HAL
open science

Analyser et valoriser un système d'information documentaire interne : vers l'utilisation de schémas de métadonnée : le cas des bases de ressources de Centre INFFO, Centre pour le développement de l'information sur la formation permanente

Stéphane Héroult

To cite this version:

Stéphane Héroult. Analyser et valoriser un système d'information documentaire interne : vers l'utilisation de schémas de métadonnée : le cas des bases de ressources de Centre INFFO, Centre pour le développement de l'information sur la formation permanente. domain_shs.info.docu. 2009. mem_00524336

HAL Id: mem_00524336

https://memsic.ccsd.cnrs.fr/mem_00524336v1

Submitted on 7 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONSERVATOIRE NATIONAL DES ARTS ET METIERS
INSTITUT NATIONAL DES TECHNIQUES DE LA DOCUMENTATION

MEMOIRE pour obtenir le
Titre professionnel "Chef de projet en ingénierie documentaire" INTD
niveau I

présenté et soutenu par

Stéphane Héroult

le 4 décembre 2009

Analyser et valoriser un système d'information documentaire interne : vers l'utilisation de schémas de métadonnées

Le cas des bases de ressources de Centre INFFO Centre pour le développement de l'information sur la formation permanente

Jury

Mme Nadia Raïs
M. Frédéric Martinoty

Promotion XXXIX

Remerciements

Je tiens à remercier chaleureusement l'ensemble des personnes rencontrées dans les unités du Centre INFFO pour leur accueil et le temps consacré à me livrer informations et explications sur leurs bases de données :

Le Département Documentation

Le Département Multimédia

La Direction Juridique Observatoire

La Mission Régions et Europe

La Mission Système d'information et Processus

Le Service informatique

Mes remerciements s'adressent également à Frédéric Martinoty, responsable du Service informatique, mon référent à Centre INFFO pour cette mission sur le Système d'Information Documentaire, ainsi qu'à Nadia Raïs, professeur associé au CNAM, pour leurs conseils avisés, leur écoute et la transmission de leurs savoirs.

Notice

HEROULT Stéphane. Analyser et valoriser un système d'information documentaire interne : vers l'utilisation de schémas de métadonnées. Le cas des bases de ressources de Centre INFFO Centre pour le développement de l'information sur la formation permanente. 2009. 208 p. Mémoire de Titre professionnel « Chef de projet en ingénierie documentaire », INTD-CNAM, 2009. (94 réf.)

L'essor des technologies de l'information, dont celles du Web, a multiplié le volume des informations et documents numériques disponibles. Une partie de ces ressources est stockée dans des bases de données, plus ou moins structurées, développées en vue de les gérer, les retrouver et les exploiter. Ce mémoire propose une démarche d'analyse du Système d'Information Documentaire (SID) d'un centre d'information sur la formation professionnelle, le Centre INFFO. L'investigation menée sur la structure de bases de production, hétérogènes, aux objets variés, ainsi qu'une enquête menée parallèlement auprès des utilisateurs/gestionnaires, au moyen d'entretiens, ont permis d'esquisser pour chaque base le modèle des données, de schématiser le workflow de production, et de formuler des constats et propositions afin d'améliorer la production. Optimiser l'exploitation et valoriser les contenus, sont les objectifs qui ont ensuite guidé la sélection de jeux de métadonnées - tels Dublin Core - et de Schémas XML associés, adaptés à chaque type de ressource. Les propositions émises tracent des perspectives d'évolution pour le SID en vue d'assurer sa flexibilité, l'enrichissement des métadonnées, l'amélioration de la structuration de l'information et l'interopérabilité.

Descripteurs :

métadonnée ; modèle de données ; base de données ; banque de données bibliographique ; hétérogénéité ; système d'information ; interopérabilité ; Dublin Core ; XML ; DTD ; valorisation ; entretien

Table des matières

UURemerciements 2
Notice 3
Table des matières 4
Introduction 10
Première partie Analyser et valoriser un système interne d'information documentaire : définition des concepts clés, présentation des enjeux 14
1 Processus et activités 15
1.1 Processus - *Process* 15
1.2 Activités 15
2 Système d'information 16
2.1 Système d'information et système informatique 16
2.2 Conception d'un système d'information : approche orientée services 17
2.3 Système d'information documentaire 21
3 Document ou Unité bibliographique ? Document électronique ou Ressource électronique ? ... 22
3.1 Document et Unité bibliographique 22
3.2 Document électronique ou Ressource électronique 22
4 Les métadonnées 24
4.1 Métadonnée : définition, fonctions, intérêts 24
4.2 Métadonnées pour les ressources numériques : types et catégories 25
5 Banques ou bases de données ? 27
5.1 Définition d'une banque de données 27
5.2 Les différents types de banques de données 27
5.3 Les avantages attendus d'une banque de données 28
5.4 Définition d'une base de données 28
5.5 Conception : de la modélisation des données à la base de données relationnelles 29
5.6 Format documentaire ou format documentaire informatique 31
6 De l'interopérabilité 32
7 L'information non-structurée : le défi de sa maîtrise et de sa gestion 35
7.1 Information non-structurée et information structurée 35
7.2 Maîtrise et gestion de l'information non-structurée : un grand défi managérial actuel 36
8 Norme ou standard ? 37
8.1 Norme 37
8.2 *Standard* - Standard 37

9 Schémas, référentiels, registres, jeux de métadonnées, profils d'application 39
10 XML - *eXtensible Markup Language*, recommandation du W3C pour une nouvelle modélisation des données 42
10.1 Le XML, *eXtensible Markup Language*, langage à balises extensible 42
10.2 Le document XML : un document structuré 44
10.3 La modélisation XML au service de la conception du système d'information 46
10.4 Recommandations XML complémentaires 47
11 Déclarations de structure XML : utilisation de DTD ou de Schémas XML 49
11.1 DTD – Définition de Type de Document 49
11.2 XML-Schema ou XSD – *XML Schema Definition* 50
11.3 Caractéristiques d'un document XML 51
12 Manipulation, traitement de documents XML 52
12.1 Manipulation des documents XML : solutions techniques 52
12.2 Autres standards XML : langages et outils pour la présentation, la transformation, la gestion des liens, le nommage des éléments et attributs 52
13 Technologie XML et documents XML : avantages, usages, contraintes 56
14 Entrepôt de (méta)données 58
Deuxième partie Le cas de Centre INFFO : association productrice d'informations et gestionnaire de données sur la formation professionnelle 59
1 Le contexte de l'étude 60
1.1 Le Centre INFFO : statut, missions, activités 60
1.2 Les orientations technologiques au Centre INFFO 62
1.2.1 Schéma Directeur Nouvelles Technologies 2008-2011 62
1.2.2 Un développement croissant des activités Web 62
1.3 Réseau informatique interne et Intranet 63
1.3.1 Services accessibles sur le réseau interne 63
1.3.2 Intranet 63
1.3.3 Intranet documentaire 64
2 Etude des modèles de données gérées dans les bases du Centre INFFO et processus métiers : problématiques majeures et objectifs visés, méthode utilisée, résultats attendus 65
2.1 Problématiques majeures et objectifs visés 65
2.1.1 La gestion d'informations dans des bases... qui sont à redécouvrir 65
2.1.2 Enquêter sur des bases, analyser un système, valoriser les ressources 65
2.2 Cheminement dans l'investigation : point méthodologique 67
2.3 Rapporter des constats, des besoins ; formuler des idées, des hypothèses ou des propositions 68
Troisième partie Investigation des bases de données de production de Centre INFFO 70
1 Recensement des Bases de données de production : documentaires et autres 71

2 Base de données LORIS 75
2.1 Présentation de la base LORIS 75
2.2 Références bibliographiques : ouvrages et articles de périodiques 76
2.2.1 Revues publiées par le Centre INFFO 76
2.2.2 Accroissement du fonds documentaire : ouvrages et articles de périodiques 77
2.2.3 Constats sur le fonds bibliographique (hors textes conventionnels) 77
2.2.4 Workflow relatif au traitement du fonds bibliographique (hors textes conventionnels) 79
2.3 Fonds juridique 79
2.3.1 Textes législatifs et réglementaires - Jurisprudence 79
2.3.2 Textes conventionnels dont les accords de branches 80
2.3.3 Workflow relatif au traitement documentaire des textes conventionnels 81
2.4 Structure de la base LORIS 82
2.4.1 Format documentaire informatique : format UNIMARC 82
2.4.2 Structure de la base LORIS 83
2.4.3 Liste des tables principales dans LORIS 84
2.4.4 Le Thésaurus de la formation 85
3 Base de données sur les Produits pédagogiques multimédias 86
3.1 Présentation de la Base Produits pédagogiques multimédias – Constats 86
3.1.1 Présentation sur le site Web du Centre INFFO 87
3.1.2 Domaines 87
3.1.3 Supports de diffusion 88
3.1.4 Evaluations 88
3.1.5 Consultations 89
3.1.6 Constats et réflexions 90
3.2 Structure de la base Produits pédagogiques multimédias : tables, modèle 91
3.3 Workflow du traitement documentaire des produits pédagogiques multimédias 93
4 Application NOEL – Nouvelle Offre en Ligne 94
4.1 Présentation de NOEL application de gestion de l'offre de formation 94
4.2 Structure des fichiers STAGES et ORGANISMES, constitutifs de NOEL 95
4.3 Modèle des données de NOEL 98
4.4 Workflow du traitement de l'offre de formation avec NOEL 99
5 Application de gestion du Thésaurus de l'offre de formation : Formacode 100
5.1 Présentation du Thésaurus Formacode – Constats 100
5.2 Structure du fichier FORMACODE 102
5.3 Modèle des données du fichier Formacode 104
5.4 Workflow relatif au Formacode 105
6 Base de données FAR – Formation et Apprentissage en Régions 106

6.1 Présentation de la base FAR 106
6.2 Constats sur FAR 107
6.3 Structure de FAR 108
6.4 Modèle des données de FAR 109
6.5 Workflow du traitement des ressources de FAR 109
7 Base Annuaire des structures et contacts 110
7.1 Présentation de la base Annuaire des structures et contacts 110
7.2 Structure de la base Annuaire des structures et contacts 111
7.3 Modèle des données de la base Annuaire des structures et contacts 112
7.4 Workflow relatif à la base Annuaire des structures et contacts 114
Quatrième partie Pour une évolution des modèles de données : optimiser la description des ressources et leur exploitation 115
1 Normes, standards et schémas de métadonnées : de nouvelles modélisations à envisager pour la description des ressources gérées au Centre INFFO 116
1.1 Norme Dublin Core Metadata 117
1.2 Schémas de description – Modèles applicatifs – Profils d'application basés sur XML : des schémas de métadonnées ouverts autorisant une exploitation variée des ressources 120
1.2.1 Schéma XML DC simple ou qualifié pour les « Fiches pratiques de la formation continue » 120
1.2.2 Norme LOM pour la description de ressources pédagogiques 121
1.2.2.1 LOM – Learning Object Metadata 121
1.2.2.2 LOM-FR, profil d'application français 123
1.2.2.3 Sup-LOMFR 124
1.2.3 Standards pour la description de l'offre de formation 125
1.2.3.1 LHÉO – Langage Harmonisé d'Échange d'informations sur l'Offre de formation 125
1.2.3.2 CDM – Course Description Metadata et CDMFR 126
1.2.3.3 MLO - Metadata for Learning Opportunities 127
1.2.4 Modèle pour les références bibliographiques : BiblioML 128
1.2.4.1 Les inconvénients et lacunes des formats MARC à l'heure du Web 128
1.2.4.2 Les successeurs de MARC : enrichir les liens entre éléments, faciliter le partage et l'échanges de données, accroître la visibilité de fonds bibliographiques 129
1.2.4.3 Focus sur les avantages du format BiblioML 130
1.2.4.4 Avenir du catalogue de la bibliothèque du Centre INFFO 130
1.2.5 Modèle pour les actualités et articles journalistiques : NewsML 131
1.2.6 SKOS - Simple Knowledge Organization System, langage de représentation pour les thésaurus 132
2 Des métadonnées pour la représentation et l'accès aux informations gérées dans les bases : perspectives, réflexions et questions 133

2.1 Perspectives pour le Système d'information documentaire (SID) : de la situation actuelle à la situation future 133
2.1.1 Etats du Système d'information du Centre INFFO, aujourd'hui et demain 133
2.1.2 Concilier plusieurs jeux de métadonnées 134
2.1.3 Processus de production, modèles de données, passerelles et convergence 135
2.1.4 Architecture du SID : distinguer production et recherche 137
2.2 Hypothèses au regard de la valorisation des données 138
2.2.1 Une collection de ressources complémentaires et enrichies à penser dans son contexte 138
2.2.2 De l'intérêt d'enrichir les métadonnées des ressources numériques internes 139
2.2.3 Les données, leur sens, leur filiation, leur complémentarité : application à Centre INFFO 140
2.2.4 Choisir des Schémas XML et suivre des normes et recommandations 141
2.3 Des solutions techniques à explorer pour développer le système et les services : propositions et remarques 142
2.3.1 Adopter la technologie XML : avantages et contraintes 142
2.3.2 XML et bases de données 143
2.3.3 Des dispositifs pour la manipulation des données du Web 144
2.3.4 Intérêts de la finesse de structuration : pour l'exploitation des données et l'optimisation des résultats de recherche 145
2.3.5 Avenir d'XML : le Web sémantique, en cours de développement 146
2.3.6 Proposer des ressources documentaires sous forme d'archives ouvertes 148
2.3.7 Se doter d'outils collectifs ou individuels pour la gestion de l'information non structurée 148
2.3.8 Créer des services (documentaires) à valeur ajoutée 149
2.4 Investigation menée sur le système d'information documentaire : problèmes soulevés et ajustements à apporter 150
2.4.1 Dresser le portrait du SID : mieux le connaître pour le faire évoluer 150
2.4.2 Connaître le workflow 150
2.4.3 Avoir des outils de suivi des bases et modèles 151
2.4.4 Appréhender les bases et leur structure 151
2.4.5 Sur la méthode d'analyse et d'enquête 152
2.4.6 Aspects méthodologiques de la conduite de projet 153
2.4.7 Etre attentif au(x) facteur(s) humain(s) : un volet à approfondir 153
Conclusion 154
Bibliographie 159
Annexes 188
Annexe 1 Sigles et acronymes relatifs aux technologies 189

Annexe 2 Organisation fonctionnelle du Centre INFFO 191
Annexe 3 Cartographie macro du Système d'Information du Centre INFFO 192
Annexe 4 Formats UNIMARC et ISBD : exemples de références extraites de la base LORIS 195
Annexe 5 Exemples d'utilisation des données pour la production de dossiers documentaires 200
Annexe 6 Site interne Webdoc sous logiciel SPIP 202
Annexe 7 Base Juriweb : production des « Fiches pratiques de la formation continue » 206
Annexe 8 Guide d'entretien pour l'enquête menée auprès des utilisateurs internes des BDD 208

Introduction

Le système d'information, ensemble structuré des informations et des fonctions utilisées par les activités de l'entreprise, mobilise les acteurs de différentes professions pour son organisation et sa gestion. Dans ce cadre, la documentation met en œuvre un ensemble de méthodes et techniques de traitement systématique de documents ou d'informations, quel que soit leur support, en vue de répondre aux besoins des usagers. Cet ensemble inclut l'acquisition, le signalement, l'analyse documentaire, l'indexation, le stockage, la recherche, la diffusion de ces documents ou informations.
La mission première d'un centre de documentation est donc de rendre l'information accessible, et permettre d'assurer la recherche et la diffusion d'une information pertinente, complète et fiable, au service de ses utilisateurs. Un système documentaire est l'ensemble des fonctions remplies par un centre de documentation en vue de répondre à ces objectifs.

L'objet d'information traditionnel était géographiquement situé et conçu selon une chaîne de production et de diffusion spécifique.
L'essor des technologies de l'information, et parmi celles-ci les techniques du Web, le développement rapide de l'information numérique, ont profondément modifié l'économie et le cycle de vie du document. Les données et les documents numériques disponibles se sont considérablement accrus. Le volume des informations a été démultiplié.
Une partie de ces informations est stockée dans des bases de données, plus ou moins structurées, que l'on a peu à peu développées en vue de les gérer, de les retrouver et de les exploiter. Mais la masse des informations variées recensées, la diversité des applications générées, la variété des structures de données, n'en facilitent aujourd'hui pas l'exploitation, chaque base ayant de surcroît sa propre interface de recherche, son propre langage d'interrogation, sa propre interface de présentation des résultats.
L'enjeu est bien aujourd'hui de faciliter l'accès centralisé à des données hétérogènes mais complémentaires. Nécessité est ressentie de centraliser l'accès aux données, de mettre en relation ces ressources, de faciliter les requêtes multibases et d'en présenter les résultats, d'améliorer l'exploitation des informations gérées et des résultats de requêtes, de valoriser davantage les fonds documentaires, par le recoupement des données disponibles et par la variété de la représentation de ces informations sous différents formats, pour différents usages, et une lecture sur des plateformes diverses.

En effet, la masse exponentielle de ressources numériques multiformes et permettant des parcours de lecture divers offre une opportunité extraordinaire d'accès, sur Internet ou à l'intérieur d'un système d'information. Mais c'est aussi une opportunité d'évolution des pratiques de production et de mise à disposition de l'information : intégration de ressources préexistantes, liens entre les ressources, activités collaboratives… Les évolutions technologiques créent de nouveaux besoins de traitement, et le système d'information pour y répondre met en œuvre un ensemble de pratiques de « curation digitale ». La curation digitale, qui vise à maintenir et ajouter de la valeur à un corpus fiable d'information électronique pour les utilisations actuelles et futures, de la création des données à leur archivage en passant par leur diffusion, est un processus qui repose sur un élément central : les métadonnées.
Ces dernières, ensembles de données structurées décrivant des ressources physiques ou numériques, sont un maillon essentiel pour l'interopérabilité de l'information et sa gestion. Elles ne constituent pas seulement des données descriptives ou d'informations d'exploitation, elles peuvent de manière effective, formalisées dans des standards, promouvoir l'interopérabilité par la « réutilisabilité » d'ensembles de données communes.

Ce mémoire propose de s'intéresser au cas d'un Système d'Information caractérisé par une hétérogénéité en termes de données et de solutions de paramétrage. Cette diversité se situe dans les systèmes de stockage (bases, systèmes de gestion de contenu, fichiers), les formats de données (tables MySQL, fichiers propriétaires) et les solutions utilisées pour gérer les différents types de données. Cette hétérogénéité a pour conséquence de rendre lourde et complexe, voire coûteuse, la mise en œuvre et l'exploitation de ces données par les applications de l'entreprise. Et utiliser cet ensemble d'applications différentes afin de pouvoir gérer cette diversité dans les types de données peut aussi entraîner une redondance tant dans les données que les outils.

L'étude commence par une réflexion sur les concepts clés qui sont nécessaires à la compréhension des thèmes afférents à l'analyse du Système d'Information Documentaire (SID) et sa valorisation. Les définitions sélectionnées clarifient les notions, la façon dont elles s'enchaînent et se complètent, telles que nous les concevons dans la suite du mémoire, constituant ainsi une base commune d'échange entre les corps de métiers amenés à intervenir dans le système.
La seconde partie présente l'entreprise contexte de l'étude, le Centre INFFO - Centre pour le Développement de l'Information sur la Formation permanente, les problématiques majeures auxquelles est confronté son SID composé de bases nombreuses. Puis la méthode utilisée pour son analyse, visant particulièrement une amélioration de l'interopérabilité, entre éléments techniques du système et entre les ressources, et une valorisation des données.

Nous nous intéressons à un ensemble de bases de données de production, en procédant d'une part à une investigation de leur structure, de l'organisation et de la nature des données et métadonnées gérées, et d'autre part en menant une série d'entretiens avec les utilisateurs/gestionnaires pour appréhender la production et le flux d'information de l'entrée à la sortie du système. C'est l'objet de la troisième partie du mémoire, synthétisant les informations rapportées, proposant une modélisation des structures de quelques bases ainsi qu'une schématisation des processus.
La quatrième propose d'optimiser la description des ressources documentaires et leur exploitation, décline à cet effet des modèles de données, normés ou standardisés sous forme de Schémas XML, puis en tire des conclusions pour envisager une évolution du SID et une valorisation de ses ressources.

Une bibliographie analytique proposée en fin de mémoire permet d'illustrer et d'approfondir les thèmes traités, c'est pourquoi tout au long des chapitres figurent des numéros entre crochets renvoyant vers ces références complémentaires, pour un grand nombre consultables sur le Web.

Première partie

Analyser et valoriser un système interne d’information documentaire : définition des concepts clés, présentation des enjeux

1 Processus et activités

1.1 Processus - *Process*

« Ensemble d'**activités** corrélées ou interactives qui transforme des éléments d'entrée en éléments de sortie. (NF EN ISO 9000) Les éléments d'**entrée** *[input]* d'un processus sont généralement les éléments de **sortie** *[output]* d'autres processus. Les processus d'un organisme sont généralement planifiés et mis en œuvre dans des conditions maîtrisées afin d'apporter une valeur ajoutée… ».

