléments pertinents. Il doit être complet, simple, facile à employer et fournir une vision claire de son résultat. Le moteur de recherche interne doit fonctionner rapidement et efficacement (en minimisant à la fois le « silence » -informations pertinentes que le moteur a omis de présenter en résultats- et le « bruit » -informations non-pertinentes que le moteur a malgré tout retenu en résultats-). Un index général constitue un moyen simple d'avoir rapidement accès à l'information désirée. Une rubrique « quoi de neuf dans la bibliothèque » permet également de visualiser rapidement l'information. Elle est particulièrement adaptée pour les bibliothèques numériques riches sur le plan informationnel et souvent mises à jour. Une page d'aide est un moyen d'assister l'utilisateur dans sa démarche de recherche, en cas de difficulté, ou du moins pour l'éclairer davantage. Un plan du site constitue un dernier moyen visant à faciliter la recherche d'informations ; l'objectif étant de proposer une vue d'ensemble de la bibliothèque numérique.

2.1.1.4 Pourquoi ?

2.1.1.4.1 Les besoins

- *A quels besoins la bibliothèque numérique répond-elle ?*

Pourquoi telle ou telle bibliothèque numérique existe t'elle ? Quel est son intérêt ? Toute nouvelle application informatique est légitimée par un certain nombre de besoins, auxquels elle se doit de répondre. Selon les besoins exprimés, l'application peut par exemple informer, défendre des idées, faire vendre. Généralement une bibliothèque numérique a pour vocation première d'informer ; mais il est intéressant de voir si elle combine d'autres vocations.

2.1.1.4.2 Les publics

- *Pour quels publics la bibliothèque numérique existe t'elle ?*

Les publics d'une bibliothèque numérique sont forcément variés. Le niveau de connaissances abordées, le niveau scientifique des références bibliographiques, la qualité des liens hypertextes, l'utilisation d'un vocabulaire spécialisé ou non, l'existence d'aides à la lecture et à la compréhension, la présence ou l'absence d'illustrations, sont autant d'indicateurs pour déterminer les publics visés par l'application.

2.1.2 Cas d'applications

L'ensemble de ces questions, à propos de la bibliothèque numérique et de la technologie XML, est synthétisé dans une grille descriptive combinatoire (combinaison de connaissances théoriques et d'éléments visuels). Celle-ci vient s'appliquer à chaque exemple de bibliothèque numérique présenté dans la première partie.

2.1.2.1 Le cas des bibliothèques numériques patrimoniales**Grille appliquée à GALLICA¹⁰**

QUI ?	
La source	
Nom et références	Bibliothèque nationale de France (BNF) 
Compétences	Mise en perspective et mise en valeur du patrimoine littéraire français du Moyen-Age au 19 ^{ème} siècle
Moyens	<ul style="list-style-type: none"> - <i>mise en place du programme de numérisation (1992-1998)</i> : 9 millions d'euros - <i>maintenance du programme</i> : 750.000 euros par an, pour un coût moyen de 0,15 euro par page en mode image et de 1,5 euro par page en mode texte
L'environnement	
Collaborations	<ul style="list-style-type: none"> - <i>pour l'élaboration d'un corpus de documents en mode texte</i> : INALF (Institut national de la langue française, éditeur de la base Frantext), éditeur Acamédia, éditeur Bibliopolis, éditeur Honoré Champion - <i>pour la constitution d'un corpus significatif en sciences</i> : Bibliothèque de l'Université d'Orsay, Bibliothèque de l'Ecole Polytechnique, Bibliothèque du CHU Sainte-Anne (Bibliothèque H. Ey) - <i>pour l'enrichissement du corpus d'images fixes</i> : Bibliothèque du Musée de l'Homme, Bibliothèque centrale du Muséum national d'Histoire naturelle, Médiathèque du Patrimoine et de l'Architecture
QUOI ?	
La plate-forme informatique	
Architecture technique	<ul style="list-style-type: none"> - stockage des documents sur des juke-boxes de disques optiques - diffusion des documents via un serveur Web
Le contenu documentaire	
Types de documents	<ul style="list-style-type: none"> - documents numérisés en mode image - documents numérisés en mode texte - images fixes - sons <p><i>NB : tous les documents visibles sur Internet sont libres de droits ! Mais il existe aussi quelques documents numérisés sous droits (négociés avec le Syndicat de l'édition), uniquement visibles en Intranet dans la bibliothèque.</i></p>
Taille du fonds	<ul style="list-style-type: none"> - 50.000 volumes imprimés, dont 1250 numérisés en mode texte - 80.000 images regroupées dans 10.900 dossiers thématiques

¹⁰ Voir aussi 1.1.1.2.1

	- 4 heures de documents sonores
COMMENT ?	
Approche développeurs	
Saisie de l'information	2 modes possibles : <ul style="list-style-type: none"> - traduction de données en XML, notamment à partir de notices InterMarc (selon l'exemple de la Bibliographie nationale française) - création de données en XML natif (selon l'exemple des Signets)
Traitement de l'information	Contrôle par la DTD BiblioML
Présentation de l'information	Processus de transformation XSLT , pour produire des pages soit de façon statique (pages de feuilletage par exemple), soit de façon individualisée, au cas par cas, en réponse à une recherche du lecteur
Approche utilisateurs	
Modes de recherche documentaire	<ul style="list-style-type: none"> - <i>modes traditionnels</i> : catalogue auteurs/titres/sujets, index général de domaines disciplinaires, moteur de recherche plein texte sur les ouvrages en texte intégral, la totalité des tables des matières et des légendes des images fixes - <i>modes spécifiques</i> : tables chronologiques (ex : que se passait-il en 1802 ?), parcours thématiques (tous les parcours possibles dans la culture francophone) <p><i>NB : Tous les accès sont dotés de liens hypertextes permettant de rebondir d'une notion à l'autre, d'une image à l'autre !</i></p>
POURQUOI ?	
Besoins exprimés	<ul style="list-style-type: none"> - <i>résoudre les problèmes de traitement</i> : manque de structuration (absence d'homogénéité et d'exhaustivité), limitation fonctionnelle pour la recherche - <i>résoudre les problèmes de production</i> : faciliter la mise à jour et la validation des données, éviter les doublons
Publics visés	Lecteurs du monde entier : d'après les estimations, 45% des utilisateurs viendraient de France et le reste de l'étranger. De plus, 65% de l'ensemble des utilisateurs appartiennent au monde de l'éducation, de la recherche ou des métiers du livre.

Grille appliquée au GUIDE DE L'INTERNET CULTUREL¹¹

QUI ?	
La source	
Nom et références	Ministère de la Culture et de la Communication – Mission de la Recherche et de la Technologie (MRT) 
Compétences	Missions de service public : sélection ou constitution de ressources spécialisées pour le secteur culturel, traitement (numérisation) et diffusion du Patrimoine culturel français
Moyens	Subventions de la MRT
L'environnement	
Collaborations	Prestataire technique : société AJLSM ¹²
QUOI ?	
La plate-forme informatique	
Architecture technique	Architecture SDX¹³ (Système de Documentation en XML, version 2) , bâtie avec plusieurs composantes de base : <ul style="list-style-type: none"> - Infrastructure de publication de documents XML Cocoon - Serveur Web Apache - Moteur de servlets Tomcat - Machine virtuelle Java - SGBD (MySQL, Interbase, InstantDB ou Oracle)
Le contenu documentaire	
Types de documents	Ressources hypertextes -services et serveurs- françaises et étrangères, autour des thématiques suivantes : <ul style="list-style-type: none"> - histoire de l'art - langue française et littérature - livre et édition - musées - musique - patrimoine - photographie - presse - théâtre - architecture - archives - droits d'auteur
Taille du fonds	Plus de 5.400 ressources électroniques référencées
COMMENT ?	
Approche développeurs	
Saisie de l'information	SDX ne constitue pas un outil de saisie ou de production de documents.
Traitement de l'information	C'est une plate-forme de gestion et de diffusion, permettant la

¹¹ Voir aussi 1.1.1.2.2¹² AJLSM : société spécialisée dans la gestion de l'information électronique. Voir sur : <http://www.ajlsm.com>¹³ Version gratuite SDX 1.1 à télécharger sur : <http://sdx.culture.fr/sdx>

	consultation de documents XML existants.
Présentation de l'information	Différents écrans générés à l'aide de feuilles de style XSLT : <ul style="list-style-type: none"> - écran de résultats rubrique : notices présentées directement sous l'arborescence (niveau de rubriques) - écran de résultats mot-clé : notices affichées selon un tri de pertinence qui fera apparaître en premier les fiches contenant tous les mots demandés
Approche utilisateurs	
Modes de recherche documentaire	4 modes de recherche : <ul style="list-style-type: none"> - <i>recherche libre</i> : utilisation du moteur de recherche interne - <i>recherche avancée</i> : utilisation du moteur de recherche combinée avec critères de sélection tels que choix de la langue, choix du statut de l'éditeur... - <i>recherche thématique</i> : choix d'une rubrique et/ou d'une sous-rubrique - <i>lecture aléatoire</i> : choix d'un site de la semaine <i>NB : 1 aide à la recherche : conseils à l'utilisateur pour la saisie de la requête.</i>
POURQUOI ?	
Besoins exprimés	Rendre aisément accessible aux internautes une masse impressionnante d'informations, tout en les guidant dans leurs démarches particulières : par exemple, pour les internautes intéressés par les musées, la sélection des sites, guides, projets et corpus relatifs à la numérisation d'images d'œuvres d'art, peut être précieuse.
Publics visés	Internautes s'intéressant au patrimoine culturel.

Grille appliquée à ARTISTES ET ŒUVRES¹⁴

QUI ?	
La source	
Nom et références	Centre national de la Danse (CND) 
Compétences	<ul style="list-style-type: none"> - conservation et diffusion du patrimoine documentaire concernant la danse (« développement de la culture chorégraphique ») - encadrement de formations - organisation de rencontres (stages, colloques...) - résidence d'artistes
Moyens	Subventions du Ministère de la Culture et de la Communication – DMDTS (tutelle)
L'environnement	
Collaborations	Prestataire technique : InOvo ¹⁵
QUOI ?	
La plate-forme informatique	
Architecture technique	Editeur intégré
Le contenu documentaire	
Types de documents	<ul style="list-style-type: none"> - <i>documents produits par les compagnies</i>: dossiers de présentation, supports de communication... - <i>documents produits par les lieux de diffusion</i>: programmes, dossiers de presse... - <i>documents produits en interne</i>: rencontres avec les artistes, captations... - <i>documents produits par les critiques et les chercheurs</i>: ouvrages, thèses, rapports... - documents iconographiques
Taille du fonds	(en phase de production test)
COMMENT ?	
Approche développeurs	
Saisie de l'information	<p>Modélisation en XML à partir des données documentaires fournies par les responsables de l'information</p> <p>Structuration en XML organisée en 3 niveaux hiérarchiques : œuvre/instance/représentation</p> <p>Définition de 5 types de documents XML :</p> <ul style="list-style-type: none"> - Compagnie - Artiste - Œuvre - Lieu - Thème <p>Création de groupes de balises spécifiques, avec éléments et</p>

¹⁴ Voir aussi 1.1.1.2.3

¹⁵ *InOvo* : société spécialisée dans la conception et l'architecture multimédia, en particulier XML. Voir sur : <http://www.inovo.be>

	<p>attributs : ex : pour décrire toutes sortes d'événements, même très précisément :</p> <ul style="list-style-type: none"> - <i>éléments</i>: <Event>, <EventDate>, <EventPlace>, <EventTitle>, <EventDesc>... - <i>attributs</i> : <Event type="naissance">, <Event type="oeuvre">, <Event type="exposition">...
Traitement de l'information	<p>Création d'une DTD pour chaque type de documents XML. Chaque DTD est élaborée à partir des composants suivants :</p> <p><teiHeader> <bases> <coordinates> <persDesc> <eventDesc> <adminDesc> <text> <back></p> <p><i>NB : plusieurs DTD peuvent disposer d'un même composant :</i> <i>Ex : « coordinates » pour des adresses, éléments que l'on trouve aussi bien dans un document « compagnie » que dans un document « lieu » !</i></p>
Présentation de l'information	<p>Feuilles de styles XSL offrant deux modes de consultation :</p> <ul style="list-style-type: none"> - consultation active (l'utilisateur formule lui-même sa requête) - consultation passive (l'utilisateur se laisse guider) <p><i>NB : modes de présentation aussi variés que modes de recherche possibles (voir ci-dessous).</i></p>
Approche utilisateurs	
Modes de recherche documentaire	<p>Diversité des natures de requêtes possibles :</p> <ul style="list-style-type: none"> - recherche sur balises uniques ou successives - recherche sur éléments textuels - recherche sur la présence ou non de telle valeur - recherche sur tel type d'objet ou sur l'ensemble des objets - recherche sur tel attribut de telle ou telle balise, dans tel ordre...
POURQUOI ?	
Besoins exprimés	<ul style="list-style-type: none"> - offrir des modalités variées de consultation des mêmes données - offrir des possibilités de requêtes croisées - faire communiquer les fonds d'informations (ex : base de données contacts sous FileMakerPro, base de données bibliographiques sous ExLibris...) et les différents supports - permettre la plus grande personnalisation possible
Publics visés	<ul style="list-style-type: none"> - <i>publics externes</i>: professionnels de la danse, enseignants, étudiants, journalistes, responsables de structures culturelles, grand public - <i>publics internes</i>: professionnels des départements du CND, étudiants en formation, intervenants, artistes en résidence...

Grille appliquée à ARTICLE@INIST¹⁶

QUI ?	
La source	
Nom et références	Institut de l'Information Scientifique et Technique (INIST) 
Compétences	Collecte, analyse et diffusion des résultats de la recherche scientifique internationale
Moyens	(inconnus)
Collaborations	Collaborateur intellectuel : CNRS
L'environnement	
QUOI ?	
La plate-forme informatique	
Architecture technique	Serveur Web interactif (recherche + commande) avec des sous-ensembles correspondant chacun à une description documentaire classique
Le contenu documentaire	
Types de documents	Les documents référencés sont des revues et des monographies (ouvrages, thèses soutenues dans les universités françaises, rapports, congrès) appartenant nécessairement au fonds de l'INIST ! Les domaines d'application sont les suivants : <ul style="list-style-type: none"> - sciences sociales et humaines (23%) - biologie (19%) - sciences appliquées (17%) - chimie (6%) - physique (5%) - psychologie (3%) - sciences de l'information (2%)
Taille du fonds	<ul style="list-style-type: none"> - 25.000 titres de périodiques - 560.000 numéros de périodiques - 2,5 millions d'articles de périodiques au total - 120.000 monographies
COMMENT ?	
Approche développeurs	
Saisie de l'information	Développement en XML
Traitement de l'information	Plusieurs validations possibles suivant types de documents : revue, article, monographie...
Présentation de l'information	Plusieurs présentations possibles suivant types de documents : revue, article, monographie...
Approche utilisateurs	
Modes de recherche documentaire	Choix entre 2 modes de recherche : <ul style="list-style-type: none"> - mode expert (équations de recherche) - mode simplifié Choix entre 4 bases d'interrogation :

¹⁶ Voir aussi 1.1.1.3.1

	<ul style="list-style-type: none"> - base revue - base numéro de revue - base article - base monographie <p><i>NB : de nombreuses fonctionnalités pour faciliter ou améliorer la recherche : 1 aide à la recherche très complète, 1 iconographie explicite...</i></p>
POURQUOI ?	
Besoins exprimés	Concilier deux prérogatives : diffuser les travaux de recherche scientifique et technique et proposer une offre commerciale intéressante pour commander les documents
Publics visés	<ul style="list-style-type: none"> - entreprises clientes - chercheurs « ponctuels »

Grille appliquée à NAÏADE¹⁷

QUI ?	
La source	
Nom et références	Etablissement national des invalides de la Marine (ENIM) <ul style="list-style-type: none"> - établissement public à caractère administratif - direction d'administration centrale du Ministère de l'Équipement, des Transports et du Logement, en charge de la marine marchande 
Compétences	<ul style="list-style-type: none"> - <i>en tant que direction d'administration centrale</i>: l'ENIM élabore la doctrine, propose aux Pouvoirs publics et au Parlement les réformes législatives et réglementaires relatives à la sécurité sociale des gens de mer - <i>en tant qu'établissement public administratif</i>: l'ENIM gère le régime social des marins et fait partie des régimes spéciaux de sécurité sociale
Moyens	Subventions du Ministère de l'Équipement, des Transports et du Logement
L'environnement	
Collaborations	∅
QUOI ?	
La plate-forme informatique	
Architecture technique	Intranet juridique
Le contenu documentaire	
Types de documents	<ul style="list-style-type: none"> - textes juridiques (lois, circulaires...) - jurisprudence - doctrine (notes interprétatives, lettres...) - fiches techniques ou fiches de procédure ou fiches d'information <i>NB : Tous les documents sont numérisés en mode image !</i>
Taille du fonds	Archives très importantes (cf. création de l'ENIM dans sa forme actuelle en 1930)! Donc n'ont été retenus que les documents volumineux, ou devant faire l'objet d'échanges, ou devant présenter des garanties de pérennité (ex : les circulaires)
COMMENT ?	
Approche développeurs	
Saisie de l'information	Saisie des données dans l'éditeur XML-Spy¹⁸
Traitement de l'information	Création d'une DTD spécifique pour chaque type de document : ex : pour la DTD circulaire : <ul style="list-style-type: none"> - 1 racine <circulaire> avec 1 attribut « circid » - sous racine : 6 éléments : <validation>, <identif>,

¹⁷ Voir aussi 1.1.1.4.1

¹⁸ Version d'évaluation XML-Spy 4.4 Suite gratuite pendant 30 jours. Télécharger sur : <http://www.xmlspy.com>

	<p><entete>, <zonetitre>, <zoneindex>, <corps texte></p> <ul style="list-style-type: none"> - sous <validation> : 1 élément - sous <identif> : 4 éléments - sous <entete> : 2 éléments - sous <zonetitre> : 5 éléments - sous <zoneindex> : 7 éléments - sous <corpstexte> : 1 attribut « sommaire » - sous « sommaire » : 4 éléments - etc. <p>Contrôle de la DTD et validation des documents dans XML-Spy</p>
Présentation de l'information	Écriture de feuilles de styles XSL
Approche utilisateurs	
Modes de recherche documentaire	(en cours de définition)
POURQUOI ?	
Besoins exprimés	<ul style="list-style-type: none"> - <i>résoudre les problèmes de traitement</i>: absence de normalisation, manque d'homogénéité - <i>traiter les cas particuliers propres à certains documents</i>: respect des éléments significatifs tels que noms propres ou couleurs associées aux documents
Publics visés	Publics internes : professionnels appartenant à la structure ou à la tutelle, voire personnes relevant de l'ENIM (marins professionnels français)

Grille appliquée à LEGIFRANCE¹⁹

QUI?	
La source	
Nom et références	Cellule du gouvernement français 
Compétences	Service public et gratuit de diffusion des données juridiques produites par l'Etat
Moyens	∅
L'environnement	
Collaborations	Informations juridiques provenant de différentes sources : <ul style="list-style-type: none"> - des Journaux officiels - du Secrétariat général du Gouvernement - du Conseil constitutionnel - de la Cour de cassation - du Conseil d'Etat - de la Cour des comptes - de l'Assemblée nationale - du Sénat Prestataire technique : ORT (titulaire de la concession de service public des bases de données juridiques)
QUOI?	
La plate-forme informatique	
Architecture technique	Serveur Web
Le contenu documentaire	
Types de documents	<ul style="list-style-type: none"> - journaux officiels - bulletins officiels des ministères - traités et accords - réglementation et jurisprudence - conventions collectives - textes de lois en vigueur - liens hypertextes vers des sites français et étrangers <i>NB. : documents nationaux, européens et internationaux !</i>
Taille du fonds	(indéterminé)
COMMENT ?	
Approche développeurs	
Saisie de l'information	Développement en XML
Traitement de l'information	Plusieurs validations possibles suivant types de documents : textes de lois, traités, conventions...
Présentation de l'information	Plusieurs présentations possibles suivant types de documents : textes de lois, traités, conventions...
Approche utilisateurs	
Modes de recherche documentaire	<ul style="list-style-type: none"> - recherche sur la majorité des textes juridiques : par mots du titre du document recherché ou par numéro du

¹⁹ Voir aussi 1.1.1.4.2

	<p>texte</p> <ul style="list-style-type: none"> - <i>recherche sur l'édition du Journal officiel « Lois et décrets »</i> : par date du JO, par mots du titre ou du texte (recherche avancée) du document recherché, par numéro du texte - <i>recherche sur les codes</i> : par la table des matières, par numéro d'article - <i>recherche sur les conventions collectives</i> : par mots-clefs, par titre de la convention, par numéro de la convention, par la table des matières
POURQUOI ?	
Besoins exprimés	Favoriser le développement de l'information juridique française
Publics visés	Internauts cherchant à résoudre une question de droit, qu'elle soit ponctuelle (ex : recherche d'un article de loi) ou plus longue (ex : constitution d'un dossier juridique)

2.1.2.2 Le cas des bibliothèques numériques universitaires**Grille appliquée à CODEX²⁰**

QUI ?	
La source	
Nom et références	Université de Paris X-Nanterre – Département INFOCOM 
Compétences	Formation et diffusion des savoirs
Moyens	(inconnus)
L'environnement	
Collaborations	Collaborateurs techniques
QUOI ?	
La plate-forme informatique	
Architecture technique	Serveur Web Apache couplé au moteur de servlets Jserv exécutant des servlets Java pour générer les « vues » (voir ci-dessous)
Le contenu documentaire	
Types de documents	Mémoires (maîtrise ou DEA) en Sciences de l'information et de la documentation
Taille du fonds	- environ 50 documents en juillet 2001 - croissance estimée à 20 documents par an
COMMENT ?	
Approche développeurs	
Saisie de l'information	Conversion en XML à partir de formats bureautiques natifs (ex : Word, WordPerfect)
Traitement de l'information	Validation du corpus en XML conformément à la DTD TEI Lite²¹
Présentation de l'information	Création d'un système spécifique de « vues » (ou modes de visualisation), donnant accès à des éléments des documents suivant besoins. A charge de l'utilisateur de choisir la « vue » dont il se sent le plus proche : ex : <ul style="list-style-type: none"> - <i>pour un aperçu synthétique</i>: visualisation du titre et du résumé d'auteur - <i>pour un aperçu global</i>: visualisation du titre, de l'introduction et de la conclusion - <i>pour un aperçu construit</i>: visualisation de la table des matières - <i>pour un aperçu notionnel</i>: visualisation du titre, du résumé d'auteur et de la bibliographie Ecriture de feuilles de styles XSLT pour permettre d'extraire des documents XML les fragments servant à composer les « vues ».
Approche utilisateurs	

²⁰ Voir aussi 1.1.2.2.1²¹ Voir TEI Consortium sur : <http://www.tei-c.org>

Modes de recherche documentaire	<ul style="list-style-type: none"> - index créés originellement par les auteurs des mémoires - métadonnées offrant des possibilités de recherche classiques : nom de l'auteur, titre du mémoire, directeur du mémoire, année de soutenance...
POURQUOI ?	
Besoins exprimés	Répondre à des besoins informationnels précis en développant un système tenant compte des différents types d'usages
Publics visés	<ul style="list-style-type: none"> - enseignants et étudiants - chercheurs en Sciences de l'information et de la documentation

Grille appliquée à PELLEAS²²

QUI ?	
La source	
Nom et références	Polytechnicum de Marne-la-Vallée 
Compétences	Edition des documents pédagogiques et des résultats de la recherche universitaire
Moyens	(inconnus)
L'environnement	
Collaborations	<ul style="list-style-type: none"> - prestataire technique (spécialiste de la numérisation de documents) : Jouve²³ - apports documentaires de la structure universitaire : secteur recherche, mais aussi secteur administratif et secteur pédagogique - apports financiers du Conseil régional d'Ile-de-France et des collectivités locales (pour REMUS)
QUOI ?	
La plate-forme informatique	
Architecture technique	<ul style="list-style-type: none"> - réseau de télécommunications à hauts débits : REMUS (REseau Marne-la-Vallée pour les Usages) - serveur de données Ex Libris (Cadic)
Le contenu documentaire	
Types de documents	<ul style="list-style-type: none"> - documents en texte intégral achetés auprès d'éditeurs commerciaux : revues et actes de congrès - littérature produite par les établissements universitaires et les centres de recherche : thèses, mémoires, rapports de stage, supports de cours, didacticiels, revues internes, documents pédagogiques, documents de recherche, actes administratifs, statistiques... - informations bibliographiques : banques de données bibliographiques et factuelles, signalements bibliographiques d'articles, revues de sommaires, listes d'acquisitions, DSI...
Taille du fonds	<ul style="list-style-type: none"> - plus de 300.000 documents - plus de 4000 périodiques
COMMENT ?	
Approche développeurs	
Saisie de l'information	Conversion des documents en XML à partir de formats textuels natifs (ex : RTF, LaTeX, Word). Chaque document est identifié et décrit à l'aide de

²² Voir aussi 1.1.2.2.2

²³ Jouve : société spécialisée dans le traitement numérique de l'information. Son offre se concentre sur la digitalisation de sources textuelles et d'images, l'ingénierie de systèmes de gestion de documents, la composition et l'impression d'ouvrages, la diffusion de l'information sur cédérom. Voir sur : <http://www.jouve.fr>

	métadonnées constituées à partir du modèle RDF
Traitement de l'information	Contrôle par rapport aux Schémas et DTD (TEI, ISO 12083²⁴, EAD²⁵, MathML, TCI...) reconnus suivant les types de document et validation des documents XML Ex : EAD (Encoded Archival Description) pour les documents du secteur de l'archivage, MathML pour les documents du domaine des mathématiques, TCI pour les documents du domaine des sciences humaines et sociales
Présentation de l'information	Ecriture de feuilles de styles XSL standard pour un ensemble de documents (ex : pour les mémoires de maîtrise, les DEA et les thèses) ou XSL personnalisée suivant le contexte
Approche utilisateurs	
Modes de recherche documentaire	(en cours de définition)
POURQUOI ?	
Besoins exprimés	<ul style="list-style-type: none"> - fédérer les ressources documentaires gérées localement par chaque établissement appartenant au consortium - mettre en place une filière de création et de gestion de documents multimédia ; d'où mise en place de la plate-forme de diffusion REMUS pour assurer des liaisons à haut débit entre les différents sites du consortium - garantir la cohérence du système mis à disposition des utilisateurs ; d'où enrichissement des fonctionnalités de l'Intranet et mise en place d'une administration centralisée des droits
Publics visés	<ul style="list-style-type: none"> - enseignants et étudiants - chercheurs

²⁴ Voir Norme ISO 12083 sur : <http://www.xmlxperts.com/12083xml.htm>

²⁵ Voir EAD sur : <http://lcweb.loc.gov/ead>

2.1.2.3 Le cas des bibliothèques numériques de presse**Grille appliquée à LE POINT²⁶**

QUI ?	
La source	
Nom et références	LePoint.fr 
Compétences	Site Internet de l'hebdomadaire d'actualités <u>Le Point</u> , offrant, outre la version électronique du journal, les services suivants : <ul style="list-style-type: none"> - cours de la bourse - offres d'emplois - programmes de télévision - prévisions météorologiques - itinéraires et plans
Moyens	(inconnus)
L'environnement	
Collaborations	Prestataire technique et hébergeur : Rosebud Technologies ²⁷
QUOI ?	
La plate-forme informatique	
Architecture technique	Plate-forme de gestion de contenu Ratio.Media : développée en Java, elle tourne sur un serveur Web Apache couplé à un serveur d'applications Tomcat et s'articule autour d'une couche de modélisation de données décrites avec XML
Le contenu documentaire	
Types de documents	Articles de presse numérisés sur les sujets suivants : <ul style="list-style-type: none"> - politiques française et internationale - sciences - économie - management - médias - culture
Taille du fonds	(indéfini)
COMMENT ?	
Approche développeurs	
Saisie de l'information	Traduction des articles au format Quark X'press dans un vocabulaire XML
Traitement de l'information	Utilisation d'une DTD
Présentation de l'information	Couche de modélisation de données décrites avec XML, intégrant des marqueurs pour « portaliser » la publication des articles. Cela permet de créer des visuels en fonction de thématiques pré-déterminées par les utilisateurs (cinéma,

²⁶ Voir aussi 1.1.3.2.1

²⁷ *Rosebud Technologies* : société spécialisée dans la gestion de contenu, principalement pour l'édition et les médias, en particulier la presse. Son expertise repose sur 2 technologies propriétaires que sont eGate et Ratio.Media. Voir sur : <http://www.rosebud.net/fr>

	littérature...), mais aussi en fonction du statut des articles (payant, gratuit), en fonction des rapprochements possibles avec des sites tiers (liens vers sites de télévision...)
Approche utilisateurs	
Modes de recherche documentaire	<ul style="list-style-type: none"> - 2 possibilités pour une recherche dans les articles de l'hebdomadaire : interrogation du moteur de recherche interne (bandeau supérieur) ou sélection d'une rubrique thématique (frame de gauche) - 1 possibilité pour un accès aux services annexes : choix d'un type de service (bandeau supérieur)
POURQUOI ?	
Besoins exprimés	<ul style="list-style-type: none"> - faciliter la réutilisation des données Web, leur récupération pour une application de cédéroms, leur transmission dans le cadre de contrats de production de contenu délégué - intégrer des marqueurs pour « portaliser » la publication des articles en fonction de thématiques pré-déterminées (cinéma, littérature...)
Publics visés	Lecteurs de la presse généraliste

Grille appliquée à LES ECHOS²⁸

QUI ?	
La source	
Nom et références	LesEchos.fr 
Compétences	Site Internet du quotidien économique <u>Les Echos</u> , offrant, outre la version électronique du journal, les services suivants : <ul style="list-style-type: none"> - information en direct - cours de la bourse - offres d'emplois et de stages - fiches techniques (création d'entreprises...) - bilans d'entreprises - suivi de finances et de déclarations sociales
Moyens	(inconnus)
L'environnement	
Collaborations	∅
QUOI ?	
La plate-forme informatique	
Architecture technique	Outil d'édition développé par l'équipe Internet du groupe pour assurer les flux d'informations à partir du progiciel Datox
Le contenu documentaire	
Types de documents	Articles de presse numérisés sur les sujets suivants : <ul style="list-style-type: none"> - politiques française et internationale - industrie - finances - management - consommation - distribution - services
Taille du fonds	(indéfini)
COMMENT ?	
Approche développeurs	
Saisie de l'information	Structuration et indexation de l'information en XML
Traitement de l'information	Utilisation d'une DTD
Présentation de l'information	Flux XML intelligents permettant d'adresser des informations spécifiques suivant les besoins des lecteurs clients
Approche utilisateurs	
Modes de recherche documentaire	Barre de navigation avec 5 onglets principaux : <ul style="list-style-type: none"> - « <i>Les Echos</i> » : recherche de l'édition complète du journal - « <i>Info en direct</i> » : recherche sur l'information en continu assurée par la rédaction Web

²⁸ Voir aussi 1.1.3.2.2

	<ul style="list-style-type: none"> - « <i>Finances personnelles</i> » : recherche dans la partie intégrée de l'ancien service MesFinances.fr - « <i>Bourse</i> » : recherche des cours boursiers - « <i>Entreprises</i> » : recherche sur les noms d'entreprises figurant dans l'actualité <p><i>NB : un système d'icônes de couleur a été mis en place pour signaler les accès payants ou gratuits</i></p>
POURQUOI ?	
Besoins exprimés	<ul style="list-style-type: none"> - développer du « sur-mesure » pour le compte de clients spécifiques ou d'un segment de clientèle - augmenter le nombre de visites en proposant une information à valeur ajoutée, donc en améliorant en priorité le taux de pénétration sur les cadres (cible privilégiée !) - parvenir à un équilibre entre 3 sources de revenus : publicité, abonnement en ligne et vente de contenu
Publics visés	Lecteurs de la presse économique et financière

2.2 Observations générales

Qu'apporte concrètement le langage XML aux différentes bibliothèques numériques par rapport à une autre technologie ? Quels standards s'associe-t-il et dans quelles perspectives ? En quoi le XML et les standards associés rendent-ils chaque projet spécifique, original ? Toutes les informations en caractère gras dans la grille, sont des éléments de réponse à ces questions. Voici ce qui en découle :

GALLICA : le développement de cette bibliothèque numérique s'inscrit dans une dynamique de découverte progressive du langage, sous toutes ses formes (XML natif, traduction en XML), avec toutes ses possibilités (transformations XSLT), et à partir des exemples de la Bibliographie nationale française et des Signets.

GUIDE DE L'INTERNET CULTUREL : la principale originalité de l'Internet culturel repose sur l'utilisation d'une solution informatique d'un genre particulier : SDX, plate-forme modulaire pour la gestion et la diffusion de documents XML, mais également outil entièrement composé de logiciels et de programmes libres, et de fait libre de droits. SDX a pu ainsi être aisément choisi comme support commun à un ensemble d'applications oeuvrant dans le même domaine, en l'occurrence la diffusion du patrimoine culturel.

ARTISTES ET ŒUVRES : la démarche de création en XML du CND est particulière ; en effet, elle coïncide avec la constitution d'une Médiathèque de la Danse, première structure physique de la sorte. La bibliothèque numérique est donc une vitrine de cette structure physique. Elle doit donc reposer sur une architecture infaillible (hiérarchie à partir de seulement 3 racines : œuvre/instance/représentation) et disposer d'une richesse rendue précieuse par sa présentation (multiplication des modes de recherche et de visualisation).