« Les normes NF EN ISO 9000 encouragent l'adoption d'une « approche processus » pour le management de la qualité. Elles expriment que toute activité d'un organisme peut être considérée comme un processus et exigent (NF EN ISO 9001) ou recommandent (NF EN ISO 9004) d'identifier systématiquement les processus nécessaires au système de management de la qualité. Dans un but d'efficacité de son fonctionnement, l'organisme doit déterminer la **séquence**[1] et les **interfaces** entre ces processus, assurer leur maîtrise, leur surveillance et leur amélioration. » [81]

1.2 Activités

En entreprise, plusieurs termes sont utilisés dans des sens voisins selon le contexte [81], que l'on peut présenter de la façon suivante :

- domaine d'activité, recouvre plusieurs métiers ;
- métier, recouvre plusieurs activités ;
- activité, recouvre plusieurs actions ;
- action, recouvre plusieurs tâches ;
- tâche, recouvre plusieurs opérations ;
- opération, est une partie élémentaire d'un mode opératoire.

[1] Ce mémoire n'a pas pour but d'explorer systématiquement l'ensemble de ces « **séquence** » et « **interfaces** » entre les processus à l'œuvre dans le système d'information, mais de comprendre les **activités** de production de bases de données qui sont effectuées dans le cadre de processus.

2 Système d'information

2.1 Système d'information et système informatique

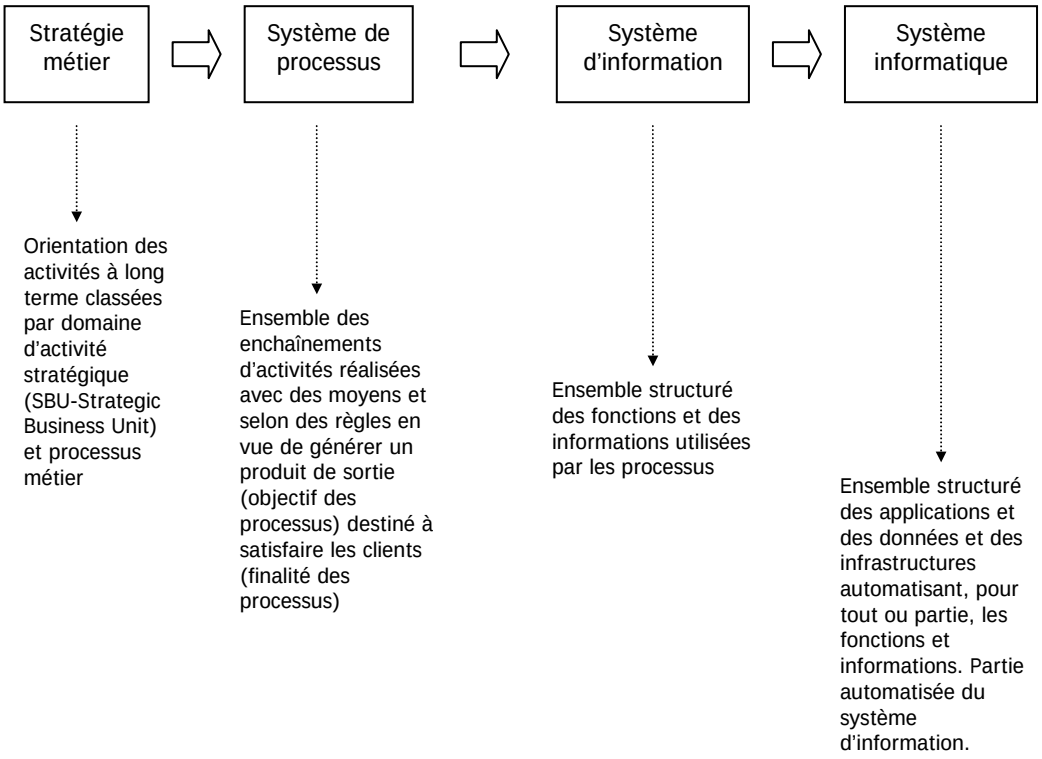
Un **système d'information** est « l'ensemble structuré des informations et des fonctions utilisées par les activités d'une entreprise ou d'un de ses segments stratégiques ». [76]

Dans nombre d'ouvrages, et dans le langage courant, on parle volontiers de système d'information pour regrouper deux notions : le **système d'information** (contenu) et le **système informatique** (contenant). « Pour dérouler toute activité ou tout processus au sein d'une entreprise », on doit traiter des informations (flux informationnels) dont la gestion est automatisée ou pas. La problématique du système informatique est plutôt technique : automatiser le traitement de l'information, alors que la problématique du système d'information est quant à elle plus stratégique. Ces deux notions sont distinctes, mais complémentaires.

« Le système d'information est composé d'un ensemble de modèles sémantiques et conceptuels (modèle métier, de données, de traitement, etc.) et de normes qui définissent comment ces concepts interagissent (règles d'urbanisme) ». [76]

La construction du système d'information est un maillon de la **chaîne d'alignement stratégique** décrite par le schéma[2] qui suit.

[2] Reproduction d'un schéma du guide [76].



La définition d'un autre auteur, J.-J. Thomasson, précise quant à elle : « Un **système d'information** représente l'ensemble des moyens humains et techniques mis en œuvre pour gérer les informations nécessaires au fonctionnement d'une entreprise. Il couvre les grandes fonctions de création, archivage, transformation et diffusion de l'information. À ce titre, il fait appel aux **techniques informatiques** de gestion des données, des traitements et des communications, ainsi qu'aux méthodes de mise en œuvre de ces techniques. »[3] [35]

2.2 Conception d'un système d'information : approche orientée services

Ce sous-chapitre n'a ni pour ambition de détailler l'**architecture d'un système d'information**, qui vise à en décrire de manière symbolique et schématique les différents composants, leurs interrelations et interactions, ni décliner tous les styles d'architecture qui, à ce jour, ont été élaborés.

---

[3] Extrait du manuel de Jean-Jacques THOMASSON et Antoine LONJON [35], « Modélisation XML », publié en version numérique en janvier 2006 : <http://www.eyrolles.com/Chapitres/9782212115215/Glossaire_Thomasson.pdf>.

Il privilégie plutôt une approche qui aide à se représenter ce qui est attendu de tout système d'information, et rapporte une méthode[4] séparant les traitements des données sur lesquelles ils portent.

Les auteurs J.-J. Thomasson et A. Lonjon [35] proposent une méthode permettant d'**esquisser et de concevoir un système d'information**, selon une **approche orientée services**, c'est-à-dire qui respecte une **architecture orientée service – SOA - *Service Oriented Architecture*.**

Cette méthode revient, en fait, à **analyser le processus de production d'informations et de documents**. [36]

Elle se distingue des méthodes fondées sur **UML - *Unified Modeling Language*** (« langage de modélisation unifié »), qui appliquent les principes de la **conception orientée objet**, où les données et les traitements sont encapsulées par des classes. L'accent mis sur les classes contribue alors implicitement « à privilégier l'analyse des traitements plutôt que celle des données échangées » [5]. [35]

Dans les **architectures basées sur les services**, la coordination entre les différentes unités de traitement, appelées « **services** », est réalisée à l'aide d'échanges d'information. L'autonomie des données échangées est privilégiée, garante de plus de **flexibilité** et de **pérennité**. En d'autres termes, est garantie la séparation des étapes de production des données, de celles d'adaptation et de mise en valeur (mise en forme, publication, indexation).

La méthode avancée par J.-J. Thomasson et A. Lonjon comprend **quatre principes de base** qui, une fois mis en œuvre, permettent de définir fonctions, sous-fonctions et services d'un système d'information[6]. [35]

- **Principe n°1**

Un système d'information repose toujours sur un même modèle de base, invariant.

L'idée première est que l'information est une matière vivante, ayant un début, une vie (croissance) et une fin. Ce **cycle de vie**, simple à représenter, propose également la représentation la plus élémentaire d'un « futur système d'information ».

---

[4] Méthode dont la mise en œuvre des étapes peut être utilisée dans le cas d'une démarche de modélisation XML - cf. chapitre 10 de cette Première partie du mémoire.

[5] J.-J. Thomasson et A. Lonjon, mettant en valeur la séparation des données et des traitements, ne considèrent donc pas l'approche dite par objets métier comme bonne, et lui opposent celle orientée service, « plus conforme aux fondamentaux de XML ».

[6] Eléments extraits du chapitre 1 du manuel de Jean-Jacques THOMASSON et Antoine LONJON [35], « Modélisation XML », publié en version numérique en janvier 2006 : consultable à l'adresse <http://www.eyrolles.com/Chapitres/9782212115215/chap1_Thomasson.pdf>.

PRODUCTION	GESTION	DIFFUSION

« **Production** » marque le début d'une information (création, récupération, extraction de granules d'information initiaux)

« **Gestion** » caractérise sa croissance (assemblage, fabrication d'une information finie à partir des granules d'information)

« **Diffusion** » marque la fin d'une information, car « dès qu'une information est diffusée, elle échappe alors au système qui l'a vu naître et passe dans un autre monde » (publication).

- **Principe n°2**

Un système d'information est un ensemble de **fonctions** et **services**.

Au sein d'un système d'information, le composant élémentaire est « un **service** capable de recevoir une information, la transformer et la transmettre à son tour à un autre composant élémentaire ». Tout composant élémentaire est constitué d'au moins deux fonctions : l'une sert à transmettre, l'autre à effectuer un traitement. Le socle précédemment représenté permet d'identifier les **fonctions** du système, de placer et empiler chaque fonction au niveau lui convenant, et de « construire l'ensemble en pensant que chaque fonction doit rendre un service potentiellement indépendant de tous les autres ».

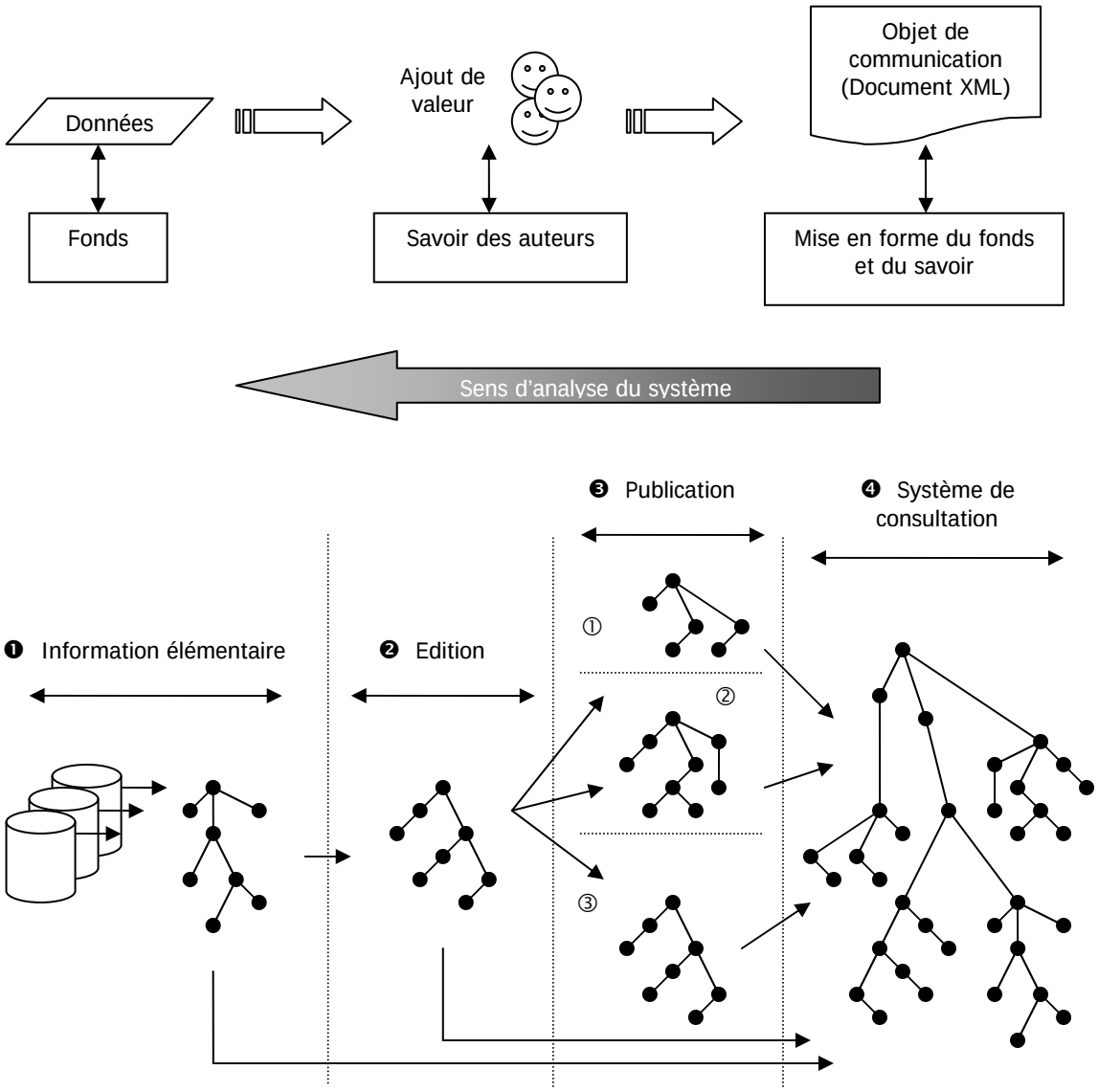
- **Principe n°3**

Un système d'information est obligatoirement **flexible**, information et rigidité étant antinomiques.

Les données doivent donc être organisées de manière à assurer au système une grande flexibilité dans le temps. Les informations doivent être articulées et transportables pour pouvoir être (ré)utilisées à plusieurs fins, sur plusieurs médias, avec différents traitements de transformation.

Identification des sous-systèmes de la chaîne de production d'information :

« l'information doit être articulée et transportable »



Ce schéma[7] met en parallèle l'**organisation humaine** et les **modèles de données** utilisés. Il résume plusieurs idées relatives à la **flexibilité** : le système doit s'adapter (en permanence) à l'évolution des organisations humaines, et aux besoins des consommateurs de l'information et de l'évolution du marché qu'ils représentent. Le système doit permettre en l'adaptant, de réutiliser une même information, ou granule d'information, ou donnée élémentaire. L'**analyse du système**, du **processus de production** se fait en partant du point « le plus stable et le plus contraint » de la chaîne (en termes de contenu, forme, lisibilité, média de diffusion), l'objet de communication qui est produit en bout de chaîne, et

[7] Reproduction d'un schéma paru dans les manuels [35] <http://www.eyrolles.com/Chapitres/9782212115215/chap1_Thomasson.pdf> et [36].

en remontant vers la source des données. S'arrêter sur chaque **état du document** autorise à déterminer les conséquences sur le **modèle de données**[8]. Les zones repérées ❶ ❷ ❸ ❹ correspondent à différents sous-systèmes de production, chacun d'eux pouvant à son tour être également découpé en sous-systèmes ① ② ③. Chaque sous-système est caractérisé par un ensemble de fonctions ainsi que par l'état des données en entrée et en sortie.

- **Principe n°4**

La **représentation du système d'information** doit être synthétique et simple ; afin de mieux appréhender la complexité du problème.

La représentation graphique finale du système doit reposer sur des règles simples qui facilitent la comparaison de différentes solutions, de différents scénarios d'implémentation. Le système va être progressivement découpé en services (ensembles de fonctions de traitement) et échanges de données entre services ; et un système d'attribution de points va permettre de noter chaque fonction et chaque service, et d'évaluer *in fine* le coût de réalisation du projet.

2.3 Système d'information documentaire

Un **système d'information documentaire** – ou **SID** – est, comme son intitulé l'indique, un type spécifique de système d'information. Comme tout système d'information, il s'agit bien d'un « ensemble d'éléments en interaction et formant un tout organisé et cohérent, mis en œuvre pour gérer, stocker et permettre l'accès à l'information ». Ces éléments sont définis tant au niveau des politiques que des procédures, des ressources matérielles et humaines. [78]

Mais un **SID** présente des fonctionnalités particulières. Il assure la gestion et la communication des **informations contenues dans des documents**, et pas seulement ceux produits par l'organisation dont il constitue une unité stratégique. « Cette gestion implique le repérage et la validation des sources et ressources d'information, la collecte, le traitement, le stockage, la diffusion, ainsi que les moyens de classement et de récupération des **documents porteurs d'information** », **internes** et **externes**, en utilisant des outils classiques, ou aujourd'hui plus systématiquement des outils informatisés.

[8] Ou le **modèle de document** : cette figure peut constituer une représentation schématique de l'évolution des structures internes d'un document XML [36], telle qu'elle sera évoquée au chapitre 12 de la Première partie de ce mémoire.

3 Document ou Unité bibliographique ? Document électronique ou Ressource électronique ?

Un système d'information documentaire a donc pour objet **les informations contenues dans des documents**. A l'heure du Web, la notion de **document,** ni même plus celle de **document électronique,** ne sont cependant désormais suffisantes pour rendre compte des **types de données** que le SID gère ou est amené à gérer. Et les évolutions technologiques modifient et élargissent aussi les publics ciblés, tout comme les publics potentiels, et les usages de ces « données ». Voici quelques définitions, notamment empruntées au « Vocabulaire de la documentation ». [78]

3.1 Document et Unité bibliographique

Un **document** est « l'ensemble d'un support d'information, quel qu'il soit, des données enregistrées sur ce support et de leur signification servant à la consultation, l'étude, la preuve ou la trace, etc. : livre, échantillon de parfum, tissu, film, etc. Le tout constitue une unité autonome ».[9] [78]

Une **unité bibliographique** est « un document, une collection ou une partie de document, quel qu'en soit le support, formant une unité indépendante d'un point de vue bibliographique. Exemples : une série photographique, les trois volumes d'une monographie, un numéro spécial de revue, un chapitre d'un ouvrage, etc. »[10] [78]

3.2 Document électronique ou Ressource électronique

Un **document électronique** est « un document sur support électronique, au sens d'un fichier de données et/ou de programmes, nécessitant un matériel pour être consulté. Il bénéficie des possibilités de l'informatique : reproductibilité (sur plusieurs écrans en même temps), modification selon le type de code, transmissibilité via les réseaux de communication. On distingue trois modes principaux de représentation de l'information de ces documents électroniques : mode caractère, mode image, mode vectoriel. »[11] [78]

On entend par **ressource électronique**, « un document (données ou logiciels) encodé afin d'être traité par un ordinateur et considéré comme une unité bibliographique. Les ressources

---

[9] Consultable à l'adresse <http://www.adbs.fr/document-16827.htm>.
[10] Consultable à l'adresse <http://www.adbs.fr/unite-bibliographique-18988.htm>.
[11] Consultable à l'adresse <http://www.adbs.fr/document-electronique-16869.htm>.

électroniques comprennent d'une part des ressources d'information stockées en local, d'autre part celles qui nécessitent l'utilisation d'un périphérique relié directement à l'ordinateur (par exemple, un disque dur, un lecteur de cédérom), et enfin les services en ligne (par exemple, les forums ou les listes de discussions, des sites Web). Une ressource électronique peut comporter soit du texte, soit de l'image fixe ou animée, soit du son. Elle peut être aussi multimédia. »[12] [78]

La notion même de **document** et la **typologie des documents** ont évolué. On parle désormais aujourd'hui dans le monde du document électronique en ligne, avec la prépondérance du Web et de ses services transitant via le réseau Internet, de **ressource** : « celle-ci devient un concept générique désignant un ensemble de données pouvant être utilisé comme un tout et relativement à un contexte défini d'utilisation ». Le terme de ressource s'applique aussi bien à un **site Web** dans son ensemble, qu'à une **page Web**, ou encore, conséquence de l'évolution du passage du codage HMTL au codage XML, à une **portion de page**, lorsque ce fragment est autosuffisant par rapport à une utilisation particulière.

Et avec l'expansion des ressources électroniques d'information, là où classiquement en documentation on utilisait le terme d'**informations secondaires** pour évoquer les données de description des documents primaires, gérées dans des **bases de données** bibliographiques (BDD), on utilise le concept de **métadonnées**. Les métadonnées font l'objet du chapitre 4 qui suit, tandis que les BDD sont étudiées au chapitre d'après. [20]

[12] Consultable à l'adresse <http://www.adbs.fr/ressource-electronique-18512.htm>.

4 Les métadonnées

4.1 Métadonnée : définition, fonctions, intérêts

Une **métadonnée** vient du terme Grec "**meta**", qui désigne ce qui dépasse, englobe. C'est donc une donnée à propos d'une autre donnée.
La racine substantive « **donnée** », elle, apparaît comme directement liée au terme anglais *data*. « Une donnée apparaît donc comme une valeur, généralement informatique, considérée comme stable au regard du processus pris en considération ». Du point de vue du domaine information-documentation, on peut considérer que « la donnée est davantage informationnelle qu'informatique, et donc qu'elle est davantage qualifiée par le sens qu'elle véhicule que par la forme de son codage ». Une même information pourra « avoir des codages informatiques différents selon le système utilisé, alors que du point de vue info-documentaire il s'agit toujours de la même donnée ». [22]

Selon le NISO – *National Information Standards Organization* [21], les **métadonnées** sont, en sciences de l'information, des ensembles de données structurées décrivant des ressources physiques ou numériques, ou, sur un plan plus fonctionnel, "de l'information structurée qui décrit, explique, localise la ressource et en facilite la recherche, l'usage et la gestion".
Il s'agit donc d'un ensemble structuré d'informations décrivant une ressource quelconque. [67].
On parle aussi de « **méta-information** », une information sur l'information, qui est représentée par les métadonnées. [72]

Si le terme « métadonnées » est « utilisé depuis longtemps dans certains domaines ou formats de ressources (données géospatiales, images, audiovisuel), il n'est apparu que depuis une dizaine d'années dans les métiers de l'information et de la conservation, où d'autres termes correspondent à des concepts proches, catalogage et signalement par exemple. Aujourd'hui, cependant, les spécificités du document électronique en matière de cycle de vie, de gestion, de droits d'usage et de recherche d'information et les besoins d'interopérabilité entre ressources font des métadonnées un élément incontournable dans le paysage numérique[13]. [20]

[13] Voir aussi une fiche, proposée par l'Université Paris V René Descartes, qui détaille tous les intérêts que présentent les métadonnées <http://wiki.univ-

Arlette Boulogne dans son ouvrage « Vocabulaire de la documentation » publié par l'ADBS, l'Association des professionnels de l'information et de la documentation, précise ainsi les **fonctions** que peuvent remplir les métadonnées[14] au regard des ressources électroniques :

a) gestion des ressources décrites (suivi du cycle de vie : création, modification, archivage) ;
b) informations sur le contenu de la ressource pour en faciliter la découverte, la localisation, l'accès ;
c) suivi de l'utilisation et du respect des droits et conditions d'utilisation associés à la ressource.

Elles peuvent être exprimées dans le même format technique de codage que celui de la ressource qu'elles accompagnent et être disponibles en même temps qu'elle. [78]
Une autre définition complète cette approche en précisant qu'une métadonnée consiste en un ensemble structuré d'informations décrivant la forme, le contenu ou la localisation d'une ressource. En fournissant un contexte et des balises uniformes et stables, les métadonnées permettent d'accéder facilement à l'information, de l'extraire et de la comprendre. [79]

Les métadonnées peuvent être écrites actuellement selon plusieurs standards ouverts : RDF - *Resource Description Framework*, TEI - *Text Encoding Initiative*, syntaxe " meta " HTML et Dublin Core, DTD EAD - *Encoding Archival Description*, etc. [78]
Elles peuvent alors aider à organiser des ressources électroniques ou à favoriser leur interopérabilité. Ces métadonnées peuvent ne pas apparaître à l'écran, mais servent à donner des indications aux outils de recherche. [79]

4.2 Métadonnées pour les ressources numériques : types et catégories

En résumé, considérant les ressources numériques, on distingue **quatre types de métadonnées** [22] [62] que l'on peut regrouper en deux grandes classes, selon leur nature :

- **externes** :
  - de type catalogues ou bases de données externes aux ressources, utilisées séparément pour la recherche (la liaison avec le document est assurée par une métadonnée de localisation) ;
  - externes à la ressource mais fournies en même temps

---

paris5.fr/wiki/M%C3%A9tadonn%C3%A9e>. De même <http://bibliodoc.francophonie.org/article.php3?id_article=172> [15].
[14] Définition également disponible à l'adresse <http://www.adbs.fr/metadonnees-17808.htm> [consulté le 6 septembre 2009].

- **internes** :
  - encapsulées, fournies dans la ressource (cas du Dublin Core)
  - englobantes[15], pouvant inclure la ressource elle-même (entièrement ou partiellement)

Trois grandes **catégories de métadonnées**, peuvent aussi caractériser leur objet :

- **Descriptives** : visant à localiser et qualifier les ressources à la fois sur le contenant (ex : description bibliographique, identification) et sur le contenu « intellectuel » (ex : analyse par résumé et indexation) ;
- **Administratives** : facilitant la gestion et le traitement, l'inscription des droits d'usage et d'accès attachés au document, la conservation des collections numériques ;
- **Structurelles** : caractéristiques physiques et informatiques décrivant une ressource électronique pour en faciliter la navigation et la présentation ; ces *metadata* placent la ressource dans son contexte en explicitant les relations avec d'autre objets, en précisant des modalités de liaison ou d'appartenance (fait partie de, comprend, est en relation avec, indications de collection, etc.).

Ces catégories peuvent être générées automatiquement ou de manière manuelle, constituées de contenu structuré ou non, et enfin être statiques ou évolutives. [15] [18] [21] [22] [65]

[15] Cas des modèles basés sur EAD – *Encoded Archival Description* utilisés en archivage, non retenus et traités dans ce mémoire.

5 Banques ou bases de données ?

L'objet de ce mémoire est de mener une étude de bases de données recensant des références bibliographiques, des textes officiels ou des ressources pédagogiques, répertoriant une offre de formation (stages, organismes, diplômes), ou encore de bases contenant du texte intégral assurant la gestion d'informations ou d'articles diffusés sur le Web (articles journalistiques, fiches juridiques,...).

5.1 Définition d'une banque de données

Voyons dans un premier temps ce qu'est une ***banque de données***.
« Une Banque de données est un ensemble de données relatif à un domaine défini de connaissance organisé pour être offert aux utilisateurs » (cf. Journal officiel du 17 janvier 1982).
Cette définition indique les deux aspects complémentaires de la banque de données : d'une part la constitution de celle-ci et d'autre part sa structuration pour l'utilisation. [80]
L'expression ***« base de données »*** est utilisée en informatique pour décrire les systèmes de gestion des ensembles de données / fichiers de données et de leurs relations. L'expression ***« banque de données »*** est, elle, utilisée pour parler des « gisements d'information informatisée et des produits d'information issus de ces gisements et proposés à différents publics ».

5.2 Les différents types de banques de données

Le nombre d'**informations** disponibles et les moyens de les diffuser sont en constante progression : la croissance exponentielle du *World Wide Web* a encore accru ce développement en fournissant l'accès à des banques de données très variées, avec une interface commune. Ces banques, assez couramment appelées bases, se situent au cœur de l'activité des entreprises et des administrations, et bon nombre d'activités humaines sont désormais liées à l'informatique. [6]
« L'information, bien économique, est consommé par notre société en grande quantité...la durée de vie des produits, mais aussi des théories ou des modes, est courte, obligeant à aborder de nouveaux domaines. ». Les **banques de données professionnelles** se développent ainsi à un rythme élevé. [72]

Différents types d'informations peuvent être diffusés au moyen de banques de données, permettant de dresser [72] une **typologie des banques** :

- banques de références bibliographiques (intégrant des informations issues d'un traitement documentaire, à savoir une indexation au moyen d'un thésaurus par exemple ou un résumé)
- banques de données intégrales ou plein texte (full-text), proposant par exemple l'intégralité du texte d'articles de lettres ou revues spécialisées.
- banques de données factuelles : fournissant des données validées et analysées par exemple sous forme d'informations dont la mise en forme est liée à la nature même (brevets, marques...).
- banques de données d'images.

La **production d'une banque de données** est le résultat d'une mise en commun de moyens humains et intellectuels, matériels et technologiques. Différents rôles sont à distinguer : auteur, producteur, serveur, vendeur, intermédiaire de l'information, formateur, etc. [72]

5.3 Les avantages attendus d'une banque de données

De **multiples avantages** sont attendus d'une banque de données, afin d'obtenir, rapidement, des informations précises, quel que soit le domaine. [72]
- la connexion et la consultation d'une banque de données sont quasi instantanées.
- l'obtention d'une référence ou d'une information, directement utilisable, est assurée par la possibilité de recevoir rapidement l'information ou le texte intégral souhaité, par différents moyens (courriel, SMS, chat, etc.).
- l'actualisation, la mise à jour des banques de données se fait de plus en plus souvent (pour certaines, jusqu'à plusieurs fois par jour). Certaines publications sont même disponibles en ligne avant même d'être imprimées (*pre-print*).
- l'étendue du champ de recherche, assuré par l'ensemble des banques de données accessibles via le Web, est sans commune mesure avec les ressources d'un centre de documentation traditionnel.
- l'interactivité : les temps de réponse raccourcis à des requêtes, permettent d'affiner, de faire évoluer les questions, de reformuler pour obtenir plus rapidement de nouveaux résultats.
- la précision des réponses aux questions posées.
- l'enrichissement du fonds documentaire, rendu possible grâce au téléchargement des notices d'une banque externe dans la banque interne.

5.4 Définition d'une base de données

On entend par **base de données**, un « ensemble de données structurées en champs et en relations, organisé en vue de son utilisation par différents types de programmes : création,

modification, suppression, tri, recherche, édition,... L'ensemble de l'application est géré par un système de gestion de base de données (SGBD) rendant indépendantes l'évolution des données et celle des programmes. » [79]

Nous intéressant principalement dans cette étude à la structuration des données sous différents systèmes de gestion développés ou utilisés au Centre INFFO, nous privilégierons ce terme de **bases de données** (BDD).

5.5 Conception : de la modélisation des données à la base de données relationnelles

En informatique, les bases de données sont utilisées pour stocker efficacement de très grandes quantités d'informations. Ce stock d'informations est organisé et structuré de manière à pouvoir être facilement manipulé. Pour qu'une base de données constitue un système robuste et évolutif, il est nécessaire d'adopter une démarche de **conception** rigoureuse. En effet, une mauvaise conception de départ conduit à stocker des données inutiles (redondance), à générer d'autres incohérences (bruit, silence), voire même à aboutir à des données inutilisables. Par ailleurs, « une structure de données inadaptée peut provoquer des erreurs fondamentales d'expression dans l'interrogation de la base ». [6]

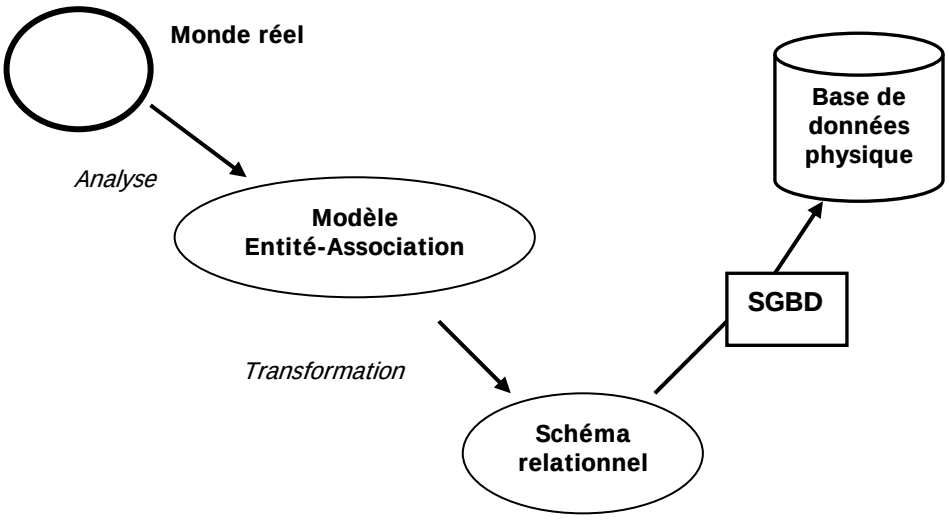
On distingue l'**organisation logique** des données qui se fait selon un modèle de données et la **structure physique** des fichiers contenant les données.

Le **modèle de données** ou ***Data model***, consiste en la « formalisation de l'architecture des données permettant de rendre compte des phénomènes pouvant agir sur elles. Pour être efficace, cette formalisation doit s'appuyer sur une syntaxe de représentation rigoureuse. » [81]

Le **modèle de données relationnel** est aujourd'hui un modèle très utilisé dans le monde des bases de données parce qu'il permet l'indépendance entre la structure physique et l'organisation logique des données.

La conception d'une **base de données relationnelles** nécessite d'enchaîner plusieurs étapes : [6]

Etapes de la conception d'une base de données



- l'analyse du monde réel, c'est-à-dire l'analyse de la situation du monde réel à considérer afin de le modéliser, sous forme de **modèle conceptuel de données (MCD)** ou **schéma conceptuel**, par exemple un **modèle Entité-Association (E/A)**[16].

- la transformation de ce modèle entité-association en **modèle relationnel** ou **modèle logique des données** (par exemple sous forme de **schéma relationnel**) qui permet de traduire la réalité de l'information vers une représentation utilisable en informatique. Il correspond à la manière de structurer l'information dans la BDD.

- et enfin le passage à l'élaboration effective de la base, constituée de tables, qui contiendront les valeurs ou données. On parle alors de **structure physique**[17]. Cette étape nécessite un logiciel appelé **Système de Gestion de Base de Données** ou **SGBD**[18]qui permet de structurer et de manipuler la BDD (organisation, contrôle, consultation, modification...).

---

[16] Ce formalisme « entité-association » utilise une représentation graphique sous forme de diagramme. On peut également utiliser un autre formalisme, **UML - *Unified Modeling Language***, pour cette représentation.
[17] On parle également de **modèle de données physique**.
[18] On parle aussi de **SGBDR** pour **Système de Gestion de Base de Données Relationnelles**.

5.6 Format documentaire ou format documentaire informatique

Le **format documentaire** ou **format documentaire informatique** est la structure des données dans la base de données. Il s'agit de la répartition ordonnée d'informations. Le logiciel retrouve, trie l'information et répond aux questions d'autant plus aisément et rapidement que l'information est structurée et organisée. Un format documentaire adapté à ses besoins permet, tout en restant conforme aux normes nationales et/ou internationales en vigueur, de saisir, d'interroger, d'éditer ou d'échanger facilement des données. « Un format documentaire adapté est celui qui :

- présente tous les champs nécessaires pour répondre à tous les besoins documentaires,
- est adaptable,
- est portable afin de pouvoir être implanté sur n'importe quel logiciel documentaire,
- est facile à utiliser,
- est compatible avec les formats et normes internationaux pour faciliter les échanges de données sans reformatage lourd. » [8]

6 De l'interopérabilité

L'interopérabilité est un des thèmes fondamentaux du management des systèmes d'information. A. Berdugo, R. Mahl et G. Jean la définissent comme la « capacité de plusieurs systèmes d'information hétérogènes à fonctionner ensemble en utilisant des standards communs et à interagir afin de répondre conjointement à la satisfaction des processus de gestion de l'entreprise ». [76]

L'interopérabilité peut aussi être définie fonctionnellement comme la capacité d'une ressource ou d'un service à s'intégrer dans un ensemble plus vaste. Et techniquement, selon le NISO – *National Information Standards Organization*, c'est « la capacité d'échanger et de partager des données entre systèmes multiples disposant de différentes caractéristiques en termes de matériels, logiciels, structures de données et interfaces, et avec le minimum de perte d'information et de fonctionnalités ». [20] [21] [65]

L'interopérabilité des systèmes est la faculté pour ces systèmes d'échanger des données, des services ou des informations, donc de communiquer entre eux, sans ambiguïté et de la façon la plus transparente possible. L'interopérabilité est rendue possible grâce à des interfaces ou des protocoles normalisés. Deux systèmes ne pourront « dialoguer » entre eux, ne seront « compatibles » que s'ils respectent les exigences imposées par les normes qui définissent les interfaces nécessaires (cas des systèmes bancaires, ferroviaires, informatiques…).
Dans le domaine du Web, l'interopérabilité porte sur l'ensemble des éléments du système : les réseaux via lesquels transitent les données, les systèmes informatiques pour les transactions, les formats des données informatiques, les formats documentaires.

Les documents numériques structurés (et les langages à balisage que constituent SGML, HTML et XML) se sont développés dans ce contexte, qui nécessite à la fois :

- une exploitation la plus automatique possible des données, imposée par des flux et des volumes d'information croissant transitant sur les réseaux, ainsi qu'une diversification des types de ressources ;
- des articulations renforcées entre les systèmes d'information (gestion documentaire, gestion des commandes et achats, gestion des droits, système d'édition, système de production d'information, etc.) imposées par une multiplication des acteurs, une diversification des usages, un développement renforcé des applications informatiques ;

- une prise en compte des usages des informations, des équipements informatiques, des modalités d'affichage de l'information propres à chaque contexte d'activité.

L'interopérabilité entre systèmes d'information documentaire (SID) présuppose que les contenus d'informations soient identifiés et décrits de manière formelle.
Les évolutions de ces dernières années, le passage de la notion de documents à celle d'objets d'information numériques, amènent les SID à enrichir leurs modèles de données pour fournir plus d'informations sur ces objets, que les traditionnelles notices bibliographiques. Plus que la description formelle de documents destinée à les citer et les retrouver dans le système, il s'agit également d'en faciliter l'exploitation automatique ultérieure, pour des usages variés.

L'interopérabilité des ressources, selon Catherine Morel-Pair [65], met en fait en jeu **trois niveaux techniques complémentaires** d'interopérabilité :

- le niveau des jeux de métadonnées[19] : une description des ressources avec des sémantiques communes issues de différents jeux standardisés (Dublin Core, MARC-XML, MODS, EAD, LOM...) ;
- le niveau du cadre générique d'implémentation de ces descriptions dans des langages structurés standardisés, interprétables par les machines (XML, RDF, espaces de nom, URL) ;
- le niveau des protocoles informatiques d'échange de ces données normalisées (HTTP, OAI-PMH...).

Cet auteur distingue également un **quatrième niveau** d'interopérabilité, plus **organisationnel** quant à lui, qui implique des pratiques communes dans l'utilisation des éléments et dans les valeurs de ces éléments : codes, vocabulaires, données d'autorité reconnus. Ce consensus se retrouve dans les « spécifications » de chaque jeu de métadonnées, qui sont formalisées dans des « registres de métadonnées » qui décrivent chaque élément (définition, hiérarchie, usage, valeurs attendues, éventuellement correspondances avec d'autres jeux, etc.) et dans des *guidelines*. [65]

XML - *eXtensible Markup Language*, est aujourd'hui au cœur de la problématique d'interopérabilité des systèmes d'information. Ce format de données « standardise le dialogue entre applications qui n'ont pas été conçues pour communiquer entre elles et s'affirme comme le format pivot des échanges interentreprises et interapplications ». [76]

Risques liés au déploiement d'un projet d'interopérabilité de systèmes d'information

Les auteurs A. Berdugo, R. Mahl et G. Jean [76] attirent l'attention sur certains risques souvent cités par des chefs de projets et la direction générale des entreprises concernées :

- des projets trop longs et qui semblent s'enliser ;
- une perte du consensus de départ sur les objectifs et les moyens mis en œuvre ;
- des résistances opposées par les salariés dont les méthodes de travail sont fortement impactées par les outils déployés et la nouvelle infrastructure.

Moyens de limiter ces risques :

- « mettre en place un projet pilote afin de mieux évaluer les difficultés » pouvant être rencontrées et se fixer des objectifs atteignables ;
- définir « le lotissement du projet de façon à obtenir à échéances régulières des *quick wins*, victoires rapides à même de redynamiser les équipes ;
- constituer une équipe projet composée de responsables métiers et de responsables informatiques, puis définir rôles et missions de chacun des acteurs ;
- impliquer le management dans le projet en lui rapportant régulièrement les *quick wins* et les difficultés rencontrées lors du déploiement du projet.

[19] Sur l'interopérabilité au niveau des Schémas notamment, lire [46] et [47].

7 L'information non-structurée : le défi de sa maîtrise et de sa gestion

7.1 Information non-structurée et information structurée

« Il est devenu coutumier de dire que 80 % des **informations** qui circulent dans une organisation sont **non-structurées** », contre 20 % qui seraient **structurées** ». Ce chiffre, qui constitue un ordre de grandeur et reflète une impression largement partagée, était avancé en octobre 2007 par un Livre blanc de trois grandes associations professionnelles du secteur des nouvelles technologies de l'information et de la communication : l'APIL – Association des Professionnels des Industries de la Langue, l'APROGED – Association des Professionnels pour l'Economie Numérique et le CIGREF – Club Informatique des Grandes Entreprises Françaises [1]. Alain Garnier va plus loin : il avance pour sa part qu'à ce jour, « 99 % de l'information est non-structurée » [7] [9].