ARTICLE@INIST : l'INIST dispose d'un fonds documentaire pluridisciplinaire, auquel il convient pourtant de trouver une cohérence, au-delà des contenus et même des types de documents, pour la diffusion électronique. L'apport du XML peut être un outil de cette nécessaire cohérence.

NAÏADE : l'offre d'éditeurs XML est forcément moins variée que celle des éditeurs HTML (Flash, Dreamweaver, Frontpage...). Pourtant, l'ENIM figure parmi les premiers organismes à avoir choisi de développer son fonds documentaire en XML à partir de l'éditeur XML-Spy. A la différence de XML-Notepad, autre éditeur XML plus simple d'utilisation, XML-Spy est l'un des premiers à permettre de lier des documents XML à des DTD. L'intérêt pour l'ENIM est d'élaborer plusieurs DTD suivant le type de document juridique.

LEGIFRANCE : l'information juridique française concerne des publics très variés, qu'il est parfois relativement difficile à recenser. S'il y a peut-être autant de besoins que d'articles de lois, il faut dès lors une solution informatique comme XML, apte à la multiplication des présentations de l'information suivant lesdits besoins.

CODEX : la transformation des documents XML par des feuilles de styles XSLT permet d'élaborer un système de consultation personnalisée de l'information, selon différents profils d'utilisateurs. Le système comprend autant de modes de visualisation, ou « vues », que de profils d'utilisateurs. L'idée est de permettre aux utilisateurs d'obtenir une réponse pertinente et exhaustive en fonction de leurs besoins informationnels.

PELLEAS : le projet XML, au confluent des secteurs de l'édition, de la documentation et de la recherche universitaire, fédère de nombreux besoins. Entre autres, le balisage XML doit correspondre aux normes utilisées dans le monde de l'édition et de la documentation comme la DTD ISO 12083 et la TEI (Text Encoding Initiative) utilisée dans le milieu universitaire en lettres et en sciences humaines.

LE POINT : pour la presse périodique, dont le traitement doit être rapide, le développement en XML requiert la solution la plus fonctionnelle, même si elle n'est pas forcément la plus « créative ». Le Point a ainsi choisi de traduire les données dans un format spécifique, Quark X'press, avant de générer des flux XML.

LES ECHOS : l'idée du « sur-mesure » vaut aussi pour la diffusion de la presse périodique. Grâce à XML, Les Echos proposent une offre individualisée d'information économique et financière suivant les besoins de clients, du moins de segments de clientèle.

Mais l'application de la grille d'analyse suscite nombre de nouvelles questions, dont les éléments de réponse devront cette fois nous permettre de déterminer un contexte privilégié de développement d'une bibliothèque numérique à l'aide du XML. Ainsi, y a-t-il un secteur professionnel qui se démarque des autres, dans le développement de bibliothèques numériques avec XML et les standards associés ? Y a-t-il un type de documentation qui est propice au choix du XML et des standards associés ? En somme, quels éléments constituent soit des forces dynamiques, soit des terrains propices à l'intégration du XML pour le développement de bibliothèques numériques ? S'il existe des prédispositions pour l'intégration du XML, faut-il y voir des indices de pertinence valables pour choisir de développer une bibliothèque numérique à l'aide de ce langage ? Dès lors, il conviendra de réfléchir aux contraintes de définition d'indices de pertinence justifiant le choix du XML et des standards associés dans les bibliothèques numériques.

2.2.1 Le cadre logique : quelles forces dynamiques ?

Il s'agit d'un ensemble de concepts liés les uns aux autres et qui doivent être utilisés concurremment de façon dynamique pour permettre l'élaboration d'un dispositif technique bien conçu, décrit en termes objectifs et dont on pourra ultérieurement déterminer l'intérêt. Le cadre logique permet de résumer les éléments d'un dispositif technique (moyens, besoins, objectifs), ainsi que les liens de causalité entre les aspects opérationnels (ressources, activités, résultats) et les considérations de développement (buts, finalités).

2.2.1.1 L'activité du secteur professionnel

Le secteur professionnel est un premier concept constitutif du cadre logique. Par exemple, le secteur culturel figure actuellement parmi les plus dynamiques dans l'intégration de la technologie XML. Il est en effet particulièrement riche de bibliothèques numériques développées avec XML et les standards associés. Y a-t-il quelques raisons particulières à cette dynamique du secteur culturel ?

Il faut savoir que la mise en ligne de documents numériques, issus du patrimoine culturel, ne relève pas d'une quelconque politique d'image ou de communication, comme ce peut être le cas pour d'autres types de documents (notamment pour la presse électronique). Elle relève au contraire d'un objectif fondamental : rendre publique les connaissances véhiculées par le patrimoine culturel. Ce qui est en jeu, ce n'est pas le caractère spectaculaire de ce qui est présenté, mais bien la valeur informative et intellectuelle du contenu. De plus, l'offre doit être variée et abondante, tant il est vrai que les publics de la culture sont sans doute les plus larges, la culture renvoyant tout autant au domaine de l'archéologie et à celui du théâtre par exemple. Or, on distingue encore mal des dénominateurs communs entre ces mêmes pendants de la culture. En terme de valeur informative, le choix du XML et des standards associés s'impose parce qu'ils permettent non seulement de préserver l'information brute des appareils de présentation (séparation du fond et de la forme, traités respectivement dans le document XML et dans des feuilles de styles CSS ou XSL), mais aussi d'accompagner l'information primaire, d'éléments d'information secondaire (métadonnées : caractéristiques retranscrites dans les balises). En terme de variété de l'offre publique, l'association du XML et des feuilles de

styles XSL permet de diversifier les présentations, suivant des profils d'utilisateurs, et ainsi de s'adapter aux besoins des différents publics.

2.2.1.2 La complémentarité des métiers

Il est intéressant de voir comment la constitution d'une bibliothèque numérique scelle la rencontre de deux modèles : la bibliothèque, modèle ancestral dont on a retrouvé des traces jusqu'au III^{ème} millénaire avant Jésus-Christ (Mésopotamie)²⁹, et l'ordinateur -ou l'informatique-, modèle né seulement au siècle dernier (le premier ordinateur, l'ENIAC, date des années 1940). Rien ne prédispose donc historiquement au rapprochement de ces deux modèles. Pourtant, ils ont en commun le souci fondamental d'organisation et de communication des savoirs. La rencontre actuelle de ces deux modèles est plus exactement celle de métiers qui travaillent à l'organisation et à la communication des savoirs, en particulier pour une bibliothèque numérique. La force d'une bibliothèque numérique tient entre autres à cette rencontre entre des professionnels de l'image et du discours (scénarisation, ergonomie d'écran, graphisme...), en particulier des informaticiens, et des professionnels de l'information-documentation (conception de bases de données, gestion documentaire...), en particulier des bibliothécaires ou des documentalistes. Mais la force dynamique d'une bibliothèque numérique intégrant XML est plus particulièrement cette capacité à rendre ces métiers complémentaires. Il est vrai que le XML est avant tout un langage informatique destiné de fait aux informaticiens. Mais le développement en XML entretient aussi de nombreuses correspondances avec les logiques documentaires : structuration des données, normalisation des pratiques, personnalisation de l'information... C'est pourquoi il semble désormais intéressant pour les professionnels de l'information-documentation de pratiquer le langage XML. Ce dernier fédère ainsi deux conceptions professionnelles qui gagnent à coopérer ; l'association d'informaticiens et de bibliothécaires ou de documentalistes est donc un élément dynamique pour le développement d'une bibliothèque numérique à l'aide du XML. Il est intéressant de voir que la réalisation des bibliothèques numériques préalablement soumises à la grille d'analyse, n'a donc pas cloisonné les pratiques professionnelles, mais a au contraire permis un véritable travail collaboratif entre des informaticiens et des documentalistes autour du XML, et y compris pour l'écriture des DTD ou des feuilles de styles.

²⁹ PALLIER Denis, *Les bibliothèques*, Paris, PUF, 1997

2.2.2 Le contenu intellectuel : quels terrains propices ?

2.2.2.1 Des documents « pré-structurés »

L'utilisation du langage XML permet d'enregistrer mais surtout d'encoder des documents d'une façon riche, même si compliquée. Cependant, certaines conditions peuvent permettre de faciliter la démarche, du moins de lui « préparer le terrain ». Par exemple, la structure originelle des documents peut être un terrain propice pour l'utilisation du XML. Certains documents, comme les travaux de recherche (mémoires de fin d'études, thèses, compte rendus, articles scientifiques...), figurent parmi les documents numériques les plus « faciles » à adapter à l'utilisation du XML et des standards associés. Ce sont des documents « pré-structurés », c'est-à-dire qui présentent a priori une structure claire, précise, à l'image de celle imposée pour tout document XML !. Une fois numérisés, il suffit alors de reprendre les éléments organisationnels de ces documents et de leur appliquer quelques règles de transcription pour les rendre utilisables à la fois pour la recherche documentaire, la publication sur le Web ou même la gestion administrative.

2.2.2.2 Des documents « complexes »

D'après ce que nous savons désormais des exemples d'intégration du XML dans les bibliothèques numériques, il semble que le langage soit particulièrement recommandé pour une documentation :

- volumineuse, car le développement en XML étant lourd, il est plus logique de l'entreprendre pour un fonds important !
- évolutive, étant donné que la séparation du fond et de la forme du document XML permet une mise à jour aisée et régulière de l'information !

De plus, nous savons maintenant que le choix du XML peut se révéler approprié dans les cas où les documents font l'objet d'importants échanges, comme pour la presse périodique, ou lorsque les documents présentent des garanties de pérennité, comme pour les ressources patrimoniales.

En réalité, forces dynamiques et terrains propices ne sont que des prédispositions pour l'application du XML et des standards associés aux bibliothèques numériques. Il est vrai qu'un cadre professionnel particulièrement dynamique, ou une nature documentaire particulièrement adaptable sont des avantages pour le développement en XML. Pour autant, il est impossible de considérer ces avantages comme des indices de pertinence, valant pour le choix du XML et des standards associés, car ils sont d'abord aléatoires. Comment alors établir un calcul de pertinence du XML ? Quelles sont les constantes et les variables qui entrent en jeu dans ce calcul et donc dans la définition d'indices de pertinence ? Enfin, quels intérêts présente cette démarche ?

3 RESULTATS DE L'EVALUATION ET DEFINITION D'INDICES DE PERTINENCE DU XML

L'évaluation consiste à examiner un dispositif technique, tel que celui issu de l'association du XML et de la bibliothèque numérique, à en mesurer l'impact et la viabilité. Les notions d'efficacité, de performance, de valeur et de qualité qui se rapportent soit à l'outil informatique, soit au contenu informationnel, découlent de l'évaluation. Puis elles induisent la notion de pertinence. Tandis que les premières notions peuvent être appréciées d'après des connaissances spécifiques sur les caractéristiques du langage informatique ou sur l'organisation de l'information dans une bibliothèque numérique, la notion de pertinence ne peut être évaluée que d'après les pratiques des usagers de la bibliothèque numérique.

3.1 De l'efficacité du langage informatique à la performance du dispositif technique

En quoi le XML est-il efficace pour la bibliothèque numérique (BN)? Autrement dit, comment le XML « aide » (enrichit, améliore) la bibliothèque numérique? Quels besoins de la bibliothèque numérique le XML comble-t-il? Du moins, quels objectifs permet-il d'atteindre?

3.1.1 Le potentiel efficace du XML

C'est sans doute la donnée la plus aisément mesurable. Elle repose en effet sur la connaissance des qualités objectives du langage XML. En quoi est-il logiquement efficace, notamment pour une bibliothèque numérique, au point par exemple que nombreux sont ceux qui pensent y voir « l'espéranto du Web »³⁰ ou « la lingua franca du Net »³¹ ?.. Il y a en réalité deux grandes qualités objectives du langage XML, grâce auxquelles nous pouvons dire qu'il est efficace : c'est à la fois un langage normalisé et ouvert.

³⁰ BERTIN Pascal, « XML, l'espéranto des serveurs Web », *Le Monde interactif*, 16 février 2000

CEZAI R Paul, « XML ou le rêve du Net-espéranto », *L'Expansion*, 13 septembre 2001, n°652

ROUMIEUX Olivier, « XML, l'espéranto tant attendu », *Archimag*, décembre 1998-janvier 1999, n°120

³¹ MOLINSKI Michael, « Une 'lingua franca' pour les données du Net ? », *Le Monde interactif*, 9 novembre 2000

3.1.1.1 Un langage normalisé

En quoi le langage XML est-il normalisé ? Quel est l'intérêt d'une telle caractéristique, synonyme d'efficacité dans notre propos ?

La structuration d'un document XML répond à certaines règles, représentations concrètes de la notion de normalisation. Parmi les règles fondamentales de balisage, il y a que :

- toute balise ouverte (balise ouvrante `<Organisme>`) doit être obligatoirement fermée plus loin (balise fermante `</Organisme>`).
- toute balise définit des données dans un format organisé suivant le modèle `<Organisme>données</Organisme>`.
- les emboîtements de balises ne doivent pas se chevaucher ; un document XML comprend toujours un élément unique, une racine, qui renferme tous les autres éléments selon une arborescence stricte.
- un attribut lié à un élément doit toujours être présenté dans une balise incluante selon la syntaxe `<Contact type=« direction générale »>`, ou `<Contact type=« service de documentation »>`...

Donc, le balisage d'un document XML répond au souci de normalisation. Mais il en va de même pour le balisage des fichiers associés, tels que les DTD. Il est lui aussi normalisé, répondant à des règles syntaxiques d'imbrication des éléments, de limitation de valeurs (par exemple, une seule syntaxe possible pour signifier qu'un élément ne contient que des caractères : `#PCDATA`)...

L'intérêt majeur de la normalisation est de rendre un système homogène, donc facilement exploitable. C'est d'autant plus important pour une bibliothèque numérique présentant des documents de plusieurs types. Par exemple, l'un des principaux enjeux du projet «Naï ade » a été de proposer une bibliothèque cohérente et unifiée, malgré la très grande hétérogénéité de son contenu documentaire : textes juridiques, jurisprudence, doctrine, fiches techniques... C'est pourquoi l'ENIM a choisi de mettre en place autant de DTD qu'elle recensait de types de documents, normalisant ainsi la saisie des données directement auprès des rédacteurs de chaque type de document.

3.1.1.2 Un langage ouvert

En quoi le XML est-il un langage ouvert ? Quel est l'intérêt d'une telle qualité, gage d'efficacité dans notre propos ?

Même si ce n'est pas forcément évident a priori, les termes « normalisé » et « ouvert » ne sont pas des qualificatifs contradictoires. Donc, même si la pratique du XML impose certaines règles strictes de balisage, elle présente aussi des qualités d'ouverture. Le langage XML autorise en effet la création de ses propres balises, indépendamment des contraintes de langue par exemple. Tandis que la plupart des langages informatiques sont écrits dans la langue de Shakespeare, le XML peut intégrer des balises en français (ou en toute autre langue). Cela va même plus loin, puisque le XML permet d'intégrer des lexiques propres à un métier, à un domaine..., et aussi la façon de les utiliser.

Par exemple, pour un fonds aussi spécifique que celui du Centre national de la Danse (Artistes et Œuvres), le XML a permis de baliser les documents non seulement en fonction de la nature de l'information structurée (œuvre, instance, représentation), mais aussi selon des usages précis, des types de publics et donc de requête ou de mode de consultation rendus ainsi possibles. Il est ainsi possible de modeler préalablement le balisage selon des interrogations, des usages évalués en amont. Anticiper les interrogations des utilisateurs potentiels à propos des artistes, des compagnies de danse, des œuvres chorégraphiques (par exemple, des questions aussi pointues que « est-ce que Dominique Dupuy a dansé des solo ? » ou « quel chorégraphe utilise l'eau comme thème dans ses œuvres ? ») permettent d'orienter le balisage des documents, de créer des balises adaptées à chaque situation et donc de répondre au mieux aux attentes.

3.1.2 Le principe de performance

Nous pourrions donner une représentation de la mesure de la performance, qui se caractériserait par le rapport suivant :

$$\text{Schéma de la performance optimale} = \frac{\text{gains réels de la BN (satisfaction offerte)}}{\text{gains escomptés de la BN (satisfaction souhaitée)}} = 1$$

Le rapport (ou la confrontation) entre les gains réels de la bibliothèque numérique suite à l'application du XML, et les gains escomptés de la BN compte tenu du potentiel du XML, permet de mesurer la performance du dispositif technique né de l'association de la BN et du

XML. Si l'observation des résultats concrets d'un projet suffit à déterminer les gains réels, où trouver trace des gains escomptés ? Ils sont notamment développés dans des cahiers des charges, documents énumérant les exigences, les clauses d'un projet, mais aussi les conditions imposées à des fournisseurs pour l'exécution d'un marché ou d'un contrat.