« On considère qu'une information est structurée si elle est utilisable directement par un ordinateur pour effectuer un calcul » - opération arithmétique ou comparaison – fournissant un résultat utilisable par l'ordinateur ou par son opérateur. [1] [5]
L'**information structurée** est répétable, systématique, calculable (ex : formulaires, factures, fiches de paie, libellés,...). [9]
L'**information non structurée** renvoie à un usage non inscrit dans un modèle (ex : courriers électroniques, majorité des contenus sur Internet). [9]
« L'objet du traitement de l'information non-structurée est de rendre des données (textes, images, sons...) calculables. Ainsi, ce n'est donc pas la nature de l'information mais "l'utilisabilité" d'une donnée qui marque la frontière entre **structuré** et **non-structuré**.
[1] [5]

7.2 Maîtrise et gestion de l'information non-structurée : un grand défi managérial actuel

La gestion de l'information non-structurée est l'un des grands défis managériaux actuels. Sa maîtrise peut induire de très importants gains de productivité et être un « véritable relais de croissance pour les entreprises ». [7]

L'ensemble de cette information non-structurée est contenue dans des **documents numériques** de forme et de nature variées, qui offrent de facto un premier niveau de structuration par leurs **métadonnées** : un nom, une adresse, une date, etc. [1] [5]

Elle recoupe une bonne part de l'**information textuelle**, **produite** et **reçue** par les acteurs de l'entreprise : courriels, rapports, notes, ou encore textes collectés dans des sources hétérogènes. [7]

L'activité d'une entreprise dépend de façon grandissante de l'acquisition, de la circulation, de l'utilisation, de la diffusion et de la conservation de cette information. La traditionnelle **GED – Gestion Electronique de Documents**, apparue avec les technologies de numérisation des documents, a évolué et l'on parle désormais d'**ECM – *Enterprise Content Management***, au périmètre plus large, qui regroupe les outils et méthodes permettant d'exploiter et de valoriser l'information non structurée, c'est-à-dire « capturer, gérer, conserver et diffuser le contenu et les documents liés aux processus organisationnels et métiers de l'entreprise. ».

La mise à disposition de l'information, dont la non-structurée, procède notamment de la « définition de **workflows** qui reflètent et organisent la circulation de l'information dans l'entreprise et les différentes actions qui y sont associées (validation, édition, distribution). Ces **workflows** modélisent les **processus documentaires** et, de ce fait, ils modélisent les processus métiers et épaulent ainsi une démarche de gestion de processus d'entreprise, **BPM – *Business Process Management*** ».

L'innovation est au cœur de la gestion de l'information non-structurée et de sa valorisation. Les **standards de représentation des documents** - à l'instar des Schémas XML tels HR-XML pour les ressources humaines, NewsML pour le journalisme, etc. – sont des éléments fédérateurs et structurants, permettent la convergence entre différents systèmes. Ils prennent en compte les **processus métiers** et s'y adaptent, créent un lien technologique entre les documents et ces processus. [1] [5]

8 Norme ou standard ?

Dans le domaine de l'information et de la documentation de nombreux **standards** et **normes** sont utilisés. Ils concernent la description d'objet documentaire, la production de documents numériques et de leurs métadonnées, l'échange de données bibliographiques, etc. [63]
De nombreux organismes sont impliqués dans l'élaboration et la publication de ces standards, recommandations et normes (association, consortium, état public, organisme international, organisation non gouvernementale…). [66]

Les termes norme et standard sont souvent confondus, utilisés l'un pour l'autre, notamment à l'occasion du passage de l'anglais au français.

8.1 Norme

« Une **norme** est un document établi par consensus et approuvé par un organisme reconnu, qui fournit, pour des usages communs et répétés, des règles, des lignes directrices ou des caractéristiques, pour des activités ou leurs résultats, garantissant un niveau d'ordre optimal dans un contexte donné.
Il convient que les normes soient fondées sur des acquis conjugués de la science, de la technique et de l'expérience et visent à l'avantage optimal de la communauté. (ISO/CEI Guide 2 :1996 et NF EN 45020) » [81]
Les normes doivent être validées, et publiées, par des organismes qualifiés reconnus comme :
- l'**AFNOR** – Association Française de Normalisation au niveau national,
- le **CEN** – Comité Européen de Normalisation au niveau européen,
- l'**ISO** – *International Organization for Standardization* au niveau international.

« La démarche normative est par essence au service du métier et de ses bonnes pratiques », fournissant selon la définition de l'UNESCO « pour des usages communs et répétés, des règles garantissant un niveau d'ordre optimal ». [66]

8.2 *Standard* - Standard

Le terme anglais ***standard***, souvent utilisé dans certaines entreprises, caractérise une norme, au sens d'un document normatif de l'entreprise, par opposition à la norme qui elle, est nationale ou internationale. [81]

Standard est alors utilisé pour nommer des règles et des prescriptions qui sont en fait des normes techniques servant à fixer les caractéristiques d'un produit, d'une activité ou d'un processus[20].
Mais ***standard*** en anglais sert également à traduire le terme français norme[21].

Le terme français **standard** désigne un « élément de référence, modèle qui doit être appliqué afin de respecter les règles qu'il décrit ou qu'il représente ». [81]

En ce sens, par exemple, un **standard** a pour objectif d'uniformiser une méthode de travail, en particulier pour assurer l'interopérabilité de systèmes, constitué d'un ensemble de règles, opérations, protocoles, caractéristiques couramment employés et consacrés par l'usage. Un standard résulte d'un consensus entre acteurs du monde économique, d'une même entreprise, d'une même corporation, ou de plusieurs métiers connexes, sans toutefois être validé par un organisme de normalisation, ne faisant par conséquent pas référence aux instances citées plus haut.

Un standard peut prendre la forme de **spécifications**, caractéristiques techniques, **recommandations**, faisant référence aux exigences techniques auxquelles doit répondre un produit, un processus ou un service... (Ex : spécifications ou recommandations du W3C - *The World Wide Web Consortium*). [37]
Mais ce qui relève parfois de spécifications d'un groupe de travail pluridisciplinaire, peut finalement aboutir à une reconnaissance officielle sous la forme de norme (ex : le Dublin Core, normalisé par l'ISO).

---

[20] Source : <http://www.granddictionnaire.com/>.
[21] Sylvie Millara précise que l'application et la prise en compte de normes internationales uniques sont confrontées à un obstacle : « à l'évidence la conception différente et même opposée d'une norme, entre l'Europe pour laquelle elle est un standard technique adopté par consensus au sein d'un organisme officiel, et les Etats-Unis pour qui une norme est un "produit", réalisé par un organisme sectoriel, en d'autres termes un "standard de fait" qui s'imposera sur le marché ». [61]

9 Schémas, référentiels, registres, jeux de métadonnées, profils d'application

Des ensembles de métadonnées sont standardisés ou normalisés par des communautés de praticiens, documentées et codifiées selon « des règles et des vocabulaires communs » pour que ces métadonnées « puissent être partagées et réexploitées dans des contextes variés ». [13] Plusieurs vocables sont très souvent utilisés dans la littérature pour parler de systèmes de représentation des métadonnées (et des données qui en sont la source) : **schémas, référentiels, registres et jeux de métadonnées, profils d'application**. Les explications et citations qui suivent tentent d'apporter des réponses, d'éclairer sur ces termes, présentés dans leurs contextes d'usage.

Schéma des données informatiques

Il s'agit d'un schéma des données sous un format adapté aux activités informatiques : ce schéma et sa documentation concrétisent le modèle conceptuel[22] ou schéma conceptuel, par exemple un modèle Entité-Association, élaboré dans une étape précédente à partir de l'analyse du monde réel. Exemples de schémas des données : la liste des « tables d'une base de données relationnelles, ou un schéma[23] normalisant le balisage des informations et de leur arborescence dans un document XML ». [13]

Les schémas de données, comme tout outil informatique, doivent être correctement spécifiés et documentés, pour pouvoir répondre aux besoins de fonctionnement, aux besoins de transparence sur les applications et métadonnées (cf. interopérabilité). Mais aussi aux besoins d'évolution fonctionnelle ou technique (cf. évolutions du Web et de ses usages).

Référentiels

Selon Cécile Kattnig [58], citant Sylvie Dalbin[24], les **référentiels** sont de périmètre et de composition variable. Ils sont « utilisés pour représenter des objets-ressources (identifiants, schémas de métadonnées, vocabulaires contrôlés), pour produire ou récupérer des données (outils logiciels) ou pour administrer (politiques, procédures, règles) ». La modélisation et la normalisation sont des éléments-clés de la convergence des référentiels.

[22] Cf. chapitre 5 de cette Première partie du mémoire.
[23] Cf. chapitre 11 de cette Première partie du mémoire.
[24] Cf. [58] et <http://www.bnf.fr/PAGES/infopro/journeespro/no-Afnor2008.htm>.

En informatique, un **référentiel** désigne « un **ensemble de bases de données** contenant les “ références ” d’un système d’information »[25]. Il est aussi utilisé pour désigner « un **entrepôt de métadonnées de ressources**, c’est-à-dire pour une base de références de ressources décrites suivant le schéma de métadonnées en référence ». [13]

Dans le cas d’une application informatique, le terme **référentiel**, également appelé **registre référentiel**, associe les outils et les données d’encodage (à caractère informatique, sous-entendu compréhensible par la machine), ainsi que les règles, bonnes pratiques ou directives. [13]

Les **référentiels de représentation des contenus** tiennent un rôle important dans l’interopérabilité des systèmes d’information hétérogènes et dans leur exploitation dans le Web sémantique[26]. La modélisation et la normalisation sont, à cet effet, des éléments-clés de la convergence des référentiels. [58]

Registre de métadonnées – ***metadata registry***

Ensemble des outils propres à la production et à l’administration d’un ensemble d’éléments de données : **jeu de métadonnées**, règles, vocabulaire d’encodage et éventuellement tableaux de concordance. Ces outils sont documentés et mis à la disposition de communautés, sans être directement exploités au sein d’une application. Un registre de métadonnées peut comporter une dimension réglementaire ou obligatoire. (D’après [13])

Profil d’application ou **profil applicatif** – ***AP - Application Profile***

Pour Sylvie Dalbin, « on parlera de **profil d’application d’un schéma donné** lorsque, au sein d’une organisation, d’une communauté d’utilisateurs ou d’une application précise, sont déclarés non seulement un jeu de métadonnées dérivé du schéma source mais aussi la codification des éléments de données et des valeurs, la normalisation des vocabulaires de codification (on parle aussi de **schémas d’encodage**) utilisés pour désigner ces valeurs, ainsi que des règles d’utilisation pour le développement d’un schéma technique ». [50]

Josiane Ducournau [16] nous en rapporte une définition plus synthétique : « Sélection d’éléments d’une norme, d’un standard ou d’une spécification formant ainsi un sous-ensemble adapté aux besoins des groupes qui l’utilisent. Le sous-ensemble d’éléments est défini pour fournir un cadre d’opération ».

---

[25] Cf. Michel Volle « Qu'est-ce qu'un "référentiel" ? » <http://www.volle.com/travaux/referentiel.htm>

[26] Sur le Web sémantique, voir le chapitre 2.3.5 de la Quatrième partie de ce mémoire.

Exemples de profils : **LOM-FR** <http://www.lom-fr.fr/> est le profil d'application français du modèle **LOM - *Learning Object Metadata***[27], schéma de description de ressources d'enseignement et d'apprentissage qui est un standard, tandis que **Normetic** <http://www.normetic.org/> est en le profil québécois.

[27] Sur LOM et LOM-FR, voir le chapitre 1.2.2 de la Quatrième partie de ce mémoire.

10 XML - *eXtensible Markup Language*, recommandation du W3C pour une nouvelle modélisation des données

10.1 Le XML, *eXtensible Markup Language*, langage à balises extensible

▪ Le langage XML, fait partie des **langages de description des formats de documents**, permettant de mettre en forme des documents grâce à des balises. Historiquement, trois principaux langages sont apparus pour réaliser le codage de textes numériques : SGML, HTML puis XML. [25] [28]

SGML – *Standard Generalized Markup Language*, langage normalisé de balisage généralisé a permis à partir de 1986 le développement d'applications éditoriales. Ce langage est apparu lourd à exploiter, mélangeant la notion de modèle de document et le format de codage des documents conformément à ce modèle. Il nécessitait un apprentissage complexe et n'était pas adapté à une utilisation quotidienne.

HTML – *Hypertext Markup Language,* langage de balisage hypertexte, a ensuite été développé à partir d'une simplification du SGML, pour être utilisé sur le réseau Internet. Le fichier contenant les textes, inclut à la fois les données à afficher et les éléments relatifs à leur présentation. Cette pauvreté fonctionnelle a toutefois permis la mise en œuvre facile de HTML, ainsi qu'un développement important des sites Internet.

XML - *eXtensible Markup Language*, traduit en français par « **langage à balises extensible** » ou bien parfois par « **langage à balises étendu** » a été développé en tenant compte des avantages des deux standards tout en minimisant leurs faiblesses, les difficultés de SGML d'une part et la « pauvreté sémantique » de HTML d'autre part. [23] [25] [37]

▪ **XML** est un **format informatique de données**, mis au point à partir de 1996 par le *XML Working Group*. Il est devenu en 1998 un *standard*, une recommandation du **W3C - *World Wide Web Consortium*** <http://www.w3.org/XML> également qualifiée de **norme XML**, au sens d'un « standard industriel ».

▪ **La norme XML** constitue un outil permettant de définir un « **langage de balisage extensible** basé sur la séparation nette entre le contenu, le style et la structure des documents. XML est un **métalangage**, ensemble de règles permettant de définir d'autres langages balisés spécialisés. »[28] [25] [78] [79]

- Le **langage XML** est ainsi qualifié de **métalangage,** car il permet de définir des formats de documents comportant des règles de grammaire strictes et de créer des documents respectant ces formats ;
- En XML les données sont **balisées**, c'est-à-dire qu'elles sont encadrées par des balises ouvrantes et fermantes, véritables **balises sémantiques** qui expriment le contenu, la signification de la chaîne de caractères ainsi délimitée. Ex : <organisme>Centre INFFO</organisme> ;
- La notion d'**extensibilité** signifie que les balises ne sont pas prédéfinies et qu'il est possible de créer ses propres balises sémantiques ;
- Les données XML **séparent le contenu et la structure de la présentation**, ce qui facilite la présentation des informations dans différents formats et applications. ».

XML fait partie des **standards actuels du Web**, ceux-ci garantissant notamment l'**accessibilité** aux informations, aux données mises en ligne :

- pour les logiciels (ex : navigateurs, systèmes d'exploitation tels Windows, Mac OS ou Linux) et les matériels informatiques (ex : MacIntosh ou PC),
- pour tous les publics, y compris les personnes présentant un handicap visuel.

[28] Définition également reprise sur le site de l'ADBS - Association des professionnels de l'information et de la documentation <http://www.adbs.fr/xml-extensible-markup-language--19063.htm?RH=OUTILS_VOC>.

Les **standards du Web**[29] garantissent ainsi notamment :

- **interopérabilité** et **portabilité** : les sites Web s'affichant de manière correcte quel que soit le terminal d'accès (ordinateur, téléphone mobile, PDA – assistant numérique personnel...). La mise en forme varie, mais le contenu reste accessible et affichable ;
- **production de contenu Web facilitée** : la séparation du contenu et de sa mise en forme autorise les changements rapides apportés à l'apparence des sites ;
- **volume des données optimisé** : la redondance des informations codées est éliminée et assure une réduction du volume des pages Web ;
- une **visibilité améliorée et accrue des données** sur le Web : **l'indexation** par les moteurs de recherche se voit enrichie par la **structuration** de l'information[30] ;
- **pérennité des documents** par la possibilité de les convertir en d'autres formats, simplifiant la migration des données vers de nombreux logiciels ou serveurs.

10.2 Le document XML : un document structuré

Un document XML est structuré, composé d'unités de stockage appelées **éléments**. Chaque élément stocke donc du contenu, encadré par des balises sémantiques qui en expriment la signification.

Ex : <organisme>Centre INFFO</organisme>

Un élément peut être qualifié par des **attributs** qui servent à nommer, identifier, définir chacun des éléments de données, à les relier entre eux, les représenter et les administrer.

Exemple :

```
<organisme>
        <nom>Centre INFFO</nom>
        <adresse>
        <rue>4 avenue du Stade-de-France</rue>
        <code postal>93218</code postal>
        <ville>Saint-Denis La Plaine cedex</ville>
        </adresse>
</organisme>
```

[29] Sur les standards du Web se reporter notamment au site en français <http://www.openweb.eu.org/> présentant des exemples concrets d'utilisation des recommandations du W3C. [30]

[30] De cette manière les moteurs de recherche sur le Web peuvent extraire directement par exemple le résumé ou le titre d'un paragraphe d'un fichier XML. « L'indexation peut ainsi

Pour exprimer des relations complexes, les éléments peuvent être imbriqués hiérarchiquement. Tous les éléments sont rattachés à un élément principal : **l'élément racine**. Les éléments structurent le contenu du document XML sous la forme d'un arbre représentant sa **structuration**, appelé **l'arbre des éléments**.

On obtient ainsi une **structure logique arborescente**. « Cet arbre part d'un élément racine unique et subdivise en éléments branches et feuilles qui correspondent aux constituants logiques du document. L'ensemble du document est structuré sous la forme d'éléments qualifiés par des attributs avec des valeurs. » L'ensemble de ces éléments ne sont pas prédéfinis : ils sont choisis par l'utilisateur en fonction du type de document à représenter (cf. chapitre 11 suivant). [25]

```
<?xml version="1.0" encoding="UTF-8"?>
  <!-- '''Commentaire''' -->
  <élément-document xmlns="http://exemple.org/"
xml:lang=";fr">
    <élément>Texte</élément>
    <élément>élément répété</élément>
    <élément>
      <élément>Hiérarchie récursive</élément>
    </élément>
    <élément>Texte avec<élément>un
élément</élément>inclus</élément>
    <élément/><!-- élément vide -->
    <élément attribut="valeur"></élément>
  </élément-document>
```

Source : <http://fr.wikipedia.org/wiki/Extensible_Markup_Language>

Un document XML par sa structure rigoureuse ressemble à une **base de données**[31] : il contient la **structure logique du document**, c'est-à-dire les **éléments de description** équivalents aux champs d'une base de données. Toutefois le fichier XML a la particularité de ne pas être lié à un logiciel particulier et son **format ouvert** sous forme de **fichier texte** (qui englobe contenus et balises) autorise n'importe qui à le retravailler.

Exemple d'utilisation : un fil RSS d'un site Web est un fichier texte au format XML permettant d'encadrer des articles – contenus d'informations de ce site – par des balises sémantiques titre, description, lien.

être plus précise ; un mot figurant dans un titre étant plus important que le même mot dans une note de bas de page. » [6]

[31] En tant que format d'échange universel, XML peut constituer « une alternative économique et fiable aux BDD relationnelles de taille moyenne ». [25] Des langages de requête, comme **XQuery**, permettent d'accéder aux composants des documents XML de la même manière que dans une BDD. [27]

10.3 La modélisation XML au service de la conception du système d'information

J.-J. Thomasson et A. London [35] écrivent, en 2006, que XML est un **nouveau type de modèle de données** qui, à ce titre, doit recevoir de **nouvelles techniques de modélisation**. Utiliser le langage de modélisation objet unifié **UML – *Unified Modeling Language***, pour concevoir des modèles XML n'est pas direct[32]. XML est bien plus qu'un nouveau format d'échange de données entre ordinateurs, bien plus qu'un alignement de balises et d'attributs. Au-delà d'un simple format d'enregistrement, « on exprime en XML des structures de données, des documents pourtant usuellement assimilés à des données non structurées, des descriptions de processus, des ressources informatiques physiques, mais également des concepts totalement abstraits, des langages de programmation », etc. XML permet de manipuler des espaces d'information à n dimensions. Les données XML sont au cœur d'applications de plus en plus interconnectées, et, par les enjeux économiques qu'elles représentent, il faut « prendre le plus grand soin » au moment de leur modélisation.

La **conception d'un modèle de données XML**[33] diffère grandement de celle d'un modèle relationnel. Un « modèle relationnel se contente de naviguer entre données et formulaires, au mieux des catalogues, tandis que XML navigue entre données et documents, dans le sens le plus large de ces deux termes » [35]. Avec XML, la frontière entre les données dites structurées, gérées dans des tables relationnelles, et les données non structurées, c'est-à-dire les documents, est abolie.
Ainsi le document XML a toute sa place aujourd'hui à s'insérer dans la chaîne de traitement de l'information. [33]
Selon la pyramide du savoir : en passant de la donnée au document, on transforme les **données** en **information**, transition nécessaire pour passer au stade de la **connaissance** et du **savoir**. J.-J. Thomasson [36] propose une analogie entre pyramide du savoir et XML au travers d'un schéma, reproduit ci-après, présentant XML comme « le ciment idéal liant les différentes couches de la gestion des connaissances ».

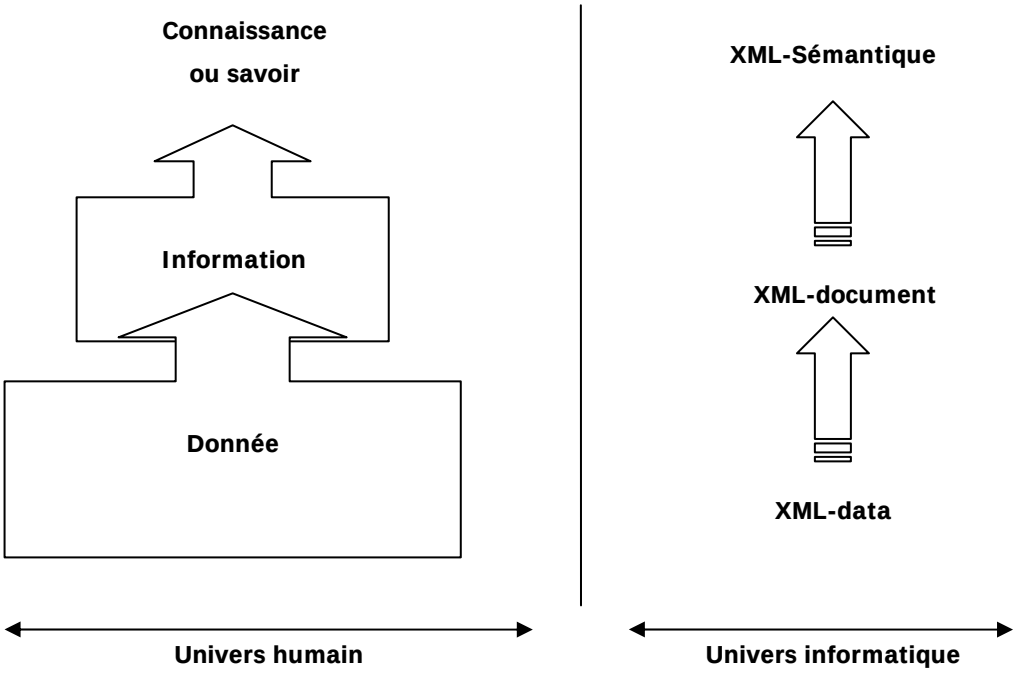
Citant Shigehisa Tsuchiya, professeur à l'université de Tokyo, spécialiste de la gestion des connaissances : « Bien que les termes données, information et connaissance soient souvent

---

[32] Selon ces auteurs, le diagramme de classes, qui en UML permet d'effectuer une représentation graphique des classes d'un système de données et les relations qui les relient les unes aux autres - est celui qui « se rapproche le plus des besoins de représentation des schémas XML. Il n'est malheureusement pas parfait pour cet usage » <http://www.eyrolles.com/Chapitres/9782212115215/Glossaire_Thomasson.pdf>.
[33] Utilisant une approche orientée services : cf. chapitre 2.2 de la Première partie de ce mémoire, portant sur la conception d'un système d'information, mettant en avant ce type d'architecture.

pris les uns pour les autres, il existe une distinction claire entre eux. Quand on donne du sens à la donnée via un cadre interprétatif, elle devient information, et quand on lit une information en lui donnant du sens via un cadre interprétatif, elle devient de la connaissance ».

Analogie entre la pyramide du savoir et XML



Toujours selon J.-J. Thomasson [36], tout projet de migration vers XML d'une chaîne de production de document nécessite une réflexion de l'entreprise sur ses modes de production d'information. La **conception du système d'information et de communication** utilisant XML devra en fait mettre en œuvre plusieurs modèles de données XML, plusieurs Schémas XML (cf. chapitre 11 suivant), conçus non seulement à partir de l'analyse de la structure interne des données, mais aussi à partir d'une réflexion sur les attentes des utilisateurs, des usagers consommateurs de l'information délivrée par le système.

10.4 Recommandations XML complémentaires

XML constitue bien un **métalangage**, un ensemble de règles permettant de définir d'autres langages balisés spécialisés – ou « vocabulaires ». Mais la recommandation XML n'est pas un ensemble de balises précises. Les balises d'un document structuré XML, leur agencement, ainsi que les styles de présentation du document ne sont pas définis dans la recommandation XML 1.0. Le métalangage XML ne permettant donc pas de développer

directement des applications, d'autres recommandations (spécifications) ou *standards*, ont été élaborées par le W3C <http://www.w3.org/standards/xml/>[34] pour permettre concrètement le développement d'applications. [37]

Les deux prochains chapitres portent respectivement sur :

- les formalismes de modélisation de structures : des documents, les DTD – Définitions de Type de Document et les Schémas XML servent à indiquer les règles à appliquer afin de structurer un document au format XML ;
- des langages complémentaires assurent la présentation, le traitement, la gestion des liens (liaisons entre documents), et le nommage des éléments et attributs des documents XML.

[34] Le site anglophone **The CoverPages** <http://xml.coverpages.org/coreStandards.html> [29] propose une rubrique dédiée à la description des grandes recommandations et standards utilisés dans les langages balisés : XML, SGML, Schémas XML, XSL/XSLT/XPath, XLink, XPointer, XML Query, CSS, SVG. Le site francophone **<XML>fr** <http://xmlfr.org/> [23], quant à lui, propose des traductions des recommandations W3C relatives à XML et ses compléments, sans valeur de référence toutefois.

11 Déclarations de structure XML : utilisation de DTD ou de Schémas XML

11.1 DTD – Définition de Type de Document

▪ La **DTD - Définition de Type de Document** ou ***Document Type Definition***, déjà utilisée dans le cadre du langage structuré SGML précurseur de XML, est un document permettant de **décrire un modèle de document XML**. Il s'agit d'un **fichier au format texte** qui décrit les règles – balises et structure – à appliquer pour structurer un document au format XML selon cette DTD[35]. [25] [34]

> Ex : la **DTD Dublin Core** définit la structure minimale pour décrire des notices de ressources électroniques. La **DTD NewsML**[36] définit la structure pour la description de messages d'actualité.

▪ Les **prescriptions de DTD** recouvrent :

- les balises ou noms des éléments qui peuvent être utilisés ;
- les attributs qui peuvent être rattachés aux éléments ;
- la mise en séquence et la structure finale du document XML.

▪ Utilisation de DTD

Une DTD permet ainsi de décrire la **structure d'un document XML**. Mais la puissance de description d'une DTD est considérée comme faible par les spécialistes du domaine, car une DTD n'indique pas par exemple la **typologie de données** contenues dans ses balises. Le **typage des données** est extrêmement limité puisque seul le type « texte » (CDATA) existe et qu'il n'y a aucun moyen de typer des dates, des nombres (entier, décimal), des chaînes de caractères soumises à des contraintes, etc. Les DTD ne sont donc à utiliser que lorsque la typologie des données contenues dans le document XML ne concerne pas ou est de moindre importance pour les traitements automatiques visés, c'est-à-dire les traitements qui seront à opérer ultérieurement sur les données gérées dans le document / le fichier XML.

---

[35] S. Rouleux présente les notions relatives aux DTD au moyen de deux exemples d'application : une DTD créée pour répertorier les dictionnaires d'une bibliothèque ; puis proposition d'une DTD adaptée aux éléments de deux pages extraites du « Dictionnaire Hachette Junior ». [34]

[36] NewsML est une spécification de l'IPTC - *International Press Telecommunications Council* <http://www.newsml.org/> ou

Actuellement, un grand nombre de DTD sont disponibles publiquement, soit pour des applications spécifiques, soit pour de grandes catégories de documents. Le site Cover Pages [29] liste de nombreuses applications XML et initiatives ayant développé des DTD, mais aussi des Schémas XML, beaucoup plus complets <http://xml.coverpages.org/xmlApplications.html>.

11.2 XML-Schema ou XSD – *XML Schema Definition*

▪ **XML-Schema** a été développé par le W3C pour décrire de manière plus fine et complète[37] la structure d'un document XML qu'une DTD, plus simple. Il est ainsi par exemple possible de spécifier **la typologie des données** (**typage**) – chaîne de caractères, entier, décimal, intervalles de temps… - que va contenir le document XML décrit par XML-Schema.
Un fichier de description de structure ***XML Schema Document*** en anglais, ou **fichier XSD**, est lui-même un **document XML**. [36]
La recommandation XML Schema[38] a été approuvée en mai 2001, puis a fait l'objet d'une seconde édition incorporant des corrections et précisions en octobre 2004. [35]

▪ Suivant la recommandation XML-Schema, de nombreux **Schémas XML** se sont développés, jouant dans une application XML le même rôle que les DTD, mais avec des possibilités plus avancées de structuration des données, et autorisant les échanges de données dans le secteur spécifique concerné. [20] [21] [29] [50] [65]

Ainsi, dans le domaine information-documentation, on peut citer :

- **MARC-XML** et **BiblioML** – *Bibliographic Markup Language*, pour la description bibliographique de deux types de format **MARC** – *MAchine Readable Cataloging* : **MARC-XML** est un schéma permettant de représenter sous forme XML l'ensemble des champs employés par le format **MARC21**, tandis que **BiblioML** est une DTD XML permettant de convertir au format XML des données saisies au format **UNIMARC** ;
- **MODS** – *Metadata Object Description Schema*, schéma XML plus simple mais moins complet que MARC-XML ;
- **ONIX** – *Online Information Exchange* : pour favoriser les échanges concernant des produits éditoriaux – livres ou autres – entre éditeurs et bibliothèques ;

<http://www.iptc.org/cms/site/single.html?channel=CH0087&document=CMS1206527546450> (cf. chapitre 1.2.5 de la Quatrième partie de ce mémoire).
[37] D'autres syntaxes de Schémas plus puissantes encore existent, non traitées dans ce mémoire : Schematron <http://www.schematron.com/> et Relax NG <http://relaxng.org/>.
[38] Voir les rubriques du site du W3C <http://www.w3.org/standards/techs/xmlschema> et <http://www.w3.org/XML/Schema>.

- **EAD** – *Encoded Archival Description* : utilisé dans le monde des archives ;
- **TEI** – *Text Encoding Initiative* : offrant un modèle général d'encodage pour tout texte littéraire ;
- **TEF** – Thèses Electroniques Françaises : pour faciliter la description, l'archivage et l'exploitation des thèses.

11.3 Caractéristiques d'un document XML

▪ Un document XML peut préciser dans son en-tête qu'il respecte une DTD ou un schéma XML donné. Cette DTD, ou ce schéma XML, est en général définie dans un fichier séparé : on parle alors de **DTD externe** ou **Schéma externe**. Le fichier/document XML contient alors non seulement la référence à la DTD qu'il respecte dans son en-tête, mais également l'indication de sa localisation.
La DTD, ou le Schéma, peut également être directement définie à l'intérieur du fichier/document XML : on parle alors de **DTD interne** ou **Schéma interne**. Toutefois ce regroupement en un même fichier de la DTD/Schéma et du code d'affichage des données spécifiées s'avère une solution peu pratique, aboutissant à des fichiers volumineux, difficiles à mettre à jour.

▪ **Deux niveaux de conformité** sont à distinguer pour un document XML :
- un document XML **bien formé** est un document ne comportant pas de DTD mais qui respecte la syntaxe, les règles de base du XML.
- Un document XML **valide** est un document bien formé qui se conforme à la grammaire, c'est-à-dire la DTD ou le schéma, à laquelle ou auquel il est associé.

▪ **La logique de structuration XML**, ayant défini séparément la présentation du document XML, permet aussi de récupérer les informations contenues dans le document pour les utiliser.
Cette récupération des données se fait au moyen d'un outil : le **processeur XML**, programme informatique également appelé **parseur** ou ***parser* XML,** ou **analyseur syntaxique**. Deux types de parseurs existent : les parseurs validants permettent de vérifier qu'un document XML est conforme à sa DTD. Les parseurs non validants se contentent de vérifier que le document XML est bien formé [39]. [34]

---

[39] A la question : pourquoi valider une DTD ? Voici ce que S. Rouleux répond : « Une DTD est un contrat entre les producteurs et les consommateurs d'information. La validation permet au producteur de vérifier qu'il produit ce qu'il a promis au consommateur, [au consommateur] de vérifier ce que le producteur lui livre, [à celui-ci] de protéger ses

12 Manipulation, traitement de documents XML

12.1 Manipulation des documents XML : solutions techniques

La manipulation des documents XML renvoie à la capacité de les traiter, les filtrer, en sélectionner des portions précises, les transformer, etc. Ce terme « manipulation » est en ceci proche de celui utilisé dans le monde des bases de données. [33]
Il existe plusieurs solutions techniques pour traiter des documents XML.

Une première solution peut consister à utiliser des **API – *Application Programming Interfaces***[40]. Ces interfaces de programmation permettent aux développeurs d'accéder aux données XML et de les traiter en utilisant des programmes écrits dans nombre de langages de programmation traditionnels, comme Java, C++, C#, Perl, PHP, etc.
Ces API, non dédiées à un type de traitement particulier, couvrent une large gamme de besoins.
Pour certains traitements complexes, « l'utilisation de ces API peut cependant s'avérer lourde et il est alors préférable d'utiliser d'autres langages spécialement conçus ». Ces langages spécialisés ont l'avantage d'opérer directement sur le modèle de données XML. [33]

12.2 Autres standards XML : langages et outils pour la présentation, la transformation, la gestion des liens, le nommage des éléments et attributs

D'autres recommandations[41] ont été formulées par le W3C pour compléter la recommandation XML, qui ne fixe aucune règle, notamment, sur la façon dont va se

---

applications et à l'auteur d'application de simplifier son travail. La détection et l'analyse d'erreur est laissée au *parser* ». [34]

[40] Comme l'API DOM - *Document Object Model* <http://www.w3.org/DOM/> qui est une recommandation du W3C (une description de DOM est proposée sur le site <http://www.dicodunet.com/definitions/developpement/dom.htm>) ; comme API SAX - *Simple API for XML* <http://www.saxproject.org/> ou encore JAXP - *Java API for XML Processing* <https://jaxp.dev.java.net/>.

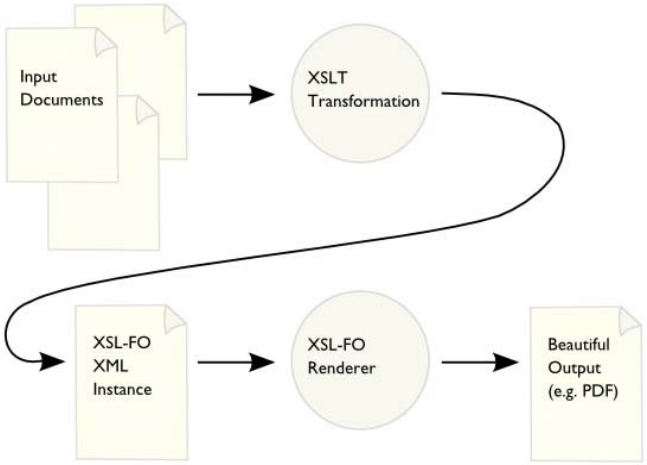
[41] La technologie XML est souvent qualifiée de « galaxie », « nébuleuse », voire même « maquis normatif ». [33]. De nombreux sites permettent de suivre l'évolution générale des techniques et recommandations XML, notamment un site francophone **<XML>fr** <http://xmlfr.org/> [23] et deux sites anglophones **The CoverPages**

présenter le document, sur la manière de traiter les objets, concernant la gestion des liens ou encore le nommage des éléments et attributs de documents XML. [33] [35] [37]

▪ **Présentation du document XML : transformation et mise en page** [25]

XSL – *Extensible Stylesheet Language* est un langage de description de feuille de style, permettant d'appliquer des styles à un document XML (pour le formatage du document XML ou sa présentation externe). En fait, **XSL** s'appuie sur plusieurs composantes, des langages complémentaires, **XSLT**, **Xpath** et **XSL/FO**, qui autorisent le retraitement d'un document XML et la modification de sa structure. XSL permet à partir d'un document XML de générer d'autres types de documents : TXT, RTF, PostScript, HTML, etc.

Pour définir la **mise en page** (affichage de textes ou graphiques), le langage **XSL/FO** formate les données de ce qui a été préalablement créé au moyen d'un langage de transformation : **XSLT**.



[Source : <http://www.w3.org/standards/xml/images/beautiful-output.jpg>]

XSLT – *XSL Transformations*[42] est un langage de transformation de document XML pour réorganiser, ajouter ou supprimer des balises et des attributs. XSLT autorise ainsi la manipulation de fichier XML dans différentes structures ou formats pour favoriser les échanges en différents systèmes, et permettre les sorties vers des périphériques variés.
Principe du **processeur XSLT** : un document (arbre) source est transformé en un document (arbre) résultat par un moteur de transformation qui applique une **feuille de**

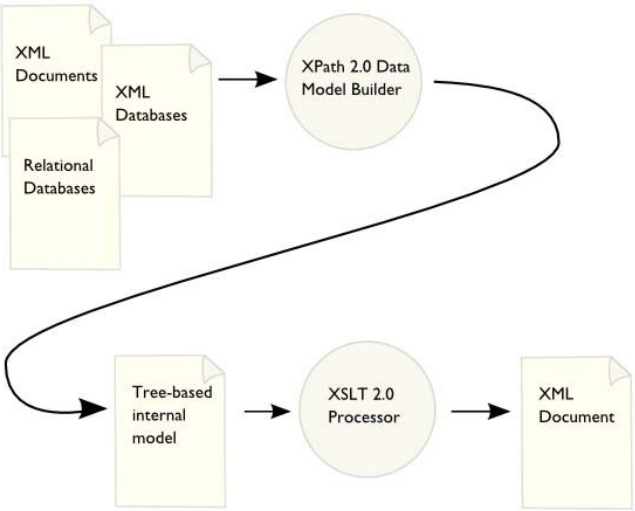
<http://xml.coverpages.org/> [29] et **XML.com** <http://www.xml.com/> [51]. **XML.org** [24], lui aussi en anglais, propose une liste dédiée aux développeurs XML-DEV <http://www.xml.org/xml-dev/>.

transformation ou de style XSLT. Ce processus offre l'intérêt de permettre de définir plusieurs feuilles de style pour un même document XML afin de varier l'affichage selon les besoins de l'utilisateur et la plateforme utilisée pour lire le fichier – ordinateur, téléphone portable, PDA, etc.

Les **feuilles de style** ou **feuilles de transformation XSLT** contrôlent la mise en page du document : elles déclarent quels éléments doivent être traités, dans quel ordre, et quel traitement leur est attribué.

La transformation peut également s'avérer nécessaire lorsqu'un document XML n'est pas « valide » par rapport à son modèle, DTD ou Schéma XML.

XSLT utilise la syntaxe et la sémantique d'un autre langage, **Xpath – *XML Path Language*** qui permet « l'extraction » et « l'expression de chemins » pour naviguer dans le document XML.



[Source : <http://www.w3.org/standards/xml/images/xdm-and-xslt.jpg>]

▪ **Gestion des liens** : [25]

Xlink et **Xpointer** utilisent également le langage **Xpath** pour identifier des fragments de documents. **Xlink** fournit une syntaxe XML pour décrire des liens hypertextes externes dans un document XML, tandis que **Xpointer** propose une syntaxe pour pointer sur des parties d'un document XML, c'est-à-dire définir des liens hypertextes internes au document lui-même.

[42] Voir le site du W3C <http://www.w3.org/standards/xml/transformation> et [28].

▪ **Espaces de noms – *Namespaces* :**

XML est extensible, c'est-à-dire qu'il permet de définir un nouveau format de document en associant et en réutilisant d'autres formats préexistants. ***XML Namespace*** est un mécanisme permettant d'éviter tout conflit lors de l'utilisation de plusieurs formats. Il permet de préciser et de qualifier les éléments de données et les attributs utilisés dans des documents XML en leur associant un **espace de nom** ou **espace de nommage**[43] (désigné par des références d'URI – *Uniform Resource Identifier*).

Les espaces de noms sont ainsi des ensembles d'éléments fermés, décrits de manière formelle, par exemple des jeux d'éléments standards ou des ensembles de valeurs. La déclaration de chaque espace de nom inclut son adresse et un préfixe devant les éléments correspondants ; ceux-ci restent donc toujours reliés à leur origine. [65]

[43] Le site <http://www.yoyodesign.org/doc/w3c/xml-namespace/Overview.html> met à disposition une traduction de la recommandation du W3C portant sur les espaces de nommage dans XML.