Par exemple, le cahier des charges pour le projet Pelleas de l'Université de Marne-la-Vallée, recommandait en 1997 l'utilisation de SGML, parce que jugé le plus apte à gérer des documentations lourdes. S'ensuit une période de deux ans, où XML apparaît, s'impose comme le nouveau choix pour Pelleas, se substituant ainsi à SGML. En effet, après un travail de vérification des pré-requis techniques de la première proposition, puis de validation des objectifs cadres, SGML est apparu trop compliqué mais surtout trop cher. Le choix du XML est alors justifié de deux manières :

- XML est un standard gratuit, donc sans droit de licence. Cela signifie que n'importe qui peut utiliser ce format de fichier, le comprendre et le lire gratuitement. Pour Pelleas, qui, parmi toutes les bibliothèques numériques que nous avons mentionnées jusqu'ici, est celle qui a le plus « d'auteurs » (appartenant au consortium), il importait vraiment de trouver un dénominateur commun accessible à tous.
- Le nombre d'auteurs fait aussi le volume et la complexité du fonds documentaire. Il convenait donc de trouver un langage suffisamment normalisé pour gérer le volume documentaire, et suffisamment ouvert pour s'adapter à la complexité documentaire.

3.2 De la valeur de l'information à la qualité de la documentation

3.2.1 La valeur informative

Qu'est-ce qui fait la valeur d'une information diffusée numériquement ? Trois critères principaux se distinguent :

- le fait qu'elle soit disponible ;
- le fait qu'elle soit fiable ;
- le fait qu'elle soit adaptable.

Comment l'apport du XML aux bibliothèques numériques peut influencer sur la réalité de les trois critères faisant la valeur de l'information ?

3.2.1.1 Disponibilité de l'information

L'information vaut pour être disponible, c'est-à-dire capable d'être retrouvée par les utilisateurs intéressés, qui pourront ensuite en disposer dans un contexte donné et en fonction de leurs propres préoccupations (suivant le contenu sémantique, l'origine...). La disponibilité de l'information se mesure concrètement grâce au critère d'accessibilité du document. Ce qui nous amène à l'interrogation suivante : comment le XML peut améliorer l'accessibilité au document, et par conséquent influencer sur la disponibilité de l'information contenue dans une bibliothèque numérique ? Notre « démonstration » est un peu comme la résolution d'un problème mathématique, dont l'énoncé serait : « sachant que l'accessibilité au document est fonction de l'identifiabilité, de la facilité d'exploitation et de la confidentialité du document, comment démontrer que le XML est un facteur d'accessibilité du document ? »

Le sous-critère d'identifiabilité suppose qu'un document stocké doit être capable de fournir des renseignements à propos de ses auteurs, de son titre, de sa date de production ou encore du type d'informations qu'il contient (œuvre littéraire, article de presse, texte juridique...), enfin toutes caractéristiques traduisant une sémantique vis-à-vis de l'utilisateur. Or, nous savons déjà que la technologie XML simplifie l'identification du document, de deux manières. La première est que les balises créées librement selon les besoins peuvent parfaitement contenir les informations requises normalement pour le catalogage et l'indexation du document, pratiques d'identification, de caractérisation du document : nom de l'auteur, titre du document, date d'édition, etc. (pour une description contextuelle), et mots-clés ou descripteurs, vedettes matières, etc. (pour une description thématique). La technologie XML dispose aussi d'un second moyen améliorant l'identifiabilité du document, grâce à la constitution de DTD qui fournissent des renseignements suivant le type d'informations contenues dans le document XML. Ces informations signalétiques peuvent être appréhendées indépendamment du document XML. Ces deux manières font que l'information signalétique identifiant le document XML peut être accessible indépendamment de l'information brute contenue dans le document XML. Et l'intégrité de l'information brute et celle de l'information signalétique (indépendante), s'en trouvent dès lors préservées.

Le sous-critère de facilité d'exploitation renvoie à l'aptitude du document à présenter tout ou partie de son contenu dans un format que l'utilisateur peut manipuler directement. La facilité d'exploitation est obtenue dès lors que l'utilisateur peut se servir de n'importe quel outil (et par conséquent celui qu'il préfère !) pour appréhender le document. Il est donc particulièrement important que le document soit stocké sous une forme normalisée qui sera acceptée par les différents outils. Or un document XML peut être visualisé sur des machines différentes, grâce à l'application de feuilles de styles qui filtrent, réarrangent, affichent l'information ; par exemple, elles peuvent engendrer des documents comprenant des graphiques pour les ordinateurs de bureau, des fichiers composés exclusivement de texte pour les ordinateurs de poche, voire même des messages pour les téléphones portables.

Enfin, le sous-critère de confidentialité du document fait écho à l'aptitude du document à ne permettre son exploitation qu'aux personnes autorisées, et dans les limites des droits qu'elles possèdent. Cette capacité est essentielle pour la presse électronique, pour laquelle la question des droits est particulièrement délicate. Or, lors de la constitution de feuilles de styles XSLT associées aux documents XML, il est possible d'intégrer des niveaux de confidentialité liés à des groupes d'utilisateurs identifiés selon leurs statuts : lecture (statut de « l'utilisateur-lecteur »), copie (statut de « l'utilisateur-client ») ou modification (statut de « l'utilisateur-administrateur »). Une telle pratique permet de contrôler l'accès aux informations véhiculées par les documents XML, suivant les droits dont disposent respectivement les auteurs, les éditeurs et les utilisateurs.

3.2.1.2 Fiabilité de l'information

L'information vaut également pour être fiable, autrement dit digne d'un intérêt et d'une confiance justifiés selon les utilisateurs. Il est en effet primordial qu'une information puisse être prouvée ou qu'un utilisateur manipulant l'information puisse justifier sa validité. Une information est dite fiable quand, s'appuyant sur des arguments, des références, elle est objective, honnête, contrôlée. La notoriété de l'auteur (la réputation de ses compétences) ou l'expérience de l'émetteur (le sérieux de son mode de production) peuvent être garantes de la fiabilité de l'information. Un document doit donc être en mesure de fournir des renseignements permettant d'évaluer la qualité de ses rédacteurs, les objectifs de sa création,

l'intérêt des procédures suivies lors de sa production... Il est impossible de déterminer directement ces données ; elles découlent forcément d'un processus d'analyse.

En quoi le XML peut-il aider à la mise en place de ce processus ? Nous pourrions dire que le langage offre aux utilisateurs des moyens non seulement pour comprendre le contenu, en faisant appel à des règles de structuration normalisée et logique, mais aussi pour l'analyser, en mettant en place des outils de contrôle et de validation du travail (DTD, Schéma).

3.2.1.3 Adaptabilité de l'information

Enfin, l'information vaut pour être adaptable, donc apte à être présentée sous des formes variables selon les besoins des utilisateurs. Pour des raisons évidentes de rentabilité du système auquel l'information appartient, celle-ci doit être intéressante à plusieurs titres ; et sa réutilisation (diversification des modes de présentation) doit demander moins d'effort que sa production. C'est sans doute le critère de valeur informative, pour lequel le XML offre le plus d'avantages. En effet, le XML est un langage modulaire, extensible (comme indiqué dans le titre !). Autrement dit, à partir de règles de base communes, il est possible d'adapter le XML selon des besoins documentaires spécifiques.

3.2.2 Le principe de qualité

Le principe de qualité de la documentation, en tant que conteneur de l'information, est par conséquent inhérent à la valeur de l'information. Comment alors décliner les critères de valeur informative en facteurs de qualité documentaire ?

Prenons la représentation mathématique d'une fonction : $y = f(x)$. Cela signifie que y est l'image de x par f . Si nous posons que y correspond à la qualité de la documentation (ce que nous cherchons à définir !), et que x correspond à la variable information (ce qui nous est déjà connu !), nous pourrions alors dire que « la qualité de la documentation est l'image de l'information, par f . » Or nous avons vu ci-avant que l'information vaut par sa disponibilité, de sa fiabilité et de son adaptabilité. Ces trois critères représentent f . Ce qui nous amène à la fonction suivante :

$$\text{Qualité de la documentation} = (\text{disponibilité} + \text{fiabilité} + \text{adaptabilité}) \text{ information}$$
$$y = f(x)$$

L'intégration de la technologie XML doit permettre de construire un système homogène de documents, tel qu'une bibliothèque numérique, dont l'apport aux utilisateurs est l'assurance de la qualité de la documentation. Cet enchaînement de compétences, à partir d'un langage efficace comme XML, peut aboutir à la notion de « document intelligent » (ou « document réactif »), capable de communiquer tout ou partie de sa sémantique et aussi de s'adapter à des profils d'utilisateurs. La capacité conceptuelle de la structuration en XML facilite, améliore, enrichit la recherche et donc la réutilisation de l'information. Cette capacité autorise en effet une interrogation sémantique et contextuelle, plus riche que la définition catalographique ou l'indexation classique. De cette façon, le système de documentation, en l'occurrence la bibliothèque numérique, peut être considéré comme un système interactif d'aide à la production de documents. L'aptitude à l'interactivité est un atout de la bibliothèque numérique. Ainsi, la possibilité de personnaliser le rapport de l'utilisateur à la bibliothèque numérique devrait toujours être incluse. L'utilisateur devrait aussi être en mesure de commenter la validité et la valeur de l'information présentée, et éventuellement faire remarquer les omissions ou les biais. Il est donc également important qu'une bibliothèque numérique fournisse aux utilisateurs, plusieurs moyens de communiquer notamment avec l'auteur du contenu ou avec l'administrateur de l'application, mais aussi avec d'autres utilisateurs. C'est pourquoi il convient plus que jamais de resituer l'évaluation du XML appliqué à la bibliothèque numérique en tant que système interactif, dans une dimension sociale. Il en va de la pertinence du XML et des standards associés...

3.3 Vers la pertinence du langage XML pour les bibliothèques numériques

La pertinence va au-delà de l'efficacité du dispositif technique, au-delà de la qualité de la documentation. Bien sûr la pertinence relève de la constitution d'un dispositif efficace et de la diffusion d'une documentation de qualité, mais elle répond à d'autres préoccupations. Dans notre cas, l'apport du XML et des standards associés dans la bibliothèque numérique est pertinent par rapport au domaine concerné (l'apport est-il utile ?), et relativement à sa présentation à l'utilisateur (l'apport est-il utilisable ?).

3.3.1 Le principe de pertinence

Notre exploration du concept de pertinence nous a amenés à étudier des théories communicationnelles, en particulier celles de Paul H. Grice³² et de Dan Sperber et Deirdre Wilson³³, avant de présenter une modélisation de la pertinence, sous la forme d'indices.

Pour Sperber et Wilson communiquer, c'est requérir l'attention d'autrui. Par conséquent, c'est laisser entendre à autrui que l'information que l'on cherche à lui transmettre est pertinente. Ils définissent ainsi un premier principe de pertinence : « une information communiquée est assortie d'une garantie de pertinence ». Sperber et Wilson montrent en quoi ce principe permet d'expliquer de quelle façon un énoncé et son contexte interagissent et déterminent la façon dont cet énoncé sera compris. Ils définissent le contexte comme l'ensemble des prémisses³⁴ utilisées pour l'interprétation d'un énoncé. Leur théorie de la pertinence se résume en deux principes fondamentaux :

- 1 : le principe cognitif de pertinence, selon lequel la cognition humaine tend à maximiser la pertinence des informations qu'elle traite
- 2 : le principe communicationnel de pertinence, selon lequel tout acte de communication communique une présomption de sa propre pertinence.

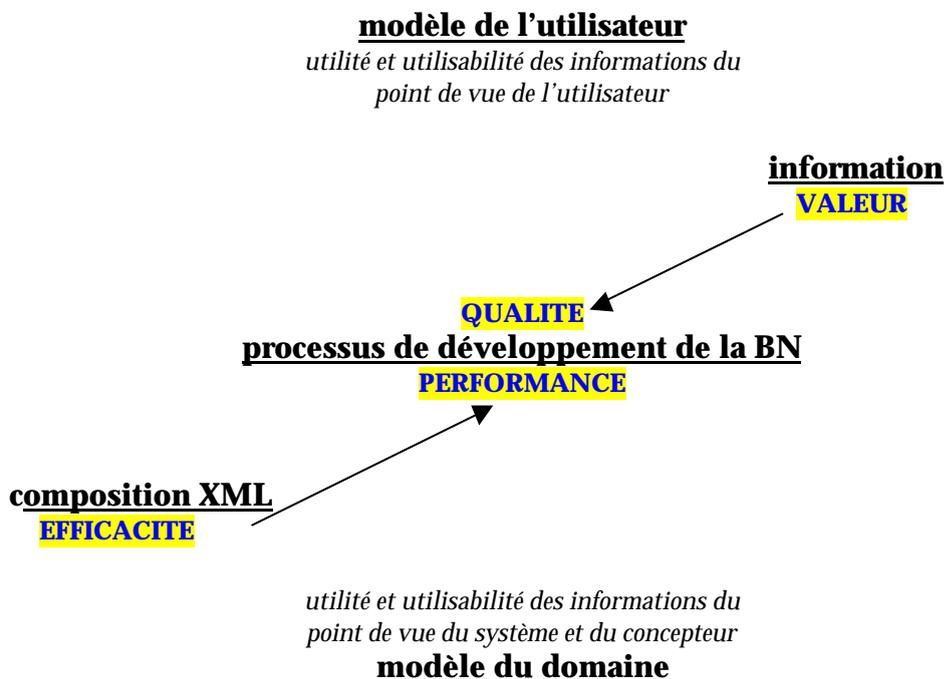
Le premier principe s'applique à toutes les adaptations mentales modulaires et à leur articulation, et découle par conséquent de considérations générales sur l'efficacité des systèmes cognitifs complexes. Le second principe suggère comment un module pragmatique exploitant la pertinence peut durer, s'enrichir, évoluer...

Or, quelle réalité ou représentation donner à ces principes dans le cadre précis de notre recherche ? Nous pourrions déjà dire que le contexte de développement des bibliothèques numériques a constitué le commencement de notre démonstration, les prémisses indispensables à l'interprétation de notre « énoncé », en l'occurrence l'apport du langage XML et des standards associés. La figure ci-dessous permet de mieux comprendre les rapports dont découlent les principes de cognition et de communication. La figure représente le processus de développement d'une bibliothèque numérique, en particulier les instances qui interviennent et bâtissent des stratégies d'évaluation de l'objet associé au langage XML, définissant ainsi des approches différentes mais complémentaires.

³² GRICE Paul H., *The conception of value*, Oxford (GB), Clarendon Press, 1991

³³ SPERBER Dan et WILSON Deirdre, *La pertinence : communication et cognition*, Paris, Editions de Minuit, 1989

³⁴ prémisses (≠ prémices) : affirmations dont on tire une conclusion ; commencement d'une démonstration



Selon ce schéma, une bibliothèque numérique naît de la concrétisation d'une pratique, la composition en XML, sur une instance, l'information. D'une part, l'efficacité de la composition XML fait la performance du processus de développement de la bibliothèque numérique. D'autre part, la valeur de l'information fait la qualité de ce même processus. De plus, le développement d'une bibliothèque numérique met en scène deux modèles : celui du domaine, émetteur de la bibliothèque numérique, et celui de l'utilisateur, récepteur de la bibliothèque numérique. Notre intention est désormais de situer la pertinence du XML dans ce schéma : quels sont les rôles des modèles dans la constitution de la pertinence ?

La pertinence de la composition XML pour l'information contenue dans la bibliothèque numérique, est fonction de deux facteurs : l'utilité et l'utilisabilité. La pertinence est ainsi fonction croissante de l'utilité en tant que possibilité d'atteindre un but visé ou d'accomplir une tâche (du moins une partie essentielle d'une tâche) grâce au langage XML. L'utilité est elle-même fonction de deux choses :

- le contexte d'énonciation, d'évolution³⁵ de la bibliothèque numérique : secteur professionnel, métiers engagés dans la création de la bibliothèque numérique, voire

³⁵ Voir aussi 2.2.1

même présence d'autres bibliothèques numériques qui peuvent la rendre redondante, complémentaire ou exprimer une contradiction ;

- la qualité de la documentation³⁶ : propriétés qualitatives inhérentes à la valeur de l'information (disponibilité, fiabilité, adaptabilité) en tant qu'unité de base.