13 Technologie XML et documents XML : avantages, usages, contraintes

La galaxie XML – XML, XSLT, Xpath… - est désormais « devenue incontournable pour permettre à la grande hétérogénéité de l'informatique de converger ». [28]

Les **avantages de XML sont nombreux**…

- Lisibilité : théoriquement, le contenu d'un document XML doit être compréhensible sans connaissance préalable, et lisible sans recours à un logiciel spécifique. Un document XML est indépendant de toute plate-forme, système informatique ou logiciel. XML permet de véhiculer à la fois les données et les informations permettant de les réutiliser.
- Auto-description : un document XML, structuré, contient les informations à propos de sa propre structure logique et sémantique (« balises sémantiques »).
- Extensibilité : un document XML peut être utilisable en tous domaines d'application.
- Arborescence : sa structure sous forme d'arbre d'éléments permet plus aisément de modéliser la plupart des problèmes informatiques.
- Portabilité / universalité : XML prend en compte tous jeux de caractères (selon la norme Unicode).
- Déployable : XML est facilement distribué par les protocoles transportant du texte (ex : HTTP).
- Intégrabilité : la logique de structuration XML définissant séparément forme et contenu, autorise toute application pouvant « *parser* » - analyser le code XML, analyser sa syntaxe – à utiliser et récupérer les données. Le format XML permet d'afficher un même document sur des applications ou périphériques différents sans avoir besoin ou nécessité de créer autant de versions du document que de représentations désirées.

… et les **usages multiples** :

- créer des documents structurés (modèle de données hiérarchique, grammaire XML).
- méta-langage de structuration, ensemble de règles « ouvertes » de syntaxe, XML autorise quiconque à créer son propre langage, garantissant qu'il soit réutilisable par d'autres personnes.
- produire des métadonnées : les modèles de données hiérarchiques, normalisés XML, peuvent représenter des structures complexes. Voir le chapitre 11 de cette Première

partie du mémoire, relatif aux Schémas XML exprimant les contraintes grammaticales complexes. Un grand nombre de Schémas XML ont été élaborés, dans des domaines d'application multiples et variés.

- échanger des données (entre bases de données gérées par différents outils informatiques).
- transmettre le contenu de bases vers le client Web (ex : contenu d'une base MySQL [28]. Le format XML permet la mise à jour des données indépendamment de la présentation et des applications automatiques de traitement. XML permet le formatage des informations, qui peuvent être ensuite affichées en HTML ou XHTML.
- conserver à long terme des documents numériques ou données électroniques : pour tout type d'information – texte, son, image, vidéo. [65]

Mais il faut également noter des **contraintes**, qui sont dues principalement : [12]

- à la complexité des technologies et spécifications de la « nébuleuse » XML elle-même ;
- à la « grande mobilité » des normes, des méthodes et des outils associés ;
- aux compétences multiples à associer ;
- à l'ouverture nécessaire aux pratiques des utilisateurs ;
- au coût des transformations des dispositifs techniques ;
- au nécessaire accompagnement au changement, notamment par la formation.

14 Entrepôt de (méta)données

Un **entrepôt de données** ou ***datawarehouse*** en anglais, est « un concept propre à l'informatique décisionnelle partant du constat que les données issues de l'informatique de production (notamment des ERP) se prêtent peu à l'analyse décisionnelle. Les *datawarehouse* permettent des requêtes complexes s'appuyant sur plusieurs tables. Leurs données sont orientées métier, non volatiles, intégrées et datées. »[44]

A. Berdugo, R. Mahl et G. Jean [76] précisent, en 2002, que « le **mécanisme du *datawarehouse*** consiste à transformer des données en informations, il s'effectue en plusieurs étapes :

- organisation des données par métiers, sujets ou thèmes ;
- intégration des données de production, de gestion et des données extérieures ;
- historisation des données ;
- garantir la non-volatilité des données. »

Avec « l'émergence de formats de données semi-structurés (tels que XML), le stockage de documents dans un entrepôt centralisé est apparu de façon naturelle comme **une adaptation des entrepôts de données** »[45].

Un **entrepôt de données XML** peut consister en un référentiel central dans lequel sont dupliquées des données provenant de différentes sources (bases). Ces dernières alors demeurent et conservent leurs données d'origine. [60]

Un entrepôt peut également contenir à la fois des objets documentaires et leur description en métadonnées. [56]

Le principe des archives ouvertes repose lui aussi sur le principe d'**entrepôts de métadonnées**. Le moissonnage des métadonnées Dublin Core (DC) dans des « entrepôts d'archives ouvertes » constitue aujourd'hui un modèle technique de plus en plus utilisé, d'accès à des serveurs de données documentaires. Un protocole particulier d'exposition et de collecte automatique de ces métadonnées est utilisé : **OAI-PMH** – *Open Archives Initiative – Protocol for Metadata Harvesting*[46] [22] [44] [55] [65]

---

[44] Source : <http://archimag.com/fr/accueil-archimag/liens-utiles/glossaire.html?user_smilelexicon_pi1[letter]=3&cHash=eb46fd2f83> [consulté le 15 novembre 2009].

[45] Source : RAVAT Franck, TESTE Olivier, TOURNIER Ronan. Analyse multidimensionnelle de documents via des dimensions OLAP. Document numérique, 2007, vol. 10, n°2, p. 85-104. ISSN 1279-5127

[46] Cf. site de l'Open Archives Initiative <http://www.openarchives.org/>.

Deuxième partie

Le cas de Centre INFFO : association productrice d'informations et gestionnaire de données sur la formation professionnelle

1 Le contexte de l'étude

1.1 Le Centre INFFO : statut, missions, activités

Le **Centre INFFO – Centre pour le développement de l'information sur la formation permanente**[47] est une association loi de 1901[48], créée en 1976, et placée sous tutelle du ministre chargé de la formation professionnelle continue (FPC)[49].
Il se trouve ainsi rattaché à la Délégation générale à l'emploi et la formation professionnelle (DGEFP), elle-même placée sous l'autorité du ministre de l'Économie, de l'Industrie et de l'Emploi et mise à la disposition notamment du ministre du Travail, des Relations sociales, de la Famille et de la Solidarité.

Son conseil d'administration est composé de représentants de l'Etat, des partenaires sociaux, des Régions et de personnalités qualifiées.
Le Centre INFFO « constitue l'échelon national de l'information et de la documentation dans le domaine de la FPC. Il a pour missions dans le cadre des politiques publiques en la matière tant nationale qu'européenne ou territoriales :
1°) D'élaborer, de capitaliser et de diffuser l'information et la documentation d'intérêt national, à destination plus particulièrement des pouvoirs publics et des partenaires sociaux. Cette mission est réalisée en liaison avec les dispositifs régionaux d'information ;
2°) De développer et de diffuser des supports d'information, de formation et de documentation à destination de l'ensemble des acteurs de la formation professionnelle. »

Le Centre a signé avec le Ministre chargé de la formation professionnelle un contrat pluriannuel d'objectifs, qui rappelle que sa mission de service public est fondée sur la conception, la capitalisation et la diffusion au niveau national de produits et de services d'information sur la FPC (droit, pratiques et environnement). Cette mission s'articule autour de trois métiers : informer, expertiser et former.

L'activité de veille permanente du Centre dans le domaine de la formation professionnelle apporte une lisibilité sur les systèmes, les pratiques et l'intervention des différents acteurs de

---

[47] Portail du Centre INFFO : <http://www.centre-inffo.fr/>.

[48] Voir également en annexe un diagramme rappelant l'organisation fonctionnelle du Centre INFFO et les effectifs employés.

[49] Décret n° 2003-479 du 4 juin 2003 modifiant le décret n°76-203 du 1er mars 1976 relatif au Centre pour le développement de l'information sur la formation permanente, paru au Journal Officiel du 6 juin 2003, page 9635.

la FPC et de l'apprentissage. Mais le contexte de sa mission de service public est en profond changement.

Parmi les évolutions majeures ces toutes dernières années : la poursuite du mouvement de décentralisation de la FPC (2004), la loi relative à la formation professionnelle tout au long de la vie (2004), le plan de cohésion sociale (2005), la loi de modernisation sociale (2002), la loi organique relative aux lois de finances (2001). En 2008 également : réforme du service public de l'emploi, modernisation du marché du travail, négociation des partenaires sociaux et de l'Etat en vue d'une réforme de la formation professionnelle, et Révision Générale des Politiques Publiques (RGPP). Et enfin en 2009, une nouvelle réforme qui a abouti à la toute récente loi du 24 novembre 2009 relative à l'orientation et à la formation professionnelle tout au long de la vie[50].

L'actuel contrat pluriannuel d'objectifs, qui couvre la période 2007-2010, implique notamment d'investir de nouveaux champs d'intervention et de nouveaux publics : les professionnels et relais d'information tournés vers l'insertion ou la réinsertion professionnelle, l'emploi, le développement des compétences, l'accès aux qualifications et le développement économique ; le grand public (indirectement via les relais, ou directement via le Portail national de l'orientation et de la formation <http://www.orientation-formation.fr/> dont le Centre INFFO est maître d'œuvre).

Parmi les nouveaux domaines d'intervention également : la valorisation de l'apprentissage, le renforcement de l'information des salariés, la formation des agents publics qui interviennent dans le secteur de la formation continue.

Parmi les actions à mener et prestations à développer : renforcer les partenariats avec les Régions, les acteurs de la décentralisation et les branches professionnelles, faire évoluer la politique de publication, renforcer les fonctions de veille technique et de capitalisation, développer des activités d'analyse et de conseil, développer le portail orientation (cité ci-avant), moderniser la gestion du Centre.

Le catalogue 2010 des produits et services du Centre INFFO[51] propose un panorama de la plupart des activités que propose le Centre :

- sessions de formation interentreprises ou sur mesure, portant sur le droit de la formation continue, ou les pratiques en usage ou développement dans le domaine,
- journées d'actualité ou d'information,
- séminaires,

---

[50] Loi n°2009-1437 du 24 novembre 2009 <http://legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000021312490>. Dossier législatif sur le site du Sénat <http://www.senat.fr/dossierleg/pjl08-578.html>.

[51] Le catalogue 2010 des produits et services du Centre INFFO est consultable à l'adresse <http://www.centre-inffo.fr/spip.php?page=zoom&id_article=4804>.

- revues d'actualité : une newsletter électronique « Le Quotidien de la formation », un bimensuel papier « L'Inffo formation », une revue d'étude bimestrielle « Actualité de la formation permanente »,
- des ouvrages, des guides techniques et juridiques, des thésaurus et dossiers documentaires, des dépliants, affiches et plaquettes d'information,
- et des sites Internet qui sont en développement selon un schéma intitulé « Galaxie Internet du Centre INFFO » (décrite au chapitre suivant).

1.2 Les orientations technologiques au Centre INFFO

1.2.1 Schéma Directeur Nouvelles Technologies 2008-2011

Un **Schéma Directeur Nouvelles Technologies 2008-2011** a été rédigé en mars 2008 par le responsable du Service Système d'Information Informatique. Il reste toutefois ouvert à des ajustements que nécessitent régulièrement les adaptations induites par les évolutions technologiques permanentes et les besoins générés par les activités du Centre. Ce schéma place l'interopérabilité au cœur de ses orientations.
Les axes de développement informatique déclinés évoquent notamment :
- le développement d'une base de recherche transversale (sur les articles des revues publiées par le Centre et les informations provenant de sites Internet du Centre),
- le renouvellement du progiciel de gestion documentaire actuel LORIS,
- un nouvel outil d'information sur le droit de la FPC (Base de données),
- le développement de la formation ouverte et à distance (dont la poursuite de bases de données sur le thème TIC et FOAD),
- une migration vers des logiciels libres (open source), notamment pour ce qui concerne la bureautique,
- le déploiement d'outils de travail collaboratifs.

1.2.2 Un développement croissant des activités Web

Un schéma esquisse la « Galaxie Internet » du Centre INFFO décrite au catalogue 2010 de ses produits et services[52]. Elle rappelle les sites Internet que gère déjà le Centre, comme son portail d'information <http://www.centre-inffo.fr/>, ou dont il est maître d'œuvre, comme le portail national orientation et formation <http://www.orientation-formation.fr/>. Figurent également des « sites dédiés » à des thématiques spécifiques.
Un site sur le droit de la formation <http://www.droit-de-la-formation.fr/> a été lancé au printemps 2009, intégrant et développant plus avant la couverture du droit de la formation

déjà opérée par le précédent portail du guide juridique « Les Fiches pratiques de la formation continue ».

Un autre site portant sur l'actualité de la formation <http://www.actualite-de-la-formation.fr/> a été quant à lui inauguré à la mi-juin 2009, marquant l'intégration des articles publiés dans la lettre électronique « Le Quotidien de la formation » et des articles provenant de la revue bimensuelle « INFFO Flash » devenant désormais « L'Inffo formation », et jusque-là publiée dans son intégralité uniquement en version papier.

D'autres sites Web, en projet, sont appelés à être lancés tout prochainement : un site consacré aux pratiques de formation, et une encyclopédie de la formation.

La « Galaxie Internet de la formation » est ainsi présentée, à terme, comme « un ensemble de sites complémentaires, articulés entre eux, spécialisés et conçus à partir des besoins des utilisateurs du Centre INFFO ».

Enfin, des blogs thématiques pourraient être amenés à être développés, à l'instar de celui ouvert à l'occasion de la Présidence Française de l'Union européenne au second semestre 2008 <http://europe-et-formation.eu/>.

1.3 Réseau informatique interne et Intranet

1.3.1 Services accessibles sur le réseau interne

Chaque salarié du Centre dispose d'un accès au Web, ainsi que d'une messagerie électronique.

Le réseau informatique interne autorise un accès à différentes parties – répertoires et sous-répertoires – des serveurs (internes) selon la nature de l'activité du salarié concerné.

Un espace intitulé « Public » où chaque salarié a un répertoire librement accessible à tous. Un espace « Projets » où les salariés ne peuvent accéder qu'aux documents relatifs aux projets dans lesquels ils sont impliqués. Un espace « Services » enfin, où chaque service, département, mission du Centre INFFO a un espace dédié seulement accessible à ses membres.

De nombreuses bases de données sont produites par le Centre INFFO, certaines hébergées en interne, d'autres accessibles sur des serveurs externes via Internet.

1.3.2 Intranet

Un Intranet a été mis en place depuis mars 2006, développé en interne. Il ne constitue pas en tant que tel un service, mais est animé par une équipe de quatre personnes qui sont à l'origine du projet.

---

[52] Cette « Galaxie Internet » est décrite pages 14 et 15 <http://www.centre-inffo.fr/spip.php?page=zoom&id_article=4804>.

Il s'est peu à peu étoffé de développements informatiques internes, notamment pour gérer certaines bases de données. Il récupère certains flux d'informations provenant de sites Web de Centre INFFO, publiés quant à eux à l'aide du logiciel SPIP.
L'Intranet ne constitue toutefois pas un portail d'accès unique aux différents outils de production ou de gestion, ni au réseau informatique interne.

1.3.3 Intranet documentaire

Sur l'Intranet, une rubrique « Sources et ressources documentaires » autorise un accès « centralisé » pour consultation : à des produits documentaires dont différentes Bases de données diffusées sur le Web via OPAC (Base de données bibliographiques, base des accords de branche), sites d'actualités quotidiennes sous logiciel de gestion de contenu SPIP (articles journalistiques du « Quotidien de la formation », portail juridique, actualités internes rédigées par la documentation), de même que des documents variés pouvant constituer une ressource documentaire sur la formation (Bibliographies, Webographies, notes de synthèse, revues de presse).
Ce site ne constitue pas en tant que tel un portail documentaire qui constituerait un point d'accès unique aux bases documentaires de production, autoriserait une recherche simultanée sur différentes sources, et permettrait une diffusion sélective de l'information automatisée, par profils.
En développement depuis 2006, l'Intranet s'étoffe en revanche petit à petit d'outils de production, développés en PHP sous forme de bases MySQL. Notamment : un outil de création de lexique, une base de production de données documentaires sur l'offre de formation (remplacement de bases sous le logiciel DORIS, utilisé jusqu'en 2007), une base des dispositifs régionaux pour la formation et l'apprentissage, un outil de gestion d'annuaire d'adresses, un espace de ressources pédagogiques et un agenda partagé.

2 Etude des modèles de données gérées dans les bases du Centre INFFO et processus métiers : problématiques majeures et objectifs visés, méthode utilisée, résultats attendus

2.1 Problématiques majeures et objectifs visés

2.1.1 La gestion d'informations dans des bases… qui sont à redécouvrir

L'une des missions du Centre INFFO – Centre pour le développement de l'information sur la formation permanente – est d'intégrer, de traiter, de consolider les informations sur la formation professionnelle, pour les présenter sous différents formats et supports de publication.

A cet effet, le Centre INFFO produit de nombreuses bases de données, riches par la variété des objets gérés, mais hétérogènes quant à leur structure, quant aux applications qui les gèrent, et non connectées entre elles pour la plupart.
Afin d'améliorer l'alimentation des données, d'apprécier la complémentarité des ressources, de mieux les exploiter, soit de valoriser davantage ce capital informationnel, il apparaît aujourd'hui nécessaire de faire le point sur les données gérées, autrement dit les « métadonnées » renseignées sur ces objets, et les éléments composant le Système d'Information Documentaire (SID).
De concert, il est tout aussi nécessaire de se pencher sur les processus métiers induits par les activités autour de ces bases, supports de la production d'informations qui sont au cœur des missions du Centre.

2.1.2 Enquêter sur des bases, analyser un système, valoriser les ressources

Ce mémoire rapporte une démarche d'analyse du SID, dans le but de l'optimiser et de valoriser les ressources gérées, visant donc à la fois :

- une présentation globale de la base : les objets documentaires gérés, l'évolution récente, la plateforme technique sollicitée ;
- une description des métadonnées et de leur organisation, par une modélisation de la structure de chaque base étudiée, autant que l'accès à leur structure et/ou à une documentation existante (cahier des charges, format de données,…) ait été permis ;

- une schématisation du workflow pour chaque base étudiée : les flux de travail des documentalistes et « producteurs » d'information, à l'œuvre dans l'alimentation et l'exploitation des données.

L'étude de l'organisation, de la structuration des métadonnées portera sur quelques bases de production significatives des ressources gérées à Centre INFFO, et qui sont actuellement produites sous différentes applications. Notamment :

- des bases sous logiciel LORIS (de références bibliographiques et textes conventionnels) ;
- une base sous logiciel développé par la société e-Motive (description de ressources pédagogiques multimédias) ;
- plusieurs applications MySQL développées et publiées en interne, ou produites à l'aide du logiciel SPIP (descriptions d'organismes de formation ou diplômes ; articles journalistiques quotidiens ; fiches descriptives de dispositifs et aides de Conseils régionaux ; annuaire d'organismes du domaine de la formation professionnelle) ;
- des bases FileMakerPro, dont les données sont en train ou sont a priori appelées à migrer sous MySQL ;
- une base de fiches juridiques accessible à l'aide du serveur d'applications ZOPE (sous SGBD PostgreSQL).

Le projet poursuivi par Centre INFFO est en effet la reprise, c'est-à-dire la refonte de ces bases disparates et/ou le reformatage de leur contenu selon les cas.
Il est notamment prévu l'acquisition d'un outil open source, pour remplacer particulièrement le logiciel de gestion documentaire LORIS, dont la version ancienne est rapidement appelée à n'être plus maintenue par son éditeur EVER Ezida. L'application choisie permettrait non seulement d'assurer les fonctions de gestion du fonds du centre de documentation, mais aussi pourrait vraisemblablement assurer une fonction de gestion électronique des documents (GED), plus globalement, pour les documents produits et/ou utilisés par Centre INFFO.

<u>Equipement logiciel actuel du Département Documentation :</u>
Le centre de documentation est équipé d'une version de LORIS datant de 2006. Il s'agit d'un logiciel propriétaire fermé ne prenant pas en compte les technologies Web, et ne correspond plus aux attentes multiples de la Documentation, en matière de traitement, de gestion et de diffusion des informations, notamment des ressources numériques. Une première ébauche d'expression des besoins a été rédigée début 2008. Le projet de changement de ce logiciel

par une solution *« full Web »*[53] permettant d'intégrer les applications gérées par le Département Documentation est inscrit depuis mars 2008 dans un Schéma Directeur Nouvelles Technologies 2008-2011. Ce schéma est actuellement en cours de révision par le Service Informatique.

2.2 Cheminement dans l'investigation : point méthodologique

La mission, qui fait l'objet de ce rapport, support de mémoire professionnel, consiste en fait à suivre un chemin inverse à la conception de base de données, procédant d'une **ingénierie inverse** *[reverse engineering]*. Nous pouvons procéder selon les étapes suivantes.

➢ Effectuer un **recensement** des principales bases de données de production produites par trois grands métiers du Centre INFFO : documentalistes, rédacteurs et juristes.

➢ Interroger ces bases afin d'en comprendre la structuration, c'est-à-dire pour chacune d'entre elles :

• évaluer le **format de données** (format documentaire informatique) et le présenter sous forme de tableau ;

• repérer si des **normes** sont respectées ou devraient l'être, si des **standards** sont suivis ;

• présenter sous forme de graphe les liens entre les **principales tables** constituant chaque base.

➢ Rencontrer en interne, en parallèle, les **utilisateurs**[54] ou gestionnaires de ces bases (parfois administrateurs restreints, et utilisateurs des métadonnées) afin de mieux comprendre quels sont les ensembles d'**actions** spécifiques, c'est-à-dire les **activités**, qui sont effectuées dans le cadre des **processus** à l'œuvre.

• obtenir des informations sur ces bases, voire des documents complémentaires (cahiers des charges, notes d'opportunité ou de recensement de besoins...) ;

• comprendre comment est organisée la production : sélection et collecte des données, analyse, rédaction, etc. ;

• éventuellement repérer le volume d'information concerné permettant d'en estimer la portée ;

---

[53] « Autorisant une recherche simultanée sur plusieurs ressources, permettant une diffusion sur profil automatisée, voire même la gestion de la numérisation de documents de bout en bout ».

[54] Appelés « utilisateurs standard » par Nicolas Larrousse [6] c'est-à-dire ceux qui utilisent le système d'information, qui y ont accès grâce aux « vues » définies par le concepteur de la base, « schémas de visualisation » différents selon les droits ou besoins des utilisateurs concernés.

• connaître l'utilisation, c'est-à-dire l'exploitation qui est faite aujourd'hui des ressources répertoriées : pour publication, pour diffusion ou par (ré)utilisation dans d'autres productions ;
• suivre le flux de données[55] et repérer les flux de travail concernés autour de chaque base (chaînage de production) : au moyen d'une représentation schématique du workflow depuis l'alimentation en données jusqu'à leur exploitation et publication, permettant de suivre le flux d'information depuis l'entrée (*input*) jusqu'à la sortie (*output*) du système d'information ;
• repérer ce qui pose problème dans cette chaîne des étapes de la production, par exemple les éventuelles ruptures de chaîne dans les processus métiers concernés ;
• et enfin relever les éventuels besoins en métadonnées complémentaires qui amélioreraient les produits et services en sortie.

➢ Faire le point sur les méthodes et technologies dans le domaine info-documentaire, actuellement disponibles, et utilisées pour structurer les (méta)données, favoriser l'interopérabilité des ressources, leur accès, leur exploitation, leur représentation, et valoriser les fonds.

2.3 Rapporter des constats, des besoins ; formuler des idées, des hypothèses ou des propositions

A l'issue de ces travaux qui visent à éclairer sur les contenus gérés par le système d'information du Centre INFFO, l'on pourrait émettre des **hypothèses**, formuler des **propositions**. Ces réflexions portent sur plusieurs axes.

➢ Repérer les éventuelles relations existantes entre bases.

➢ Evaluer la cohérence de la structuration des informations de chaque base, et évaluer les contenus d'informations renseignés.

➢ Apprécier la complémentarité des ressources et de leurs métadonnées.

➢ Rechercher une interopérabilité des ressources, en estimant le besoin d'utiliser d'autres formats de données, en étudiant l'opportunité d'adopter des modèles de données déjà standardisés ou normalisés, particulièrement en ayant recours au format XML et à des schémas de métadonnées qui respectent les principes de ces standards.

---

[55] Les **flux de données** - *data flow* sont des « transferts de données résultant de l'exécution de tout ou partie d'une action vers une autre action ». [81]

➢ Et par l'adoption de tels nouveaux schémas structurés, optimiser l'exploitation des métadonnées[56], en s'assurant pour l'avenir, notamment :

• la possibilité d'une interrogation multibases (mettant à profit XML comme format pivot) ;

• la possibilité d'une diffusion multicanale des informations (publication Web, PDF, flux RSS...) ;

• une migration future des données facilitée ;

• et éventuellement le stockage même des données en XML natif.

[56] L'exploitation des métadonnées « n'est pas un processus isolé, mais le moyen par lequel la gestion documentaire s'intègre à l'ensemble du système d'information et aux différents référentiels de l'entreprise ». [18]

Troisième partie

Investigation des bases de données de production de Centre INFFO

1 Recensement des Bases de données de production : documentaires et autres

Après enquête auprès de services producteurs et entretiens menés auprès d'utilisateurs et/ou gestionnaires, nous avons pu dresser ci-dessous un tableau recensant les **bases de production de Centre INFFO**, qui permet d'en offrir une vue synoptique. N'y figurent pas de nombreuses bases MySQL contenant les données publiées sur les sites Web produits par le Centre à l'aide du logiciel de publication de contenu **SPIP** (Système de publication pour l'Internet partagée <http://www.spip.net/>).

Nom de la base et/ou Type d'objet	Type de base	Service gestionnaire	Utilisation des données Publication sur le Web
Références bibliographiques	Logiciel LORIS	Documentation	• **Base bibliographique** référençant Périodiques, Articles de périodiques et Ouvrages <http://www.centre-inffo.fr/Base-bibliographique.html>, dont sont issues les données de la : • **Base Articles des revues INFFO Flash** et **Actualité de la formation permanente** <http://www.centre-inffo.fr/Base-des-revues-Actualite-de-la.html>.
Textes conventionnels : - Accords de branches et - Accords interprofessionnels	Logiciel LORIS	Documentation	• **Base CPNFP** [Comité Paritaire National de la Formation Professionnelle] **sur la négociation collective en matière de formation professionnelle** <http://www.centre-inffo.fr/Base-textes-conventionnels.html>.
Etat des négociations de branche	Base FileMakerPro (en cours de changement ?)	Documentation	• **Suivi de l'actualité de la négociation de branche en matière de formation** : liste des accords de branche publiés au Bulletin Officiel des Conventions Collectives (BOCC), ou des accords signés mais non encore publiés et accessibles au format PDF. Deux accès à cette liste : <http://www.centre-inffo.fr/spip.php?page=accords> et <http://www.droit-de-la-formation.fr/spip.php?page=accords>.
Thésaurus de la formation	SGBD Pick (accessible via Wintegrate)	Documentation	• **Outil maison** développé sous Pick (en basic) pour la **gestion du Thésaurus de la formation**, dans ses deux dernières éditions 2000 puis 2006, qui a fourni les données pour la publication papier. Application pour l'instant figée. [cf. sous-chapitre suivant 2.4.4]. Une version des données 2006 est aussi disponible sous tableur Excel (sans historique des versions antérieures du thésaurus).

Nom de la base et/ou Type d'objet	Type de base	Service gestionnaire	Utilisation des données Publication sur le Web
VetDet – Vocational Education and Training Data Entry Tool	Logiciel MS-Access	Documentation	• Références bibliographiques consultables en ligne, intégrées à la **base VetBib du CEDEFOP** – Centre européen pour le développement de la formation professionnelle <http://libserver.cedefop.europa.eu/F?local_base=ced01>.
Produits pédagogiques multimédias pour la formation des adultes	Application développée par eMotive	Documentation	• **Base des Produits pédagogiques multimédias** pour la formation des adultes <http://www.centre-inffo.fr/La-base-de-donnees-de-produits.html> encore hébergée sur le site Web <http://ressourcesv2.e-motive.com/frontblocks/produits/> de l'ancien organisme ALGORA. Fiches descriptives de produits ou ressources Web, et pour certaines d'entre elles, accompagnées de fiche(s) d'évaluation. • Une **sélection thématique des nouveautés recensées** est également proposée environ tous les deux mois : <http://www.centre-inffo.fr/Selection-thematique-des,2683.html> • Des **dossiers documentaires** au format PDF valorisent les produits recensés par thèmes : **Etat des lieux des ressources pédagogiques** <http://www.centre-inffo.fr/Etat-des-lieux-des-ressources.html> ; **Fichiers doc express** <http://www.centre-inffo.fr/Fichiers-express.html> Les références de la BDD sont également consultables via : • la **BCRP – Base coopérative d'information sur les ressources pédagogiques**, qui constitue un métamoteur de recherche <http://www.centre-inffo.fr/BCRP-Base-Cooperative-d.html> sur quatre bases de données réalisées par différentes institutions (partenariat Centre INFFO – Agrosup Dijon/EDUTER, les CNR – Centres Nationaux de Ressources de l'Éducation Nationale, et le CARIF de la région Lorraine INFFOLOR). • et, pour certaines données sur la formation professionnelle automobile : sur le **site educauto.org** du Centre national de ressources pour la formation automobile (partenariat entre l'Education nationale et l'ANFA – Association nationale pour la formation automobile) qui propose un formulaire de recherche <http://www.educauto.org/OutilsMultimedia/multimed.php>.
Application **NOEL – Nouvelle Offre en Ligne**	Base MySQL	Documentation	• **Organismes de formation** proposant des formations dans le domaine des métiers de la formation et de l'insertion professionnelle <http://www.centre-

Nom de la base et/ou Type d'objet	Type de base	Service gestionnaire	Utilisation des données Publication sur le Web
			inffo.fr/spip.php?page=organismes-de-formation>. • **Titres et Diplômes** dans le champ des métiers de la formation des adultes <http://www.centre-inffo.fr/spip.php?page=stages-metiers-formation>. • **Stages agréés par l'Etat** au titre de la rémunération au plan national <http://www.centre-inffo.fr/spip.php?page=stages-agrees>.
CQP – Certificats de Qualification Professionnelle	Base MySQL (ex Base FileMakerPro)	Documentation	• Liste de **CQP – Certificats de Qualification Professionnelle** par branche d'activité professionnelle accessible sur le site **Certif info**, référentiel national de certifications <http://www.certifinfo.org/>.
Base de données **Métiers de branche**	Base MySQL (ex Base FileMakerPro)	Multimédia / Documentation	• Affichage des métiers des branches professionnelles, en correspondance avec des domaines de formation (Formacode) sur le Portail Orientation et Formation <http://www.orientation-formation.fr/>.
Gestion du **thésaurus Formacode de l'offre de formation**	Base MySQL	Documentation	• Publication en version papier du **Formacode® 9.0** (édition 2008) ; • Publication en version informatique **Formacodex V9.0** (édition 2008) • **Base Formacode** consultable en ligne <http://www.centre-inffo.fr/Le-Formacode-R-thesaurus-de-l.html>.
Formation et Apprentissage en Régions (FAR)	Base MySQL	Mission Régions et Europe	• Recensement des aides régionales en vigueur en matière de formation professionnelle et d'apprentissage, ainsi que des contrats de projets et d'autres outils de programmation gérés par les Conseils régionaux. La **base FAR** est consultable en ligne : <http://www.centre-inffo.fr/spip.php?page=baseregions>.
Annuaire des structures et contacts : anciennement appelé « Adresses utiles »	Base MySQL	Direction Juridique Observatoire	• **Annuaire des acteurs de la formation** <http://www.centre-inffo.fr/spip.php?page=annuaire&recherche_libre=Annuaire> mais dont les données sont aussi utilisées pour : • le site **Droit de la formation** <http://www.droit-de-la-formation.fr/spip.php?page=adresses>. • et le **Portail Orientation-Formation** <http://www.orientation-formation.fr/>.
Webdoc : rédaction d'actualités sur la formation repérées sur le Web	Base MySQL produite via SPIP	Documentation	• sous logiciel SPIP, en diffusion interne, à l'exception des **Textes officiels** diffusés au jour le jour sur le site **Droit de la formation** <http://www.droit-de-la-

Nom de la base et/ou Type d'objet	Type de base	Service gestionnaire	Utilisation des données Publication sur le Web
			formation.fr/spip.php ?page=textesofficiels-rubrique>.
Le Quotidien de la formation : articles de périodique d'actualité	Base MySQL produite via SPIP	Agence d'information	• **Lettre électronique quotidienne** diffusée sur abonnement par courriel. • Diffusion d'une sélection d'actualités sur le **site officiel du Centre INFFO** <http://www.centre-inffo.fr/-Actualite-du-quotidien-de-la-.html>. • Diffusion depuis juin 2009 sur le site **Actualité de la formation** <http://www.actualite-de-la-formation.fr/>
Juriweb : base de production du guide juridique « Les Fiches pratiques de la formation continue »	Via le serveur d'application ZOPE (SGBD PostgreSQL)	Direction Juridique Observatoire	• **Base Fiches Pratiques** sur le site **Droit de la formation** <http://www.droit-de-la-formation.fr/> • Publication papier annuelle **« Les Fiches pratiques de la formation continue »** (sous la forme de deux volumes) <http://www.centre-inffo.fr/-Les-Fiches-pratiques-de-la-.html>.
Recodification	Base MySQL	Direction Juridique Observatoire	• Rubrique **Recodification du Code du travail** (partie législative) sur le site **Droit de la formation** : <http://www.droit-de-la-formation.fr/?Recodification-du-code-du-travail,1030>.
Droit de la formation	Base MySQL produite via SPIP	Direction Juridique Observatoire	• sous logiciel SPIP, production du site **Droit de la formation** <http://www.droit-de-la-formation.fr/?-Actualite-juridique->, comprenant notamment une rubrique **Actualité juridique** (présentation détaillée de textes officiels). S'y adjoint la **mise à jour des « Fiches pratiques de la formation continue »** provenant de la base Juriweb. • **alerte mensuelle par courriel** : Fiches Pratiques mises à jour, dossier du mois, Actualités juridiques, calendrier, mini guide.

Dans la suite du chapitre, les bases qui ont fait l'objet d'une enquête pour les besoins de cette étude sont présentées l'une après l'autre. Celles dont le **nom** est surligné en **jaune** ont fait l'objet d'une enquête moins approfondie et figurent en annexes du mémoire. Tandis que celles dont le **nom** est surligné en **gris** n'ont pas fait l'objet d'une investigation.
Les éléments descriptifs suivants sont proposés, dans la mesure de leur disponibilité, et après recueil effectué auprès des personnes gestionnaires et utilisatrices.

- une présentation de la base
- sa structure (principales tables, champs) ainsi qu'un graphe des liens entre tables lorsque cela est possible. La modélisation des données proposée est alors présentée sous la forme d'un schéma simplifié basé sur **UML – *Unified Modeling Language***, « langage de modélisation objet unifié »[57] (type diagramme de classes sans indication des cardinalités)
- une schématisation du workflow depuis l'entrée des données jusqu'à leur exploitation et publication (schéma du processus métier).

[57] Sur le standard UML, qui propose un modèle conceptuel de représentations des données, voir par exemple le site <http://uml.free.fr/index.html>.

2 Base de données LORIS

2.1 Présentation de la base LORIS

Nom de la base Type d'objet	Type de base	Service gestionnaire Utilisateurs	Utilisation des données Publication sur le Web
Références bibliographiques	logiciel LORIS	Documentation Utilisateurs : 1 Bibliothécaire + 10 documentalistes	• **Base bibliographique** référençant Périodiques, Articles de périodiques et Ouvrages <http://www.centre-inffo.fr/Base-bibliographique.html>, dont sont issues les données de la : • **Base Articles des revues INFFO Flash** (devenu **L'Inffo formation** en 2009) et **Actualité de la formation permanente** <http://www.centre-inffo.fr/Base-des-revues-Actualite-de-la.html>.
Textes conventionnels : - Accords de branches et - Accords interprofessionnels	logiciel LORIS	Documentation Utilisatrice : 1 documentaliste	• **Base « CPNFP »** [Comité Paritaire National de la Formation Professionnelle] **sur la négociation collective en matière de formation professionnelle** <http://www.centre-inffo.fr/Base-textes-conventionnels.html> puis <http://www4.centre-inffo.fr/v2/cpnfp/recherche.phtml>.

La base de données bibliographiques, gérée sous le logiciel LORIS, référence différents types de ressources décrites selon le même format UNIMARC, mais que l'on peut distinguer au moyen des différentes grilles de saisie. Le tableau ci-dessous les liste ainsi que le nombre de notices actuelles.

Nom de la grille de saisie	Nombre de notices (au 16/11/2009)	Commentaire
LIV	9692	Ouvrages
PER	665	Titres de périodiques
REV	17194	Articles de périodiques : y compris revues publiées par le Centre INFFO : le bimensuel « INFFO Flash », devenu en 2009 « L'inffo formation », et le bimestriel « Actualité de la formation permanente ».
ACC	2831	Textes conventionnels : accords de branche et accords interprofessionnels.
LEG	2970	Textes législatifs et réglementaires
JUR	1	Jurisprudence.

2.2 Références bibliographiques : ouvrages et articles de périodiques

La base de données bibliographiques produite par le Département Documentation du Centre INFFO recense plus de 25 000 références d'ouvrages, d'études, d'articles de revues sur la formation professionnelle continue et son environnement.

Pour chacune de ces références, elle propose un descriptif (titre, auteur(s), nom de la revue et/ou éditeur, année, date, pagination), enrichi de descripteurs (mots-clés du Thésaurus de la formation 2006) et d'un résumé.

Cette base de données est historique et propose des informations à partir de 1969.

Un OPAC (*Online Public Access Catalog*) proposé par le logiciel LORIS, permet sa consultation via le site Web du Centre INFFO <http://www.centre-inffo.fr/Base-bibliographique.html>. A noter : l'achat en ligne de certains ouvrages est facilité par un lien proposé sur le site de la Librairie Lavoisier (indiqué sur la référence).

2.2.1 Revues publiées par le Centre INFFO

La revue bimestrielle « **Actualité de la formation permanente** » est publiée depuis 1972.
Le bimensuel « **INFFO Flash** », publié depuis 1976, a été rebaptisé en 2009 « **L'Inffo formation** ».
Ces deux périodiques font l'objet d'une base de données publiée séparément du reste du fonds bibliographique. Elle recense près de 5000 articles <http://www.centre-inffo.fr/Base-des-revues-Actualite-de-la.html>.

Chacune des références présentées dans la base propose un descriptif (titre, auteurs(s), numéro de la revue, pagination, année de publication) enrichi de descripteurs (mots-clés issus du Thésaurus de la formation) et d'un résumé.

Les notices des articles de la revue « Actualité de la formation permanente » publiés à partir de 1999 sont complétées par le texte original en format PDF. Les textes originaux des articles d'« INFFO Flash » sont proposés en format PDF à partir de l'année 2003. Les articles des deux dernières années 2008 et 2009 sont simplement signalés par une notice descriptive, ces revues faisant l'objet d'une commercialisation.

2.2.2 Accroissement du fonds documentaire : ouvrages et articles de périodiques

	2004	2005	2006	2007	2008
Accroissement des ouvrages [1]	+ 427	+ 394	+ 352	+ 243	+ 276

	2004	2005	2006	2007	2008
Accroissement des articles de périodiques (hors revues du Centre INFFO) [2]	+ 998	+ 774	+ 561	+ 294	+ 301
Accroissement des articles des revues du Centre INFFO : « INFFO Flash » et « Actualité de la formation permanente »	+ 318	+ 342	+ 587	+ 525	+ 827

[1] [2] Cette diminution continue du nombre d'articles de revues et d'ouvrages référencés chaque année dans LORIS correspond à une politique systématique d'allègement de la base de données.

2.2.3 Constats sur le fonds bibliographique (hors textes conventionnels)

La fonction GED n'a été utilisée, quasiment, que pour les articles de revues publiées par Centre INFFO, jusqu'à il y a peu de temps. Avec l'apparition à la mi-2009 du site « Actualité de la formation » <http://www.actualite-de-la-formation.fr/> devant contenir les articles à venir du Quotidien de la formation et de l'Inffo formation, mais aussi reprendre les articles publiés dans INFFO Flash, cette fonction GED est de moins en moins sollicitée.

Le traitement intellectuel des articles de revues s'appauvrit. Cela correspond à une politique systématique mise en place il y a trois ans. En effet, de moins en moins d'articles de revues, hormis ceux publiés par Centre INFFO, font l'objet d'un traitement documentaire. Et pour ceux qui sont sélectionnés, ils ne font pas tous l'objet d'un résumé, seule une indexation étant systématiquement effectuée.

La production de bibliographies et de dossiers documentaires :

Certaines productions nécessitent, en plus des résultats de requête dans la base LORIS, de recourir largement à des compléments d'informations extraits d'autres bases de données internes (Webdoc, le Quotidien de la formation, les Fiches Pratiques de la formation continue…), et issues d'Internet – bases de données externes, sites Web, revues en ligne, etc. Les documents primaires téléchargeables sur Internet, sélectionnés et exploités ne font alors pas systématiquement l'objet d'un référencement dans LORIS.

Les produits documentaires, bibliographies et dossiers, du Centre sont pour une grande part enregistrés sous forme de notices dans la base LORIS, mais ne sont pas attachés en GED en tant que tels. Pour les dossiers documentaires, cela tient au fait que jusqu'à peu, ils n'étaient pas produits pour diffusion au format numérique, seule la biblio-webographie étant au format Word.

La production de dossiers documentaires directement en PDF via Acrobat 9 Professionnel a commencé en 2009, et permet de faire évoluer les choses.