Quant à l'utilisabilité, elle désigne la possibilité pour l'utilisateur de percevoir et d'interpréter l'apport du XML, mais aussi d'utiliser la bibliothèque numérique. L'utilisabilité est elle aussi fonction du contexte d'énonciation, d'évolution de la bibliothèque numérique (comme l'utilité !), mais aussi de la modalité de celle-ci, notamment toutes les caractéristiques des documents contenus³⁷ dans la bibliothèque numérique : type, quantité... Donc, la pertinence de la composition XML pour le développement d'une bibliothèque numérique est le résultat de la combinaison des deux termes. Elle concerne l'effet signifié par l'interprétation de ladite bibliothèque numérique par l'utilisateur dans un contexte particulier. D'où la nécessité du modèle de l'utilisateur pour caractériser et évaluer la pertinence de la composition XML, pourtant issue du modèle du domaine concepteur. D'après le modèle de l'utilisateur, dans une logique coopérative avec le modèle du domaine, nous avons ainsi pu déterminer trois indices essentiels de pertinence, valables pour l'évaluation du XML et des standards associés dans une bibliothèque numérique.

³⁶ Voir aussi 3.2.2

³⁷ Voir aussi 2.2.2

3.3.2 Les indices de pertinence

Il importe désormais de nous situer dans une nouvelle approche, celle de l'utilisateur, et non plus dans celle de l'invention technique. Quelques chercheurs, à l'instar de Jacques Perriault³⁸ dans son approche des logiques d'usage, ont ainsi consacré une analyse de l'innovation technique à rebours, où la création passe du pôle de l'invention à celui de l'utilisateur.

3.3.2.1 La notion de contextualisation

Pour être pertinent, le XML doit être « contextualisé », c'est-à-dire renvoyer à une réalité sociotechnique. Il faut en effet arriver à une certaine contextualisation du XML appliqué aux bibliothèques numériques, car le langage informatique donne trop souvent l'impression d'exister, de circuler au-delà de réalités sociales.

Contextualiser, c'est d'abord resituer le langage XML dans l'histoire de la communication informatique, comprendre les raisons de son émergence, ses forces et ses faiblesses, les comparer avec celles d'autres langages (HTML, SGML...), voire d'autres solutions (base de données relationnelle...). Comment inscrire le XML dans une histoire alliant un cadre technique et un cadre social, plutôt que de présenter le langage comme une solution révolutionnaire, voire même incomparable ? En d'autres termes, il faut relativiser le discours « mythique » vantant les mérites universels du XML, relativiser l'idée selon laquelle il ouvrirait une page de l'histoire de la communication informatique totalement vierge. Plus nous parlerons de « nouveau radical », plus il faudra se méfier. Le caractère révolutionnaire du XML est une idée reçue, parmi d'autres³⁹. L'application du XML aux bibliothèques numériques n'est ni facile, ni bon marché ; de plus le choix du langage n'a pas forcément été une évidence. Son développement est aussi porteur de lourdeur, parfois même de rigidité. Pourtant, presque paradoxalement, avoir conscience de tout cela permet de rendre le XML pertinent, notamment dans des cas d'application à des bibliothèques numériques. C'est pourquoi il nous faut systématiquement sortir du discours « mythique » et entrer dans une logique d'évaluation, comme il en existe par ailleurs pour toute activité technique, sociale, ou même économique, pour connaître les forces et les faiblesses du XML pour les bibliothèques numériques, pour étudier ses coûts, pour comprendre enfin ses performances, voire ses contreperformances. En somme, contextualiser le XML, c'est informer, expliquer, mais aussi

³⁸ PERRIAULT Jacques, *La logique d'usage. Essai sur les machines à communiquer*, Paris, Flammarion, 1989

³⁹ DHENIN Cyril, « Idée reçue n°3 : XML, c'est l'espéranto du Web », *JNet Solutions*, 13 mars 2002

critiquer ; c'est donc mettre de la distance avec son efficacité technique pour réaliser que le XML sera un jour dépassé par d'autres langages. Ces démarches sont des préalables indispensables à la compréhension du XML ; il faut en effet opérer une relativisation, au même titre qu'une réglementation du langage, afin que celui-ci s'inscrive dans le temps de l'histoire de la communication informatique et cesse d'apparaître comme étant hors de réalités sociales. Rendre pertinent le XML pour les bibliothèques numériques, c'est arriver à prendre conscience que le langage informatique agit sur les bibliothèques numériques, en particulier sur leur information et leur communication, lesquelles sont des réalités humaines et sociales bien plus compliquées que les dispositifs techniques qui les transmettent.

3.3.2.2 La notion de réception

Pour être pertinent, le XML doit être perçu, « réceptionné ». Comment rendre alors le XML ostensible, palpable par les utilisateurs des bibliothèques numériques ? En somme, comment ceux-ci peuvent-ils savoir, se rendre compte d'une présence du XML ?

La notion de réception assoit la légitimité et l'acceptation de la part des ses utilisateurs, des bibliothèques numériques qui ont pris l'initiative d'intégrer le XML et les standards associés. Ensuite, nous verrons que cela doit amener à l'appropriation des « nouvelles » bibliothèques numériques par des publics divers. Le XML ne saurait être pertinent s'il ne renvoie pas à une appropriation des dispositifs techniques qu'il engendre : c'est le passage à l'usage. Or, l'apport du XML aux bibliothèques numériques n'est ni intuitif, ni transparent. Pourtant il renvoie à une forme informatique nouvelle par de très nombreux aspects, entre autres modes de production et modes de circulation, dont les utilisateurs doivent avoir conscience. Cela implique la maîtrise cognitive et technique d'un minimum de savoirs et de savoir-faire. Elle est une condition sine qua non à la réception du XML et par voie de fait des nouveaux dispositifs techniques. En conséquence, les utilisateurs doivent développer un certain niveau de compétences spécifiques leur permettant de s'approprier les bibliothèques numériques. Il est donc nécessaire de percevoir, même indirectement, le XML qui nourrit les bibliothèques numériques. En effet, si l'on n'a pas les compétences pour apprendre à apprendre, les systèmes d'information et de communication demeureront imperméables. Rendre pertinent le XML pour les bibliothèques numériques, c'est donc aussi donner les moyens aux utilisateurs de percevoir l'apport du XML, afin de comprendre que les bibliothèques numériques ne se (re-)construisent pas ainsi de manière aléatoire, hors de tout contexte, hors de tout besoin et de toute perspective.

3.3.2.3 La notion d'usage

Pour être pertinent, le XML doit initier un, voire plusieurs usages. En ce sens, la notion d'usage transcende celle de réception, en lui donnant une conséquence matérielle. Comment qualifier alors l'apport du XML et des standards associés, en terme d'usage des bibliothèques numériques ? Quels phénomènes sociaux, déterminés d'après l'analyse des bibliothèques numériques, impliquent par exemple le recours au langage XML ?

La notion d'usage est liée à l'utilisation de dispositifs techniques, mais telle que construite par les recherches empiriques en sciences humaines, elle se démarque des stricts comportements d'utilisations. Elle donne les moyens d'observer, de classer et d'interpréter des phénomènes sociaux qui impliquent le recours à des objets techniques. En premier lieu, il faut renoncer à l'ambition de pouvoir s'affranchir vite et simplement de multiples contraintes économiques, techniques, scientifiques, etc.⁴⁰ liées au développement conjoint du langage XML et des bibliothèques numériques. Mais en contrepartie il faut élever les exigences de précision terrain par terrain, dans l'explicitation de démarches qui ont affaire avec :

- du flou : on ne peut pas présager de la pérennité du XML (quand sera-t-il remplacé par un nouveau langage encore plus « puissant » ?) ;
- des registres discursifs mixtes : on doit tenir compte à la fois des discours des informaticiens et des discours des documentalistes, car leurs approches du langage XML sont complémentaires ;
- des dynamiques multiples enchevêtrées : on doit pouvoir concilier la dynamique du secteur professionnel, la dynamique des métiers impliqués, et même la dynamique intrinsèque des bibliothèques numériques (celle relative au contenu ou à la présentation) ;
- des ambiguïtés.

Ces exigences de précision montrent que l'offre XML n'est pertinente que si elle se construit sur une rationalité prenant en compte les usages des bibliothèques numériques. Cette rationalité qui présage de la pertinence du XML, se fait nécessairement en deux temps. Le premier temps est celui de la performance techniciste⁴¹, qui est la démarche d'impact de la technique sur le social. Autrement dit, la technique agit sur le social, l'application du XML aux bibliothèques numériques engendrent de nouveaux usages (nouvelle appropriation des

⁴⁰ Voir aussi 1.2.1

⁴¹ Voir aussi 3.1.2

dispositifs). Le second temps est celui de la cohérence sociotechnique, qui est une démarche de rétroaction entre le social et la technique. Autrement dit, le social sur lequel la technique a eu un impact, vient rétroagir sur cette même technique. Seul le social pourra faire en effet la réussite des bibliothèques numériques, et par conséquent la pertinence du XML pour celles-ci. Ces deux temps de la rationalité construisent une sociologie des usages liés aux bibliothèques numériques. Dans le premier temps, les usages sont le fruit d'une sorte de négociation sociotechnique entre des concepteurs et des utilisateurs, via des études de besoins, des analyses de faisabilité... Dans le second temps, les utilisateurs finissent par s'imposer grâce à l'impact sociologique lié à l'apport du XML aux bibliothèques numériques.

Nous réalisons alors que l'application du XML aux bibliothèques numériques est le résultat de l'interaction entre des éléments : l'activité des différents acteurs, concepteurs et utilisateurs et le cadre sociotechnique. Dans L'innovation technique⁴², Patrice Flichy montre ainsi que « l'histoire de la technique et des usages se construit selon trois lignes entremêlées, celle du hasard, de la nécessité et du vouloir humain, individuel et collectif. » Quant à la pertinence de cette application du XML, elle a peut-être moins à voir avec le hasard, étant donné les coûts qu'elle peut engendrer, mais elle repose bien sur la connaissance des contraintes sociotechniques (notion de contextualisation) et sur l'activité des utilisateurs (notions de réception et d'usage), que l'activité des concepteurs se fait forte de solliciter.

⁴² FLICHY Patrice, *L'innovation technique. Récents développements en sciences sociales : vers une nouvelle théorie de l'innovation*, Paris, La Découverte, 1995

CONCLUSION

Notre étude nous a permis de dégager un certain nombre de contraintes stratégiques d'évaluation, regroupées sous la forme d'enjeux pour l'application du langage XML dans les bibliothèques numériques. Ces enjeux sont de trois ordres :

- enjeux industriels et économiques
- enjeux scientifiques
- enjeux sociaux

Les enjeux industriels et économiques : le positionnement matériel

La constitution d'enjeux industriels et économiques résulte principalement de deux facteurs :

- la prise en compte de besoins ou d'attentes latentes mal couvertes par l'état actuel de la technologie et des produits, et révélées par la proposition de solutions innovantes telles que le langage XML ;
- des propositions innovantes qui se doivent d'être concrètement nouvelles, suffisamment attractives et séduisantes pour engendrer des usages nouveaux, notamment pour les bibliothèques numériques.

Les enjeux scientifiques : la nécessaire réglementation

Le concept de langage ne suffit pas, à lui seul, à fonder une théorie de la communication. Un énoncé peut très bien être conforme au code (la syntaxe) et ne produire aucun effet si son récepteur ne lui connaît aucune pertinence. Le sens découle à la fois du code et de ce que le récepteur est capable d'inférer à partir de tout ce qui constitue, pour lui, le contexte d'emploi de ce code : la situation présente, mais aussi l'inscription « historique », les connaissances mobilisées... Cette démonstration générale s'applique parfaitement au cas du XML et des bibliothèques numériques. Bien sûr le langage informatique suscite une forme de communication pour les bibliothèques numériques, mais celle-ci ne se résume pas à la performance du système de transmission informatique ; elle renvoie à l'implication de modèles sociologiques (utilisateur et domaine). Dans son ouvrage Internet, et après ?⁴³, Dominique Wolton écrit ceci : « si la communication se réduisait à un échange performant,

⁴³ WOLTON Dominique, *Internet, et après ? : une théorie critique des nouveaux médias*, Paris, Flammarion, 2000

rationnel, rapide et libre d'informations, on ne parlerait plus 'de problèmes de communication, ni 'd'incommunicabilité' ». En réalité, il convient de distinguer la logique de la transmission, et son complément l'efficacité du XML, de la logique de la communication et la nécessité d'intercompréhension entre toutes les instances, depuis les concepteurs aux utilisateurs des bibliothèques numériques. La réglementation sert cette intercompréhension et rappelle ainsi que la communication ne peut jamais se réduire à une problématique de la transmission, c'est-à-dire à sa seule dimension technique. Les enjeux scientifiques renvoient donc à la mise en place d'une réglementation spécifique à l'application du XML dans les bibliothèques numériques. Elle est le moyen de rappeler la nécessité de penser la pertinence du XML pour les bibliothèques numériques dans une dimension sociotechnique, et de ne pas la réduire à la dimension purement technique.

Les enjeux sociaux : les significations d'usage

Les enjeux sociaux impliquent une réflexion sur la réception et les usages. Ces deux notions montrent que la technologie peut être évaluée au-delà de son efficacité technique. Dans l'article « L'intégration sociale des TIC : une sociologie des usages », Philippe Mallein et Yves Toussaint⁴⁴ ont montré que l'insertion sociale d'une technologie d'information et de communication, telle que nous pouvons considérer le langage XML, son intégration à la pratique des usagers, dépend certes dans un premier temps de ses qualités techniques intrinsèques, de ses performances objectives et de sa sophistication. Mais elle dépend surtout des significations d'usages projetées et construites par les utilisateurs sur le nouveau dispositif technique. En quelque sorte, et de manière plus ou moins consciente, les utilisateurs effectuent une sorte de procès en légitimité de l'usage du dispositif technique proposé. De l'issue de ce procès dépendront les conditions d'appropriation sociale de la nouvelle technologie, conditions représentatives de la pertinence du XML.

⁴⁴ MALLEIN Philippe, TOUSSAINT Yves, « L'intégration sociale des technologies d'information et de communication : une sociologie des usages », *Technologies de l'information et société*, Paris, Dunod, 1994, vol.6, n°4

En tant que phénomène d'abord technique, le développement des bibliothèques numériques avec le langage XML nous a donc permis de voir qu'il y a un enchevêtrement dynamique, même si instable, entre le pôle des technologies et celui de l'environnement social. En d'autres mots, tout objet ou dispositif technique porterait l'empreinte des rapports sociaux qui le sous-tendent et, en même temps, la dimension technique traverserait de plus en plus largement possible l'ensemble des formes sociales. Selon ce principe, notamment développé par Serge Proulx⁴⁵, lorsque des concepteurs proposent aujourd'hui une bibliothèque numérique intégrant le XML, s'ils veulent intéresser, mobiliser, impliquer des professionnels (documentalistes, bibliothécaires, archivistes...), des acteurs industriels ou même tout simplement des individus (utilisateurs finaux), ils doivent montrer que leur innovation technique est porteuse d'une nouvelle vision de la société, et surtout que cette innovation apportera des changements souhaitables pour la société. En d'autres mots, les innovateurs en matière technologique doivent se faire les porte-parole d'une vision sociale portée par leur innovation technique... sans malgré tout verser dans l'utopie et le mythe d'une solution universelle. Rendre pertinent le XML ne signifie pas en faire la seule voix d'une vision sociale, entre autres de transparence et d'interopérabilité des bibliothèques numériques. Au-delà des indices de contextualité (inscription du XML dans un cadre de référent sociotechnique), de réception (acceptabilité du XML par des utilisateurs) et d'usage (appropriation par ces mêmes utilisateurs), la pertinence du XML repose alors sur la nécessité de ne jamais généraliser ni son utilité, ni son usabilité. Dans cette logique particulariste, il serait désormais intéressant de donner une transcription concrète de ces indices de pertinence pour le XML, par exemple en développant des enquêtes, voire même en accompagnant et en observant réellement les utilisateurs des bibliothèques numériques dans leur compréhension du langage XML et dans leur appropriation des dispositifs techniques. Cela nous permettrait notamment d'offrir un regard concret et croisé entre les attentes des utilisateurs et les moyens des concepteurs.

⁴⁵ PROULX Serge, « La construction sociale des objets informationnels : matériaux pour une ethnographie des usages », Montréal, Université du Québec : présentation faite au Département des communications, avril 2000

BIBLIOGRAPHIE GENERALE

Les références mentionnées ci-dessous sont autant d'orientations théoriques pertinentes pour la compréhension des bibliothèques numériques, du langage XML et des pratiques d'évaluation et de réception consécutives à l'application du XML dans les bibliothèques numériques.

LES BIBLIOTHEQUES NUMERIQUES

Les contenus :

DESRICHARD Yves, « Bibliothèques électroniques », *Bulletin des bibliothèques de France (BBF)*, 2001, t.46, n°1, p.111-113.

DUPOIRIER Gérard (coordinateur), « Les bibliothèques numériques », *Document numérique*, Paris, Hermès Science Publications, 1998, vol.2, n°3-4.

LE MOAL Jean-Claude, HIDOINE Bernard (coordinateurs), *Bibliothèques numériques : cours INRIA, 9-13 octobre 2000, La Bresse*, Paris, ADBS Editions, 2000.

Le traitement :

Groupe interuniversitaire de recherche en Sciences de l'information et de la communication, « Normes et documents numériques », *Solaris*, décembre 1999-janvier 2000, n°6.

JACQUESSON Alain, RIVIER Alexis, *Bibliothèques et documents numériques : concepts, composantes, techniques et enjeux*, Paris, Editions du Cercle de la Librairie, 1999.

LUPOVICI Christian, « La chaîne de traitement des documents numériques », *Bulletin des bibliothèques de France (BBF)*, 2002, t.47, n°1, p.86-91.

LE LANGAGE XML

La technique :

Dynamiser vos informations avec XML, Groupement français de l'industrie de l'information (GFII), Paris, 23 octobre 2001.

XML : pour quoi faire ?, Association des professionnels de l'information et de la documentation (ADBS), Paris-La Défense, 4 décembre 2001.

BOSAK Jon, BRAY Tim, « Le langage XML », *Pour la science*, juillet 1999, n°261.

LAZINIER Emmanuel, « Des documents qui prennent vie, ou comment XML va transformer la façon dont nous concevons et exploitons nos documents », Paris, Ministère de la Défense : présentation faite à CALS Europe 98, 17 septembre 1998.

MIREMONT Liliane, «eXtensible Markup Language », *Bulletin des bibliothèques de France (BBF)*, 2001, t.46, n°2, p.109-110.

Les applications :

CHAUDIRON Stéphane, IHADJADENE Madjid, ROLE François, « Exploitation de l'indexation par les auteurs pour le filtrage de documents structurés », *Filtrage et résumé automatique de l'information sur les réseaux – Actes du 3^{ème} Colloque du Chapitre français de l'ISKO*, sous la direction de CHAUDIRON S. et FLUHR C., Nanterre 5-6 juillet 2001, Université de Paris X, p.39-47.

CHAUDIRON Stéphane, IHADJADENE Madjid, ROLE François, « CodeX : un système pour la définition de vues multiples guidées par les usages », *Document électronique dynamique - Actes du 3^{ème} Colloque international sur le document électronique (CIDE 2000)*, sous la direction de GAIO M. et TRUPIN E., Lyon 4-6 juillet 2000, Europa Productions, p.71-81.

LA DEMARCHE D'EVALUATION

Principes de pertinence :

CALABRETTO Sylvie, PINON Jean-Marie, POULLET Line, RICHEZ Marc-Antoine, « De la qualité de l'information à la qualité de la documentation », *Document numérique*, 1998, vol.2, n°1, p.37-52.

GRICE Paul H., *The conception of value*, Oxford (GB), Clarendon Press, 1991.

SPERBER Dan et WILSON Deirdre, *La pertinence : communication et cognition*, Paris, Editions de Minuit, 1989.

Stratégies d'évaluation :

CASANOVA Xavier, COHEN Joëlle, « L'écran efficace : une approche cognitive des objets graphiques », *Documentaliste – Sciences de l'information*, 2001, volume 38, n°5-6, p.272-289.

COOKE Alison, *A guide to finding quality information on the Internet : selection and evaluation strategies*, London, Library Association publishing, 2001.

CRAMPES Jean-Bernard, *Interfaces graphiques ergonomiques (conception et modélisation)*, Paris, Ellipses, [Technosup], 1997.

LAFOUGE Thierry, « Evaluation des systèmes d'information documentaire », Université de Lyon 1 – Projet Recodoc.

QUIVY Raymond, VAN CAMPENHOUDT Luc, *Manuel de recherche en sciences sociales (2^{ème} édition)*, Paris, Dunod, 1995.

L'INNOVATION TECHNIQUE

L'ANALYSE DES USAGES

Théories de la réception :

BERTRAND Gisèle (directeur), « Les nouvelles technologies : quels usages, quels usagers ? », *Dossiers de l'audiovisuel*, Paris, mai-juin 2002, n°103.

BIGOT Régis, *La diffusion, l'usage et l'acceptabilité des nouvelles technologies en France*, Paris, Crédoc [Rapports], 2001.

GUICHARD Eric (directeur), *Comprendre les usages de l'Internet*, Paris, Editions Rue d'Ulm, 2001.

JACQUESSON Alain, « De la difficulté à utiliser les bibliothèques numériques », *ABF bulletin d'information*, juillet-septembre 2000, n°188.

LE COADIC Yves-François, *Usages et usagers de l'information*, Paris, Nathan, collection 128, 1997.

Théories de l'appropriation :

AKRICH Madeleine, « Les utilisateurs, acteurs de l'innovation », *Education permanente*, 1998, n°134.

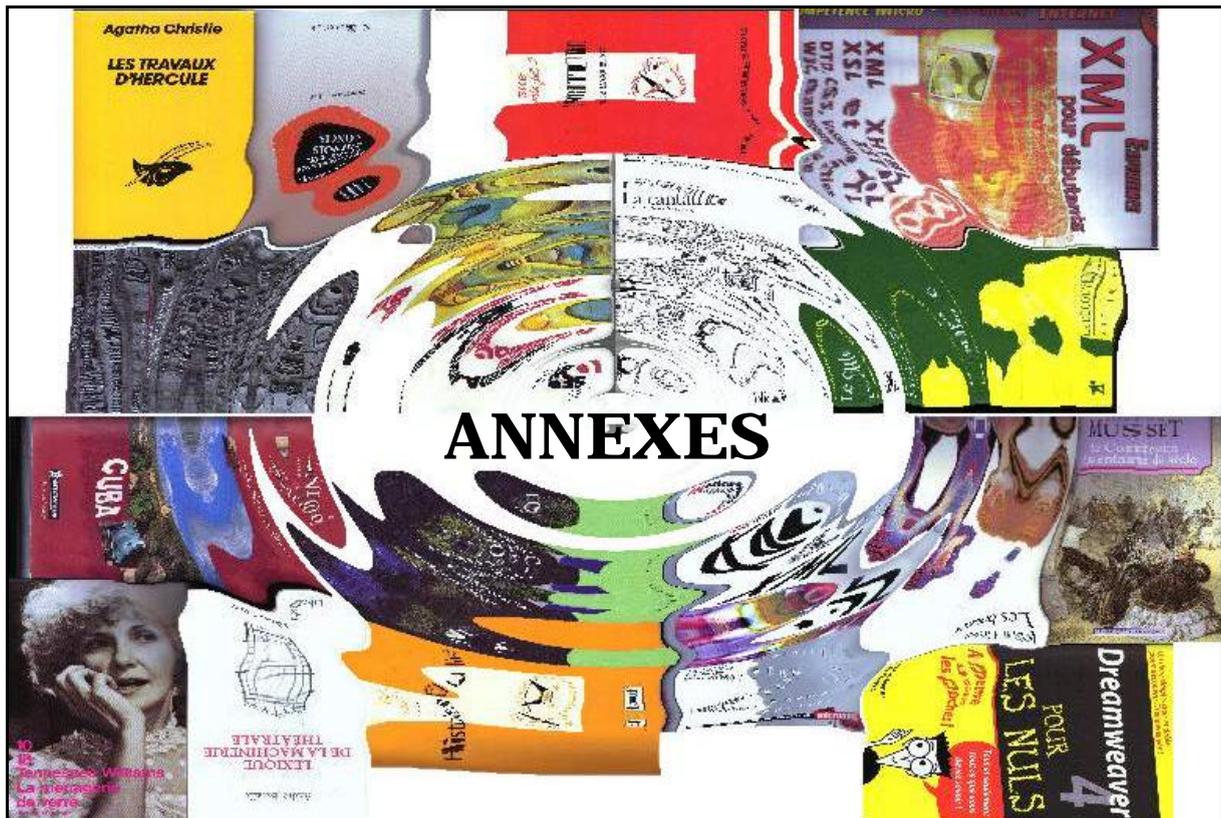
FLICHY Patrice, *L'innovation technique. Récents développements en sciences sociales : vers une nouvelle théorie de l'innovation*, Paris, La Découverte, 1995.

MALLEIN Philippe, TOUSSAINT Yves, « L'intégration sociale des technologies d'information et de communication : une sociologie des usages », *Technologies de l'information et société*, Paris, Dunod, 1994, vol.6, n°4.

PERRIAULT Jacques, *La logique d'usage. Essai sur les machines à communiquer*, Paris, Flammarion, 1989.

PROULX Serge, « La construction sociale des objets informationnels : matériaux pour une ethnographie des usages », Montréal, Université du Québec : présentation faite au Département des communications, avril 2000.

WOLTON Dominique, *Internet, et après ? : une théorie critique des nouveaux médias*, Paris, Flammarion, 2000.



ANNEXE 1 : Revue de presse : « quelle actualité pour les bibliothèques numériques intégrant XML ? » 95

ANNEXE 2 : Rencontres professionnelles : « quelle (autre) actualité pour les bibliothèques numériques intégrant XML ? » 103

ANNEXE 3 : Grille d'analyse : modèle descriptif 109

ANNEXE 4 : Glossaire récapitulatif : terminologie liée à XML 111

ANNEXE 5 : Guide de sites Web : ressources électroniques sur le XML et les bibliothèques numériques 113

Annexe 1

REVUE DE PRESSE

« Quelle actualité pour les bibliothèques numériques intégrant XML ? »

GALLICA :

LE HIR Pierre, « La BNF numérise ses collections rares pour les mettre en valeur », *Le Monde interactif*, 7 mars 2002 96

PELLEAS :

CROCHET DAMAIS Antoine, « Comment l'Université de Marne-la-Vallée gère sa base documentaire avec XML », *Journal du Net*, 7 décembre 2001 98

LE POINT :

CROCHET DAMAIS Antoine, « Comment Le Point passe de HTML à XML », *Journal du Net*, 23 janvier 2002 100

LES ECHOS :

GUERRIER Philippe, « LesEchos.fr : place à la 'transversalité' de l'information », *Journal du Net*, 21 mai 2002 102

NB : Autres références non-inclus

ARTICLE@INIST :

GRESILLAUD Sylvie, « La recherche et la commande de documents sur le Web de l'INIST », *ABF bulletin d'informations*, 2^{ème} trimestre 2000, n°187

CODEX :

CHAUDIRON Stéphane, ROLE François, IHADJADENE Madjid, « Mise en ligne des mémoires de thèse : le cas de CodeX », *ABF bulletin d'informations*, 2^{ème} trimestre 2000, n°187

Le Monde.fr

La BNF numérise ses collections rares pour les mettre en valeur

€ LE MONDE | 06.03.02 | 12h31

Plus de 50 000 ouvrages, 450 périodiques et 80 000 images sont déjà disponibles sur le site de la bibliothèque en ligne Gallica. Le catalogue encyclopédique initial s'enrichit aujourd'hui de collections spécialisées et de fonds jusqu'alors dispersés.

La bibliothèque virtuelle sans murs ni livres, dont quelques visionnaires avaient éviscéré lors de la conception la conception de la Bibliothèque nationale de France (BNF), restera sans doute longtemps encore une utopie. Au grand soulagement des bibliophiles attachés - passion ou raison - à la matérialité des ouvrages. Il n'empêche que, parmi les grandes bibliothèques mondiales, l'institution nationale est l'une des plus résolument engagées dans la numérisation de ses collections, aux côtés de la bibliothèque du Congrès américain, de la New York Public Library, de la British Library de Londres, de la Bodleian Library d'Oxford ou encore des centres documentaires des grandes universités américaines (Cornell, Yale, Berkeley) ou allemandes (Göttingen, Munich).

Gallica, la bibliothèque numérique en ligne de la BNF, ouverte en 1997 et d'accès gratuit, réunit aujourd'hui 54 000 ouvrages et 450 titres de périodiques, soit 16 millions de pages, complétées par 80 000 images. Elle enregistre un très beau succès, puisque chaque jour 27 000 pages sont "feuilletées" et plus de 10 000 images affichées sur son site (<http://gallica.bnf.fr>). De surcroît, les cyberlecteurs téléchargent quotidiennement plus de 500 000 pages sur les disques durs de leurs ordinateurs.

Après les années pionnières, "la Bibliothèque nationale entre dans une deuxième phase de numérisation", indiquent Caroline Wiegandt, directrice des services et réseaux, et Catherine Lupovici, directrice du département de la bibliothèque numérique. La première étape visait à la constitution d'un fonds à vocation encyclopédique, universel non par le nombre de volumes, mais par la couverture de l'ensemble des champs du savoir - de la littérature aux sciences et techniques en passant par l'histoire et la philosophie -, du Moyen Âge au début du XXe siècle. Ce corpus généraliste lui ayant permis "de se faire la main et de maîtriser les contraintes liées à la numérisation", l'équipe de la BNF a désormais entrepris d'intégrer à Gallica des collections spécialisées. Après la publication en ligne de dossiers consacrés à Proust - avec notamment le manuscrit du *Temps retrouvé* - ou au thème de l'utopie, une série "baladeuse" a été ouverte au printemps 2001, avec les *Voyages en France* qui seront suivis dans les prochaines semaines des *Voyages en Afrique*, puis des *Voyages en Italie*.

Une collaboration avec les sociétés savantes régionales se traduit déjà par la mise sur le réseau de près de 600 000 pages extraites de 36 revues publiées, entre 1750 et 1924, en Aquitaine et en Lorraine. D'autres projets portent sur les documents relatifs aux galeries théâtrales au XIXe siècle, les manuscrits enluminés, ou encore les incunables.

"La numérisation est un moyen de valoriser et de faire connaître des collections exceptionnelles", se félicitent Caroline Wiegandt et Catherine Lupovici. Elle permet de mettre à la disposition des chercheurs et du public des ouvrages rares ou précieux - parfois trop fragiles pour être consultés sous leur forme imprimée -, de réunir sur un même support des fonds dispersés entre différents sites, ou encore de constituer des ensembles thématiques balayant plusieurs disciplines. Elle constitue aussi un outil de préservation à long terme du patrimoine documentaire, qui n'a toutefois pas encore convaincu tous les bibliothécaires. "La pérennité des supports numériques suscite encore des interrogations dans la profession", constatent les deux directrices.

Aussi la BNF procède-t-elle pour l'instant à un double traitement de ses documents, qui sont numérisés et microfilmés au cours d'une même opération, grâce à un banc de reproduction mixte associant une tête de scannage et une caméra analogique. L'établissement s'apprête d'ailleurs à acquérir une deuxième installation de ce type. A terme cependant, le microfilm, considéré aujourd'hui par certains professionnels comme un moyen de conservation plus fiable, sera sans doute abandonné au profit du seul support informatique. Pas question, toutefois, de supprimer les documents dès une fois ceux-ci numérisés : "Ce serait aussi absurde que de jeter un tableau sous prétexte qu'on en possède une belle photo ! En même temps que nous allons par exemple numériser les journaux du XIXe siècle, qui sont en mauvais état, nous allons les restaurer, les renforcer et les ranger avec soin."

Les opérations de numérisation, sous-traitées pour le "tout-venant" mais réalisées en interne pour les documents précieux, s'effectuent de façon assez semblable à celles de microfilmage, dans des ateliers où la température et l'hygrométrie sont strictement contrôlées. "La nouveauté majeure, qui demande un énorme travail complémentaire, est l'intégration de métadonnées permettant d'exploiter les ouvrages ou

les collections numérisées. En effet, si les documents sont, pour des raisons de coût, numérisés en mode image (un fac-similé restituant l'aspect visuel de l'original), les tables des matières et les légendes sont saisies en mode texte, ce qui permet d'y rechercher des mots-clés et d'accéder aux têtes de chapitre, pages ou illustrations correspondantes. Mais le travail d'indexation est beaucoup plus riche et donc plus complexe qu'avec les documents traditionnels.

Il ne s'agit plus seulement de rédiger une notice de présentation d'un ouvrage, assortie d'une référence permettant de le retrouver sur une étagère. Chaque page, chaque image doit être identifiée. Des informations de nature juridique, indiquant si le document est ou non libre de droits, doivent être intégrées. Surtout, des liens doivent être établis avec d'autres documents, afin de tirer pleinement parti des fonds numérisés.

C'est ainsi que la série des *Voyages en France* offre une multiplicité d'entrées, par genre littéraire (fiction, récit, guide, correspondance, document administratif...), siècle, moyen de transport (à pied, à cycle, en véhicule hippomobile...), motif de déplacement (exil, guerre, voyage professionnel, sentimental, de santé...), type de voyageur (artiste, enfant, étranger...) ou nationalité. De même la future collection "Mémoires" sera-t-elle pourvue de liens qui renverront, lorsque l'on cliquera sur "Mallarmé", à son portrait par Gauguin, à la traduction par le poète du *Corbeau* de Poe, ou au *Prélude à l'après-midi d'un faune* de Debussy inspiré de son œuvre.

Déjà, la Bibliothèque nationale prépare la "troisième étape" de son programme de numérisation, annoncent Caroline Wiegandt et Catherine Lupovici. Il lui faut désormais, non plus seulement mettre en valeur ses propres collections, mais "s'insérer dans des réseaux" en développant des coopérations avec les grandes bibliothèques mondiales. Cela, afin d'éviter les doublons inutiles, mais surtout de mettre en commun leurs ressources et leurs collections, entre lesquelles pourraient librement voyager les internautes. Ainsi se dessine l'architecture, bien éhée, d'une bibliothèque virtuelle universelle.

Pierre Le Hir

Internet, Tolbiac et Richelieu

La bibliothèque en ligne Gallica ne réunit qu'une partie (55 % des imprimés et 35 % des images) du fonds numérisé de la BNF. Seuls sont disponibles sur Internet les documents libres de droits, dont les auteurs sont morts depuis plus de 70 ans, les autres pouvant être consultés sur 80 postes de travail aménagés sur les sites de Tolbiac et de Richelieu. Au total, le fonds numérisé de la BNF comprend 100 000 ouvrages imprimés, dont 80 % de livres et 20 % de périodiques, soit 32 millions de pages. S'y ajoutent 300 000 documents iconographiques, ainsi que équivalent de 1 000 heures de son et de 500 heures de vidéo. Ce catalogue s'accroît de 10 000 volumes et 20 000 images par an. Il ne représente cependant qu'une fraction du patrimoine de la BNF, riche de plus de 13 millions de livres et imprimés ainsi que de 350 000 collections de périodiques, sans compter plusieurs millions d'estampes, photographies, affiches, cartes, plans ou documents sonores et visuels.

[← Retour](#)

Droits de reproduction et de diffusion réservés; © Le Monde 2002
Usage strictement personnel. L'utilisateur du site reconnaît avoir pris connaissance de la licence de droits d'usage, en acceptant et en respectant les dispositions.
Politique de confidentialité du site.

Comment l'Université de Marne la Vallée gère sa base documentaire avec XML

Par le JDN Solutions (Benchmark Group)

URL : http://solutions.journaldunet.com/0112/011207_pelleas.shtml

Alors qu'Internet a été investi dès son lancement par le monde universitaire, les évolutions des normes du réseau n'ont par la suite pas toujours été suivies par les chercheurs et enseignants. Force est de remarquer néanmoins l'émergence récente de plusieurs programmes lancés par divers rectorats, utilisant les derniers standards Web. Aux Etats-Unis, c'est notamment le cas de l'Université de l'Indiana et de l'Université Syracuse (New York). En France, l'initiative revient à l'Université de Marne-la-Vallée avec une première expérience de grande ampleur autour du Web sémantique. "Le projet Pelleas a été lancé en 1997 suite à une demande conjointe de notre président et de la Direction de l'Information Scientifique et Technique du ministère de l'Education Nationale", explique Christian Lupovici, directeur du Service commun de la Documentation de l'université. Au total, l'objectif fixé est alors de réaliser une bibliothèque électronique propre à l'établissement, mais également de jouer le rôle de site pilote en vue d'un déploiement éventuel sur d'autres campus.

XML : un choix considéré comme naturel

Prenant acte des différents formats documentaires généralement utilisés en 1997 (HTML, PDF, etc.), le premier cahier des charges du projet recommande l'utilisation de SGML (Standard Generalized Markup Language), le langage historique pour gérer des documentations lourdes. "Un visualiseur propriétaire était néanmoins nécessaire à tout accès client", rappelle Christian Lupovici. Suit une période de deux ans, lors de laquelle le ministère effectue un travail de vérification des pré-requis techniques de la proposition, puis de validation des objectifs cadres. "Ce laps de temps a permis aux spécifications d'évoluer -et notamment de prendre en compte l'émergence de la technologie XML", commente le responsable du programme. A l'instar de SGML, l'avantage de XML (eXtended Markup Language) réside dans ses qualités d'ouverture. "Cette différence est particulièrement perceptible concernant l'interfaçage", insiste-t-on à l'Université de Marne-la-Vallée. Sans nécessiter de pilote propriétaire, XML peut recourir à des feuilles de style particulières (XSLT) pour assurer le mode de publication des données.

Au delà de la nécessité d'un langage de marquage, les objectifs du projet imposaient la couverture de l'ensemble de la chaîne de traitement des documents universitaires : depuis l'indexation, en passant par le stockage et la diffusion. Afin d'accompagner la définition des spécifications techniques, ainsi que les phases de développement et de déploiement, l'université fait appel à Jouve : une société de services spécialisée dans le traitement de l'information électronique et papier.

Ex Libris marié à d'autres modules sur mesures

"Après avoir comparé différentes solutions disponibles sur le marché, nous avons finalement recommandé le système de base de données documentaire Ex Libris (Cadic)", précise Patrick Kalifa, responsable du suivi du projet Pelleas chez Jouve. Choisi comme socle du système pour sa souplesse, ce produit associe une base de données à une interface de programmation d'applications (API). "Ce qui nous a permis d'effectuer aisément les développements complémentaires qui s'imposaient", précise le porte-parole.

Plusieurs fonctions ont notamment été ajoutées par le prestataire. En amont figure la brique d'acquisition des données. Celle-ci s'appuie sur des DTD (Document Type Definition) ou des schémas XML normalisés (EBIND, TEI, etc.) pour reprendre des documents XML existants et leur attribuer un typage (titre1, titre2, chiffre, etc.) propre au monde de la gestion documentaire. "Ici, il s'agit par exemple de manuels universitaires, de revues spécialisées, ou encore de thèses et autres mémoires de recherche", pointe Christian Lupovici. "Notez que des DTD différentes peuvent être utilisées en fonction des contextes : EAD pour l'archivage, MathML dans le domaine des mathématiques, TCI (Text and Coding Initiative) dans celui des sciences humaines et sociales, etc." Qu'en est-il des fichiers non textuels (image, PDF, son, etc.) ? "Leurs meta-données sont généralement extraites automatiquement, puis encapsulées en XML au format Dublin Core -un dérivé du fameux RDF", détaille Patrick Kalifa.

Un lancement prévu pour septembre 2002

En partant de scripts livrés avec Ex Libris, Jouve a également mise en place les interfaces finales de la plate-forme Pelleas. En partie générées en PHP, celles-ci font intervenir un moteur XSLT assurant la transformation au format HTML des documents XML et autres tables navigationnelles stockés dans la base Cadic. Associée aux fonctions de gestion des droits d'Ex Libris ainsi qu'au moteur de recherche d'Hummingbird, la solution propose au final un système d'accès à l'ensemble de la base Pelleas.

"Via un protocole spécifique à l'univers des bibliothèques (Z3950), il est également possible d'interroger nos bases de prêts relatives aux documents papier", confie Christian Lupovici. C'est également sur cette technologie que l'université prévoit de s'appuyer pour se connecter aux applications d'autres universités. Pour l'heure, la plate-forme de Marne-La-Vallée, qui tourne actuellement sur 2 serveurs HP 9000, est en

pleine phase de tests. Le lancement officiel est prévu pour septembre 2002.

[[Antoine Crochet Damais](#), JDNet]

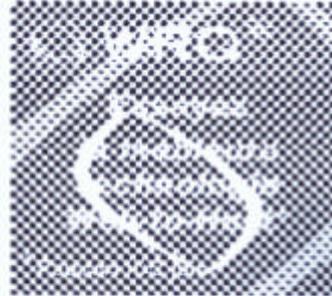
Pour tout problème de consultation, écrivez au webmaster
Copyrights et reproductions : [Données personnelles](#)
Copyright 2001 Benchmark Group 4, rue Diderot
92156 Suresnes Cedex, FRANCE

Comment le Point passe de HTML à XML

Par le JNNet Solutions (Benchmark Group)

URL : http://solutions.journaldunet.com/0201/020123_lepoint.shtml

A l'origine, le site du journal *Le Point* s'adossait à la première version de la plate-forme de gestion de contenu Ratio Media. Basée sur le couple PHP/MySQL, la solution de Rosebud Technologies impliquait alors d'intervenir directement dans le code HTML de la page d'accueil à chaque mise à jour. Une opération longue et fastidieuse que devait prendre en charge le webmaster... voire le fournisseur en cas de modifications graphiques plus profondes. Fixé courant 2001, l'objectif principal de la nouvelle version du site (en ligne depuis le 22 décembre) est de s'assurer un meilleur contrôle du système de publication -afin notamment d'être moins dépendant du prestataire-, mais également de s'appuyer sur le langage XML pour améliorer la hiérarchisation de l'information.



Un cache optimise le temps d'accès aux pages

"Rosebud, qui est notre prestataire technique historique sur le projet, a lancé une mise à jour de sa solution juste au bon moment", commente Sébastien Lelieur, webmaster du site du journal. Développée en Java, la nouvelle édition de Ratio Media (voir notre article sur le sujet) s'articule justement autour d'une couche de modélisation de données décrites à l'aide du langage de balisage XML. "Ce qui nous permet d'intégrer des marqueurs pour 'portaliser' la publication des articles en fonction de thématiques pré-déterminées (cinéma, littérature, etc.)", détaille le porte-parole. Le fameux standard facilite aussi la réutilisation des données Web, qu'elles soient récupérées pour une application de CD-Rom ou encore transmises à l'Européenne de Données dans le cadre d'un contrat de production de contenu délégué.

Hébergé chez Rosebud, le site du Point tourne sur un serveur Web Apache couplé à un serveur d'applications Tomcat chargé d'exécuter Ratio Media. Le tout installé sur une machine Sun bi-processeurs. Basé sur une seconde machine (PC), un noyau Linux supporte la base de données -qui reste pour l'heure l'application Open Source MySQL. En vue d'améliorer le temps d'affichage des pages, un module livré avec Ratio Media se charge de stocker à la volée les contenus les plus consultés sous forme statique dans un cache intermédiaire : une mesure qui évite aux utilisateurs d'avoir à lancer systématiquement l'ensemble des requêtes nécessaires au chargement d'une page Web.

Un couche de modélisation de données XML

Afin de mettre à jour la nouvelle version de son site, *Le Point* commence par traduire ses articles au format Quark Xpress dans un vocabulaire XML. Prise en charge par Rosebud à partir d'un jeu de balises préalablement défini (ou DTD), la génération des fichiers XML est contrôlée ensuite par l'équipe du journal par le biais de l'éditeur XMetal. Une fois cette phase préparatoire menée à bien, les documents sont transmis via une connexion FTP sur le serveur distant, puis automatiquement intégrés à la base de données MySQL. "Pour finir, il ne reste plus qu'à mettre à jour la page d'accueil et la publier", ajoute Sébastien Lelieur. Une opération qui s'effectue via l'interface d'administration de Ratio Media. Entièrement visuel, ce backoffice permet de définir entièrement le contenu éditorial de la Une : depuis les entrées et leur positionnement dans la page en passant par l'appel des titres, des chapeaux et des images. Sans compter le choix du statut "payant" ou "gratuit" de chaque article. Principaux avantages de l'outil selon le responsable : "la possibilité effective de se décharger de l'ensemble de la procédure sur un 'non webmaster', ainsi que la grande souplesse en matière d'ergonomie -notamment autour de la création de nouvelles rubriques dans la DTD".

Aux côtés de la gestion éditorial pure, le système d'administration permet également de créer des visuels relatifs à des partenariats avec des sites tiers autour de contenus thématiques (programme de télévision, météo, etc.), puis d'intégrer des liens vers ces derniers dans des blocks de la page d'accueil. "Si l'on désire effectuer des modifications plus importantes au sein de la page d'accueil, Ratio Media donne accès au code HTML des principaux blocks", complète Sébastien Lelieur.

Une version pour préparer la reprise du marché...

Sur la question du bilan des premières semaines de mise production de cette V2, le responsable du Point reconnaît un certain retard dans phase de migration de la base de données initiale (HTML) vers la nouvelle (XML). Cependant, il souligne "une meilleure performance générale de la nouvelle version". Parmi les points d'amélioration notables figureraient notamment la vitesse d'affichage des pages -qui aurait été multipliée par cinq-, ainsi qu'une bonne capacité de montée en charge de l'architecture. "Nous encaissons sans problème des pics d'audience de près de 8000 visiteurs uniques par jour", note-t-il. Moins chère que la mouture précédente, la version actuelle du site nécessiterait aussi un nombre plus faible d'intervenants. "Avec l'affinage de la segmentation des sujets, cet ingrédient entre dans notre réflexion concernant le retour sur investissement du projet", insiste Sébastien Lelieur. Et pourquoi la catégorisation du contenu est-elle à ce

point critique ? "Elle assure l'optimisation de l'offre d'espaces destinée aux partenaires potentiels", conclut le porte-parole... Dans l'attente de la reprise du marché de la publicité en ligne.

[Antoine Crochet Damais, JDNet]

Pour tout problème de consultation, écrivez au webmaster
Copyrights et reproductions . Données personnelles
Copyright 2001 Benchmark Group 4, rue Diderot
92156 Suresnes Cedex, FRANCE

LesEchos.fr : place à la "transversalité" de l'information

Par le JDNet (Benchmark Group)

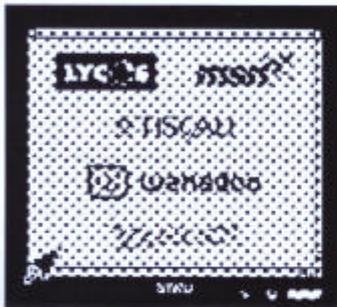
URL : <http://www.journaldunet.com/0205/020521lesechos.shtml>

Mardi 21 mai 2002

Philippe Jannet, le directeur des éditions électroniques des Echos, est ravi : la nouvelle version du site du quotidien économique constitue l'aboutissement d'un chantier entamé il y a deux ans. Il portait sur l'optimisation de l'information proposée aux lecteurs du quotidien en ligne et aux clients des Echos. "Nous voulions assurer une transversalité totale de l'information, explique Philippe Jannet. L'idée est de proposer un maximum de liens à partir d'une information. En clair, lorsqu'un article est paru sur telle entreprise, l'internaute aura non seulement le lien hypertexte vers le site de l'entreprise concerné, son cours de Bourse mais aussi les autres articles en lien dans la chaîne d'information en continu et ceux liés au secteur de l'entreprise et à ses concurrents."

En savoir plus sur...

► **Philippe Jannet**
► **Articles JDNet**
[LesEchos.fr intègre MesFinances.fr](#)
(02/01)



L'information, devenue l'unité de base, est indexée en XML. "Le nouvel outil donne de la flexibilité dans le traitement et la diffusion de l'information. Ces flux de XML intelligents sont utiles, notamment lorsque l'on veut envoyer des informations spécifiques à une société cliente qui effectue une veille particulière", commente le patron Internet des Echos. C'est l'équipe Internet du groupe qui a redeveloppé un outil d'édition pour assurer le flux des informations à partir du progiciel Datox, avant que la rédaction électronique - une vingtaine de journalistes - se l'approprie.

Du côté du site, la navigation a été simplifiée. Une nouvelle barre de navigation, avec cinq onglets principaux, a été montée : "Les Echos" (reprenant l'édition du journal), "Info en direct" (information en continu assurée par la rédaction Web), "Finances personnelles" (qui constitue la partie intégrée de l'ancien service MesFinances.fr), l'espace "Bourse" et une nouveauté : l'onglet "Entreprises" qui permet d'effectuer une recherche sur les noms d'entreprises figurant dans l'actualité. Un système d'icônes de couleur a été mis en place pour signaler les accès payants ou gratuits. "On nous a longtemps reproché d'avoir un site extrêmement riche mais dont la navigation faisait défaut. Nous espérons avoir trouvé une formule satisfaisante avec cette nouvelle version", indique Philippe Jannet.

Le site des Echos enregistrait 2,5 millions de visites et 16 millions de pages vues en avril 2002 (source Cybermétrique) et revendique la place de premier site d'information économique et financière. "Nous avons compensé la perte d'audience enregistrée du côté de l'espace Bourse par la hausse de la partie information, commente Philippe Jannet. Notre objectif n'est plus d'augmenter notre nombre de visites. Tout en maintenant notre audience actuelle, nous voulons en priorité améliorer notre taux de pénétration sur les cadres, notre cible privilégiée."

Le site

► [Lesechos.fr](#)

Philippe Jannet estime que le Web des Echos parviendra à un équilibre entre ses trois sources de revenus - publicité, abonnements en ligne et vente de contenu - d'ici la fin de l'année. La tendance actuelle est à une "granularisation des offres", c'est à dire développer du "sur mesure" pour le compte de clients spécifique ou d'un segment de clientèle. Mais le directeur des éditions électroniques des Echos ne détecte pas d'amélioration significative du côté des revenus publicitaires. A l'instar des autres entités Internet du groupe Pearson (propriétaire, entre autres, du Financial Times et des Echos), LesEchos.fr, ouvert en 1996, a entamé une politique de rationalisation de ces coûts afin de parvenir à un équilibre financier. En juillet 2001, le groupe Pearson avait indiqué vouloir réduire de 16% les coûts de ses activités en ligne pour compenser en partie la chute des revenus publicitaires.

En savoir plus sur...

► **Philippe Jannet**
► **Articles JDNet**
[LesEchos.fr intègre MesFinances.fr](#)
(14/02/01)

[Philippe Guerrier, JDNet]

Pour tout problème de consultation, écrivez au webmaster
Copyrights et reproductions : Données personnelles
4, rue Diderot - 92156 Suresnes Cedex, FRANCE -
Hébergement: ELUXUS

Annexe 2

RENCONTRES PROFESSIONNELLES

« Quelle (autre) actualité pour les bibliothèques numériques intégrant XML ? »

GALLICA :

Gallica ou la bibliothèque à l'écran, LUPOVICI Catherine, Rencontres multimédia 'Nouveaux usages, nouveaux usagers', 10 décembre 2001 104

GUIDE DE L'INTERNET CULTUREL :

L'Internet culturel : culture et numérisation, vers de nouveaux espaces publics, DALBERA Jean-Pierre, Séminaire européen 'Culture et Internet', 22 septembre 2001 105

ARTISTES ET ŒUVRES:

La Médiathèque du Centre national de la Danse : un projet bâti sur une logique de travail en réseau, SEBILLLOTTE Laurent, Journée d'études 'Réseaux, échanges, partenariats, 14-17 mars 2002 106

CODEX :

CodeX : un système pour la définition de vues multiples guidées par les usages, CHAUDIRON Stéphane, CI DE 2000 'Bibliothèques numériques', 4-6 juillet 2000 107

PELLEAS :

Pelleas : une plate-forme universitaire pour la gestion des documents électroniques, LUPOVICI Christian, 9^{ème} forum TICE/NTE 'L'information scientifique et technique à l'heure des TICE : accès, recherche, exploitation', 8 février 2001 108



Cité de la Musique - Paris
<http://www.cite-musique.fr>

RENCONTRES MULTIMÉDIA DE LA CITE DE LA MUSIQUE
JOURNÉE D'ÉTUDE - LUNDI 10 DÉCEMBRE 2001
"NOUVEAUX USAGES, NOUVEAUX USAGERS :
L'INFORMATION MUSICALE SUR INTERNET"

La journée est consacrée aux nouveaux usages et modes de pratique de l'Internet. La question sera traitée sous un double aspect :

- interroger les transformations d'usages professionnels : la recherche et la sélection des informations, leur validation, enfin leur retraitement dans un but documentaire se font à l'aide du même médium technique. Aux yeux du profane comme du spécialiste, le Web apparaît comme un gisement quasi inépuisable d'informations permettant une autonomie maximale de l'utilisateur dans sa recherche. Mais en parallèle on suppose, pour un usage bien tempéré, l'intercession d'un médiateur documentaire (humain ou technique : formalisation en bases de données, portails ou moteurs de recherche) pour recomposer cette masse d'informations à l'aide notamment des formes classiques du savoir : l'usage de types, de catégories ou de classifications hiérarchisés.
- interroger les pratiques des usagers : vis à vis des médias, les modèles culturels se caractérisent aujourd'hui par un triple déplacement : le passage d'une réception de masse à la réception individuelle, de la passivité à l'interactivité, du spectacle à la communication. Face au projet de structurer l'informe, d'ordonner l'indistinct, d'accorder de la pertinence, donc de la valeur, aux messages en circulation incessante sur les réseaux qui correspond à la fonction même d'un professionnel de la documentation, les usagers semblent se détourner de cette fonction d'usage ou mettent en place leur propre stratégie de recherche et de navigation.

Gallica ou la bibliothèque à l'écran,

par Catherine Lupovici, Directeur du Département de la bibliothèque numérique de la BNF.

A quel moment la fonction première de la bibliothèque, c'est-à-dire la mise à disposition de documents à l'aide d'outils documentaires, est-elle abandonnée au profit d'une édition plus détaillée ? Jusqu'où vont les métiers relevant de la bibliothèque et où commencent l'utilisation et le droit de l'utilisateur par rapport à cela ? C'est une vraie question et un débat permanent à l'intérieur de la bibliothèque. Doit-on parler, par exemple, pour les textes éditoriaux de Gallica, d'édition, ou simplement d'une extension du métier de bibliothécaire qui situerait sa fonction de médiateur des savoirs dans le contexte d'une "salle de lecture sur Internet" par rapport à son travail dans l'espace physique de la bibliothèque ?

On le voit, la mise en ligne génère énormément de questions. Il est essentiel de comprendre, à l'aide de l'observation des utilisateurs et par tous les moyens possibles, la façon dont le métier évolue, jusqu'où va l'offre de contenus en ligne et d'aide à l'appropriation de ces contenus.



Présidence belge de l'Union européenne
<http://www.wallonie.be>

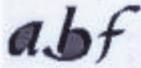
COMMISSION EUROPÉENNE, DIRECTION GÉNÉRALE EDUCATION & CULTURE
SÉMINAIRE EUROPÉEN - SAMEDI 22 SEPTEMBRE 2001
"CULTURE ET INTERNET"

Internet et culture, art et réseau : un regard transversal sur les interventions produites au séminaire.

En ouverture du séminaire "Culture & Internet" (à Mons, Belgique), le Ministre de la Culture de la Communauté française Wallonie-Bruxelles, Président du Conseil Culture de l'Union européenne, Rudy Demotte - dont on connaît l'intérêt pour le développement du secteur multimédia - a rappelé la réflexion menée à l'échelon européen sur la place de la culture, encore trop petite à ses yeux et comme à ceux de la plupart des opérateurs culturels, mais aussi sa mise en œuvre sur Internet.

L'Internet culturel : culture et numérisation, vers de nouveaux espaces publics,
par Jean-Pierre Dalbéra, chef de la Mission de la recherche et de la technologie (MRT) au Ministère de la Culture et de la Communication.

Pour la dynamique de diffusion du patrimoine culturel sur le Web il est impératif de moderniser les outils des services publics afin de faciliter l'accès à la culture mais aussi d'inciter à de nouveaux usages, également à propos de matières patrimoniales mises sur le réseau. Cela implique des coopérations internationales accrues, de nouvelles possibilités de plate-forme d'échanges. Certaines institutions ou associations culturelles ont commencé à mettre ce désir -qui est aussi un besoin- en pratique à la faveur de contenus artistiques électroniques, de coproductions et de mise en commun de savoir-faire et de technologies.



Association des bibliothécaires français
<http://www.abf.asso.fr>

CONGRÈS DU GROUPE DES BIBLIOTHÈQUES D'ART DE L'ABF
JOURNÉES DE FORMATION ET D'ÉTUDE - 14-17 MARS 2002
" RÉSEAUX, ÉCHANGES, PARTENARIATS "

Le 7^e congrès des Bibliothèques d'art de l'Abf se situe dans la continuité thématique des deux précédents congrès de Paris et de Strasbourg qui traitaient respectivement des nouvelles technologies en 1997 et d'Internet en 1999. Signe que le numérique et Internet continuent à bouleverser le monde des bibliothèques d'art et qu'ils continuent à motiver très singulièrement leurs actions, offrant notamment en matière de traitement et d'accès aux images des outils aux potentialités encore jamais égalées. Le congrès 2002 de Bordeaux progresse donc dans l'exploration de ces territoires tandis que la notion de réseau devient prépondérante dans tous les nouveaux programmes. L'idée de réseau, c'est donc aussi bien Internet en tant que réseau des réseaux que les réseaux qui le nourrissent et que les partenariats suscités par le numérique. Les échanges et les partenariats, ce sont encore les coopérations plus classiques d'associations professionnelles comparables au Groupe des bibliothèques d'art de l'Abf, elles-mêmes par une dynamique de réseau.

La Médiathèque du Centre national de la danse (CND) : un projet bâti sur une logique de travail en réseau,

par Laurent Sebillotte, Responsable de la Médiathèque du CND.

La réflexion préalable et la démarche fondatrice de la future médiathèque du Centre national de la danse (qui ouvrira en 2003 à Pantin) ont été fondées sur une logique de partenariat avec d'autres institutions spécialisées dans la danse ou appartenant au domaine de la lecture publique, dans le but de contribuer au développement d'un solide réseau documentaire national consacré à la danse. Cette volonté a conduit à inscrire au cœur du projet de la médiathèque du CND un rôle d'incitation et de soutien envers des initiatives variées menées aussi bien par des médiathèques municipales ou des centres de documentation d'établissements scolaires que par des institutions et des centres-ressources consacrés à la danse ou par des professionnels du spectacle en région. Cette logique de "réseau" est par ailleurs également présente dans la dimension internationale des activités développées par le CND, notamment par l'adhésion à des associations professionnelles internationales et par l'instauration de relations privilégiées avec des institutions étrangères analogues.

(JEUDI 14 MARS 14h00-14h30)

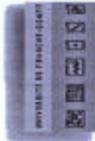


Colloque international sur le document électronique
<http://infodoc.unicaen.fr/cide/>

CIDE, COLLOQUE INTERNATIONAL SUR LE DOCUMENT ELECTRONIQUE
4-6 JUILLET 2000
"BIBLIOTHEQUES NUMERIQUES"

CIDE 2000 a eu lieu à Lyon. Co-organisé par le GREYC-Université de Caen et le PSY de l'Université de Rouen, il a plus particulièrement insisté sur la conception et l'utilisation de systèmes d'information documentaire ainsi que sur les aspects dynamiques du document. Les sujets développés ont concerné notamment le cycle de vie du document et ses aspects ergonomiques et perceptifs.

CodeX : un système pour la définition de vues multiples guidées par les usages,
par Stéphane Chaudiron, François Role et Madjid Ihadjadene.
Concilier les problématiques liées à la mise en ligne des travaux universitaires et les enjeux de la consultation des bibliothèques numériques.
(MERCREDI 5 JUILLET 16h00-16h30)



Université de Franche-Comté
<http://www.univ-fcomte.fr>

9^{ème} FORUM TICE / NTE
UNIVERSITÉ DE FRANCHE-COMTÉ - JEUDI 8 FEVRIER 2001
"L'INFORMATION SCIENTIFIQUE ET TECHNIQUE A L'HEURE DES TICE :
ACCES, RECHERCHE, EXPLOITATION"

Pelleas : une plate-forme universitaire pour la gestion des documents électroniques,

par Christian Lupovici, Directeur du Service commun de la documentation de l'Université de Marne-la-Vallée.

Contrairement au document papier, la gestion du document électronique exige la maîtrise de l'ensemble de son cycle de vie, de la conception à la conservation, en passant par la structuration, l'accès et l'exploitation sur le long terme. Le système Pelleas développé par la bibliothèque de l'Université de Marne La Vallée, prend donc en compte le document depuis sa création jusqu'à son exploitation scientifique, administrative et juridique. Il définit une architecture de campus électronique en fonction des services à offrir aux utilisateurs et en fonction du choix des normes de format de document, de structure, de métadonnées et de protocoles. Pelleas est conçu pour gérer tous les types de documents et de formats et pour gérer des flux d'information très importants.

Annexe 3

GRILLE D'ANALYSE

Modèle descriptif

QUI ?	
La source	
Nom et références	Indicateurs de renommée. Renseignent sur la légitimité de la bibliothèque numérique
Compétences	Indicateurs d'expertise. Renseignent sur la crédibilité de la bibliothèque numérique
Moyens	Indicateurs de coût. Renseignent sur la potentialité (capacité) de la bibliothèque numérique
L'environnement	
Collaborations	Indicateurs de participation. Renseignent sur l'extensibilité professionnelle de la bibliothèque numérique
QUOI ?	
La plate-forme informatique	
Architecture technique	Indicateurs techniques. Renseignent sur la fonctionnalité de la bibliothèque numérique
Le contenu documentaire	
Types de documents	Indicateurs typologiques. Renseignent sur la diversité documentaire de la bibliothèque numérique.
Taille du fonds	Indicateurs volumétriques. Renseignent sur la densité informationnelle contenue dans la bibliothèque numérique
COMMENT ?	
Approche développeurs	
Saisie de l'information	Indicateurs organisationnels. Renseignent sur la logique structurelle de la bibliothèque numérique
Traitement de l'information	Indicateurs de validation. Renseignent sur la cohérence intrinsèque de la bibliothèque numérique
Présentation de l'information	Indicateurs graphiques. Renseignent sur la logique visuelle de la bibliothèque numérique
Approche utilisateurs	
Modes de recherche documentaire	Indicateurs d'interactivité (d'orientation). Renseignent sur les possibilités d'exploitation de l'information contenue dans la bibliothèque numérique
POURQUOI ?	
Besoins exprimés	Indicateurs d'intentionnalité. Renseignent sur la causalité de la bibliothèque numérique (ni fortuite, ni gratuite)
Publics visés	Indicateurs d'adaptabilité. Renseignent sur l'étendue sociale de la bibliothèque numérique

Annexe 4

GLOSSAIRE RECAPITULATIF

Terminologie liée à XML

Les documents XML

Attribut : permet de définir les caractéristiques d'un élément, par des couples nom/valeur. Le nombre d'attributs n'est pas limité.

Ex : si l'on considère que les deux caractéristiques d'un livre sont le titre et l'auteur, l'élément sera <Livre> et les attributs <Livre titre=« La cantatrice chauve »> et <Livre auteur=« Eugène Ionesco »>

Balise (tag) : permet de délimiter des données dans les langages de balisage tels que XML ou HTML.

Élément : brique de base d'un document XML, composée :

- d'une balise de début, contenant éventuellement des attributs et leurs valeurs
- de données
- d'une balise de fin.

Ex : <Nom>Smith</Nom>

Les standards associés

CSS (Cascading Style Sheet) : feuille de style en cascade : définit la mise en page de documents XML, ou même HTML, en fixant les attributs d'affichage (police, taille, couleur...). Elles sont dites « en cascade » car elles peuvent être imbriquées.

DTD : (Document Type Definition) : établit la structure d'un document XML, c'est-à-dire l'arborescence (ordre et imbrication) et le caractère obligatoire ou facultatif des éléments qui composent un document. De plus, chaque élément est décrit. Une DTD peut être intégrée dans un document XML ou définie dans un fichier externe. Dans le deuxième cas, elle peut servir à plusieurs documents XML. Si un document XML respecte les contraintes spécifiées dans la DTD, on dit qu'il est valide.

Ex : <!ELEMENT Livre (Préface ?, (Chapitre)+> indique que chaque élément 'Livre' comporte un élément 'Préface' optionnel (signe ?) et un ou plusieurs éléments 'Chapitre' (signe +)

Schéma : tend à remplacer la DTD, à combler ses manques en définissant les types des données et les relations entre les éléments d'un document XML et ses attributs. De plus, un Schéma respecte la syntaxe XML.

XSL (eXtensible Style Language) : feuille de style extensible : langage de présentation compatible avec CSS dont il est un sur-ensemble. Une feuille de styles XSL permet de mettre en page et de reformater (ou réordonner) le document XML auquel elle est associée.

XSLT (eXtensible Style Language Transformation) : langage inclus dans XSL et permettant de transformer un document XML ayant une structure (et éventuellement une DTD) différente.

Annexe 5

GUIDE DE SITES WEB

Ressources électroniques sur le XML et les
bibliothèques numériques

RUBRIQUE	TITRE	URL	CONTENU
Information sur le XML et les standards associés			
	XML.com	http://www.xml.com	Actualités du XML : revue de presse, calendrier...
	The XML Industry Portal	http://www.xml.org	Actualités du XML : revue de presse, calendrier...
	<XML>fr	http://xmlfr.org	Actualités du XML : revue de presse, calendrier...
	XML francophone	http://www.chez.com/xml	Assistance pour le développement d'applications en XML : tutoriels, outils,...
	XML techno	http://www.xmltechno.com	Assistance pour le développement d'applications en XML : tutoriels, outils,...
	Mutu-XML	http://www.mutu-xml.org	Partage d'informations sur les réalisations actuelles du XML : savoirs, compétences..
Information sur les bibliothèques numériques	CIDE [Colloque international sur le document numérique]	http://infodoc.unicaen.fr/cide	Promotion de la recherche dans le domaine de la conception, de la modélisation, de la perception et du traitement du document numérique

				Valorisation et coordination du potentiel de recherche universitaire et privée sur le document numérique
		ISDN [Institut des sciences du document numérique]	http://isdn.enssib.fr	
Organismes professionnels				
		ABF [Association des bibliothécaires français]	http://www.abf.asso.fr	Information et formation des professionnels de l'INFODOC
		ADBS [Association des professionnels de l'information et de la documentation]	http://www.adbs.fr	Information et formation des professionnels de l'INFODOC
		GFII [Groupeement français de l'industrie de l'information]	http://www.gfii.asso.fr	Information et formation des professionnels de l'INFODOC
		W3C [World Wide Web Consortium]	http://www.w3.org	Spécification des règles pour Internet, et notamment pour le XML