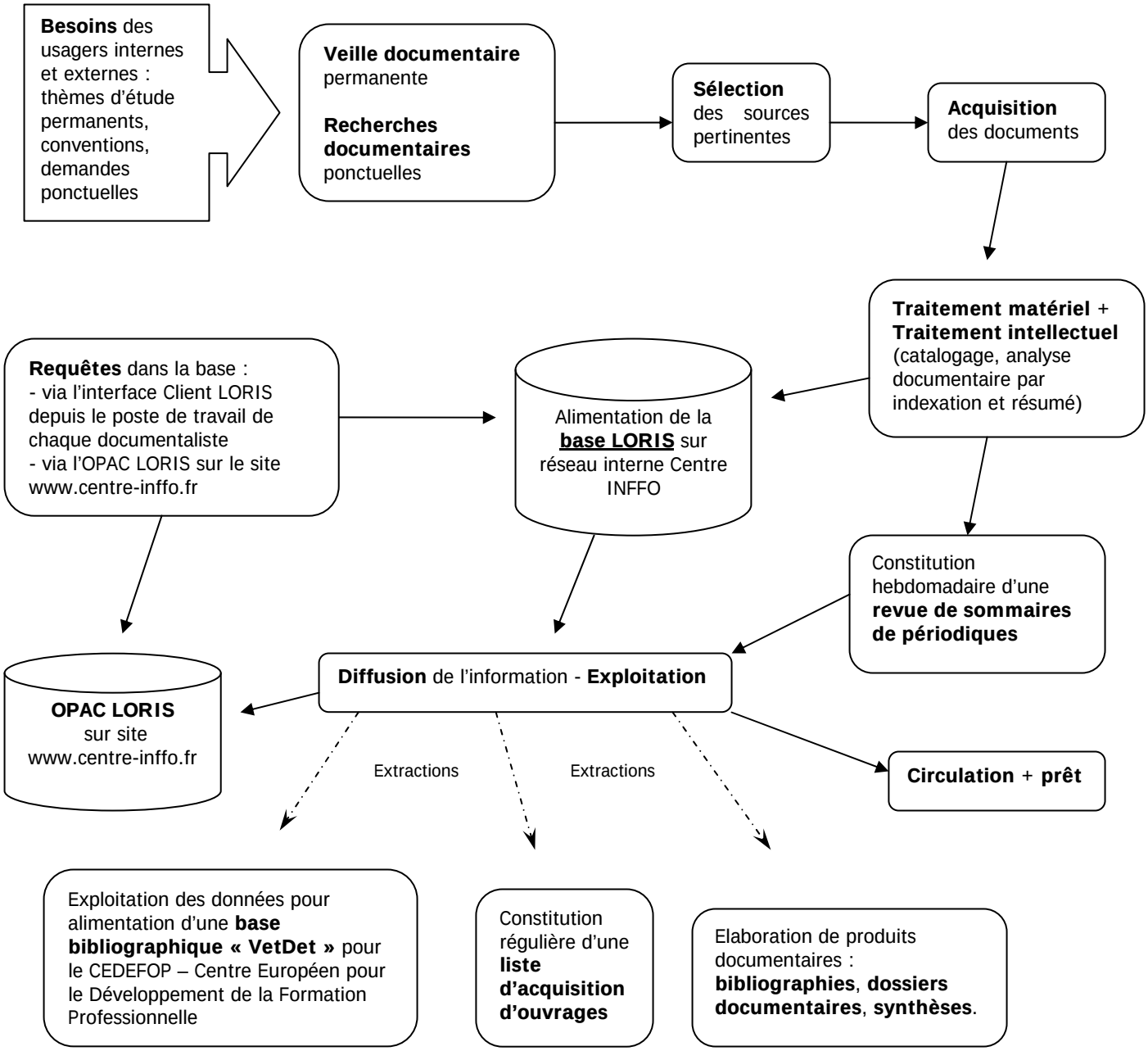
Publication en ligne de produits documentaires, situation actuelle :

- sur le site Intranet pour quelques bibliographies, aux formats PDF dynamique et HTML : une petite étude menée en 2008 par l'équipe Intranet a permis de formaliser quelques règles à observer, de façon à pouvoir correctement convertir les fichiers Word initiaux. Sont ainsi optimisées leur recherche via le moteur du site Intranet (nom du fichier, métadonnées du champ propriétés) et leur présentation différente selon le format (table des matières, en-tête et pied de page, pagination…).
- sur le site Internet de Centre INFFO pour certains dossiers documentaires ou webographies, en rubrique « Ressources documentaires » ou dans d'autres rubriques selon les thématiques (Région, Europe, Formation ouverte et à distance, site en anglais, blog Europe & formation, etc.) au format PDF statique ou dynamique.
- Toutefois, ces hébergements ne sont pas encore systématisés, que ce soit sur Intranet ou sur Internet.

Bibliographies dynamiques : il y aurait nécessité de produire des bibliographies dynamiques par requêtes prédéfinies dans la base LORIS.

Chaîne de production de dossiers documentaires papier et numériques : voir deux schémas qui illustrent l'utilisation des informations internes et externes, en annexe 5 de ce mémoire.

2.2.4 Workflow relatif au traitement du fonds bibliographique (hors textes conventionnels)



2.3 Fonds juridique

2.3.1 Textes législatifs et réglementaires - Jurisprudence

Historiquement, avant 2001, la base de données bibliographiques était gérée sous logiciel de gestion de bibliothèque OPSYS. Lors de la migration des données, les notices descriptives de **textes législatifs et réglementaires** produites dans les années 90, qui étaient alors présentes, ont été reprises sous LORIS mais l'alimentation n'a plus été continuée.

De même une grille de saisie de notices descriptives pour les **textes jurisprudentiels** avait été créée, elle aussi reprise, mais n'a jamais été utilisée.

2.3.2 Textes conventionnels dont les accords de branches

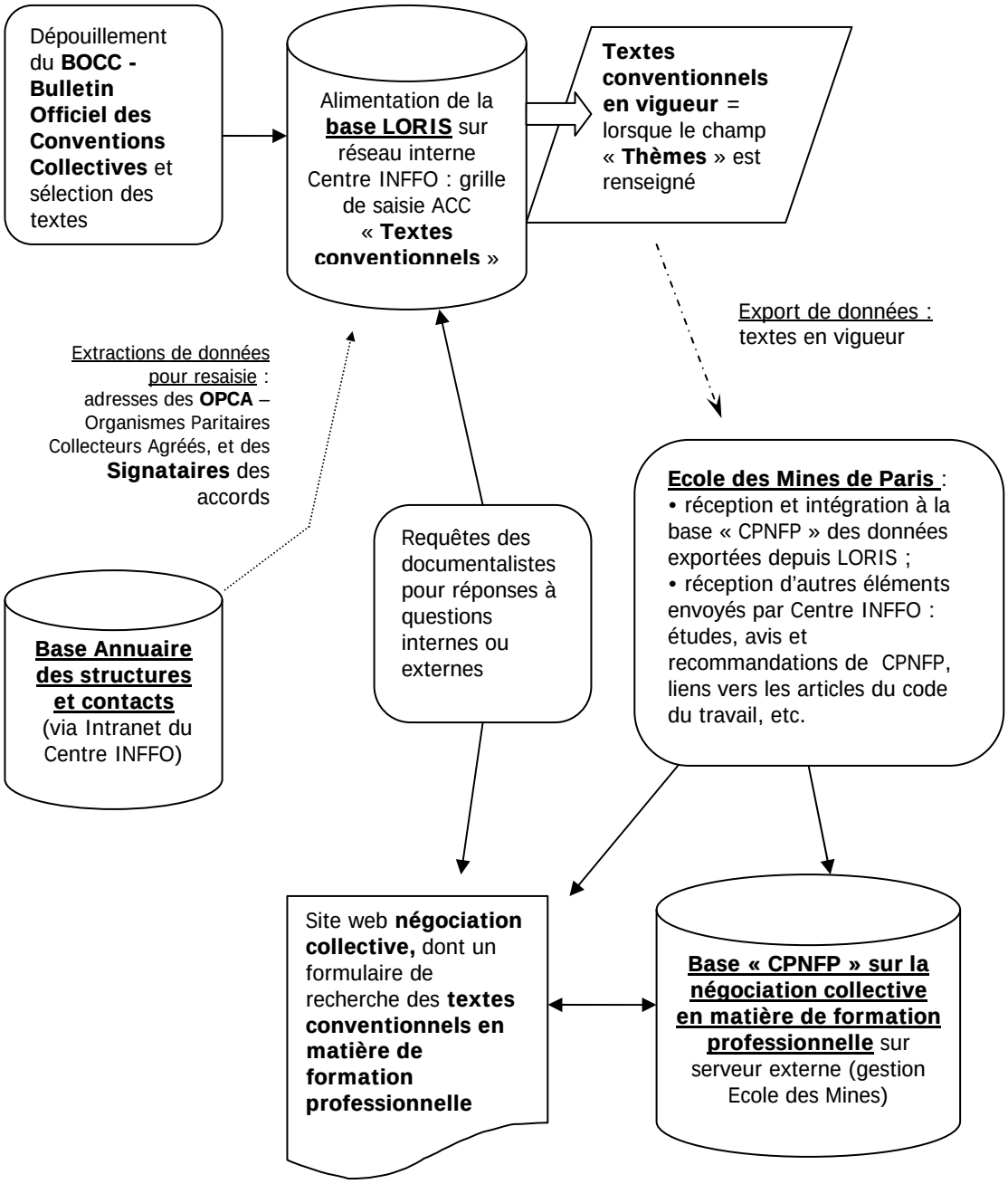
A partir de 1998, le Centre INFFO a constitué pour le compte du **CPNFP – Comité Paritaire National de la Formation Professionnelle** – une **base de données sur la négociation collective en matière de formation professionnelle**.

C'est pourquoi dans LORIS, une grille spécifique a été élaborée pour saisir les **textes conventionnels**, qui comprennent : les **accords de branche** ainsi que les **accords interprofessionnels** portant sur la formation professionnelle, l'emploi (égalité professionnelle, séniors, gestion prévisionnelle des emplois et des compétences, handicapés,...), la qualification professionnelle, et la certification (création de diplômes).
Le format UNIMARC ne permettant pas de décrire parfaitement ces textes, certains champs ont été « détournés ». Ce fut notamment le cas pour l'indexation supplémentaire au moyen de thèmes, et l'indication des organisations signataires des accords. Pour l'avenir, en cas de changement de format des données, il faudra veiller à la reprise des contenus de ces zones UNIMARC dont l'utilisation a été détournée.
En outre, les accords de branche sont indexés à l'aide de codes et libellés de la NAF – Nomenclature des Activités Françaises répertoriés dans une table dédiée, liée à une table référençant les CCN – Conventions Collectives Nationales.

La **base de données sur la négociation collective en matière de formation professionnelle** (« accords de branche ») accessible depuis le site du Centre INFFO, <http://www.centre-inffo.fr/Base-textes-conventionnels.html> puis <http://www4.centre-inffo.fr/v2/cpnfp/recherche.phtml> n'utilise pas l'OPAC de LORIS, mais est le fruit d'une application construite par l'École des Mines de Paris, reprenant les données exportées depuis LORIS, et adjointes d'informations complémentaires envoyées par Centre INFFO :

- les accords nationaux interprofessionnels sur la formation professionnelle
- les accords de branche et leurs avenants ainsi que leurs arrêtés d'extension
- une synthèse sur les mécanismes de la négociation collective
- les articles du code du travail auxquels il est fait référence dans les accords
- une présentation et les études réalisées par le CPNFP
- les avis et les recommandations du CPNFP
- le champ, le taux de collecte et les coordonnées des OPCA (Organismes Paritaires Collecteurs Agréés)
- les coordonnées des signataires et des OPCA ainsi qu'un glossaire des sigles utilisés dans les accords

2.3.3 Workflow relatif au traitement documentaire des textes conventionnels



2.4 Structure de la base LORIS

2.4.1 Format documentaire informatique : format UNIMARC

Le format de catalogage utilisé par la bibliothèque de Centre INFFO est le format **UNIMARC** – *UNIversal MARC*. Défini dans le cadre de l'IFLA – *International Federation of Library Associations*, UNIMARC fait partie d'une famille de formats internationaux de référence utilisés dans de très nombreuses bibliothèques à travers le monde : les formats **MARC** – *MAchine Readable Cataloging Record*[58]. Ce sont des formats documentaires informatiques, c'est-à-dire qu'ils listent les champs nécessaires à la description bibliographique en vue d'un traitement informatique. [72]

Structure du format UNIMARC

UNIMARC se présente sous la forme d'une grille de saisie structurée dans laquelle les éléments d'information bibliographiques sont stockés sur deux niveaux :

- dans des zones ou champs de données, identifiées par un libellé et un code numérique, appelé étiquette, sur 3 caractères : 001 à 999.
- dans des sous-zones ou sous-champs : identifiés par un séparateur ($ suivi d'une lettre)

En UNIMARC, l'information bibliographique est découpée en 999 champs regroupés dans 10 blocs logiques : [11]

0XX – numéros d'identification : numéro de la notice, ISBN, ISSN…

1XX – informations codées, regroupant des caractéristiques propres au document : type, date de saisie, date de publication…

2XX – informations descriptives : zones de l'ISBD à l'exception de l'ISBN et des notes

3XX – notes : texte libre destiné à préciser ou compléter l'information figurant dans la description bibliographique

4XX – liens normalisés vers d'autres notices sous forme numérique ou textuelle

5XX – titres associés : titres autres que le titre propre (titre parallèle, etc.)

6XX – analyse de contenu : indexation matière, indices de classification

7XX – responsabilités intellectuelles : auteurs personnes physiques, auteurs personnes morales

8XX – données internationales : source de la donnée

9XX – données locales : données propres à la gestion locale et ne faisant pas l'objet d'échange

[58] Le site de la BNF – Bibliothèque nationale de France présente le format UNIMARC <http://www.bnf.fr/pages/zNavigat/frame/infopro.htm?ancre=normes/no-acuni.htm>. Développé par l'IFLA pour permettre l'échange de l'information bibliographique informatisée et servir d'interface entre les formats MARC nationaux. Il est maintenu au niveau international par le Comité permanent UNIMARC - (Permanent UNIMARC Committee - PUC) de l'IFLA <http://www.ifla.org/en/unimarc>. En France, il est géré par le CFU - Comité français UNIMARC <http://cfu.enssib.fr/>.

UNIMARC, comme les autres formats MARC, respecte les recommandations de la **norme ISO2709** pour la structure des enregistrements, en vue de l'échange de données bibliographiques entre bibliothèques. Toutefois le Centre INFFO n'utilise que peu cette fonction d'export de données bibliographiques selon ISO2709.

Affichage des notices

Le format UNIMARC respecte le format **ISBD** – *International Standard Bibliographic Description*, ensemble de règles normatives également élaborées par l'IFLA. ISBD, lui, n'est pas un format lisible en machine. Il définit les éléments devant figurer dans une notice d'identification et de description bibliographique ; prescrit leur ordre de présentation ; définit une ponctuation pour les délimiter, cette ponctuation étant utilisée comme un codage ; et donne des règles pour leur transcription à partir des sources d'information[59].
UNIMARC est un format de saisie « éclaté » où un champ correspond à une information. A l'affichage, l'utilisation de la norme ISBD permet de présenter de manière lisible toutes les informations nécessaires à l'identification et la description d'un document sous forme de notice, concaténant notamment différentes informations provenant de différents champs UNIMARC, et les ordonnant selon une ponctuation précise.
Ainsi Centre INFFO utilise le format d'affichage ISBD, pour la visualisation de notices dans l'interface de saisie du logiciel LORIS ou sur son catalogue publié sur le Web <http://www.centre-inffo.fr/Base-bibliographique.html>, des références bibliographiques enregistrées sous le SIGB en format UNIMARC.

L'annexe 4 de ce mémoire présente des exemples de notices bibliographiques extraites de la base du Centre INFFO, telles qu'elles apparaissent dans l'interface de saisie LORIS. Elles traduisent un format documentaire adapté, utilisant les champs UNIMARC nécessaires à ses besoins. Elles sont proposées dans les deux formats UNIMARC et ISBD à titre de comparaison.

2.4.2 Structure de la base LORIS

Au Centre INFFO, LORIS est géré par le système ORACLE. Seuls quelques champs sont interrogeables sous ORACLE. Le « champ Blob » de LORIS appelé T_DORIS, présent dans chaque table de la base, comprend en fait l'ensemble des champs UNIMARC contenant les informations relatives aux références bibliographiques, et il n'est pas visible sous ORACLE, parce qu'il est au format propriétaire d'EVER, éditeur de LORIS.

Il y a dans LORIS près de 60 tables ainsi qu'environ 180 listes de contrôle (listes de valeurs contrôlées).

Un répertoire nommé DORIS_UNIV offre une liste de l'ensemble des **tables LORIS** : LORIS T.xxxx (plus les index X.... etc.). Ce répertoire n'est accessible que par un administrateur dans System C : → Program Files → LORIS 3.5.3.

[59] Voir la documentation relative à l'ISBD - Description bibliographique internationale normalisée, sur le site de la BNF – Bibliothèque nationale de France <http://www.bnf.fr/pages/zNavigat/frame/infopro.htm?ancre=normes/no-isbd.htm>.

Une liste nommée LORIS UNIMARC.CTRL offre une liste de l'ensemble des **champs UNIMARC**. Cette liste des champs est accessible dans le logiciel :
par le menu « accès multicritères » → « Table Paramètres » → Identifiant (i) = « LORIS UNIMARC.CTRL .

Chaque champ UNIMARC est lié à des tables et listes, permettant théoriquement de retrouver l'ensemble des liens entre les données.

Sous logiciel tableur : l'on pourrait dresser un tableau des champs UNIMARC. Chaque champ faisant l'objet d'une ligne avec au regard de chacun, indiqué dans plusieurs colonnes : si le champ concerné est présent dans une table, s'il est contrôlé par une liste, si un contenu est renseigné dans la base, etc.

Et à partir de cet inventaire général des tables et de leurs liens :
- générer un graphe des tables et de leurs liens.
- présenter une forme d'abstraction de manière plus globale des données gérées dans LORIS.

Il apparaît que le travail d'investigation approfondi sur la structure de la base LORIS nécessiterait un grand nombre de jours pour couvrir les 900 champs UNIMARC, ce qui pourrait représenter environ 1500 champs et sous-champs.

2.4.3 Liste des tables principales dans LORIS

Le tableau ci-dessous ne liste que les **tables principales** contenant les différents types de données gérées sous LORIS : **objets documentaires** décrits et **autorités** rattachées.

Nom de la table	Nombre de notices (au 16/11/2009)	Commentaire
ABONNEMENT	727	Tous abonnements confondus, actifs ou anciens.
AUTEUR	16 345	
CANDIDAT DESCRIPTEUR	194	
CANDIDAT THEME *	8	
CATALOGUE	36 128	
CODE NAF *	776	Dont environ 450 CCN - Conventions Collectives Nationales indexées.
COLLECTION	1 160	
COLLECTIVITE	2 349	
CONVENTION COLLECTIVE *	372	
EDITEUR	2 109	
ENTREPRISE	849	Liste de noms d'entreprises destinée à l'indexation de textes décrivant des pratiques de formation.
ETAT DE LA COLLECTION	216	Abonnements, noms des périodiques.
MINISTERE *	118	

OPCA *	101	
SIGNATAIRE *	1 642	
THEME *	100	
THESAURUS	2273	Descripteurs du « Thésaurus de la formation ».

* Tables utilisées pour les textes conventionnels (accords interprofessionnels et accords de branche).

2.4.4 Le Thésaurus de la formation

Le **Thésaurus de la formation**, dans sa dernière édition 2006, est utilisé pour indexer l'ensemble des ressources décrites dans la base LORIS. Cet outil documentaire, langage structuré, a été réalisé en collaboration par le Centre INFFO, l'AFPA – Association pour la Formation Professionnelle des Adultes ainsi que les CARIF - Centres d'Animation et de Ressources de l'Information sur la Formation <http://www.centre-inffo.fr/Thesaurus-de-la-formation-2006.html>.

Sa structure se compose de :

- une liste alphabétique générale
- 28 listes thématiques, adjointes chacune d'un schéma fléché,
- 12 listes additionnelles

Près de 2300 descripteurs et 2000 non-descripteurs sont répertoriés, adjoints de 128 définitions et 104 notes d'application. Trois types de relations sémantiques sont distingués : relation d'équivalence, relation hiérarchique, et relation d'association.

L'outil informatique ayant servi à produire le **Thésaurus de formation** lors de ses deux dernières éditions, en 2000 puis en 2006, est basé sur celui créé en 1995 pour produire le **Thésaurus Formacode**. Il s'agit d'une application maison développée sous le SGBD Pick (en Basic) accessible via le logiciel d'émulation Wintegrate. L'application Pick pour la gestion du **Formacode** a été abandonnée en 2008, un nouvel outil ayant été développé en interne utilisant PHP et MySQL (pour produire la dernière édition de ce Thésaurus Formacode 2008). En revanche, même si elle est figée, l'application Pick de gestion du **Thésaurus de la formation 2006** demeure.

Le **Thésaurus de la formation** est actuellement implanté sous LORIS dans le format UNIMARC.
Il n'a pas fait l'objet d'une mise à jour depuis 2006, mais dans le format UNIMARC de LORIS, un champ « Candidat descripteur » permet de prendre en compte les nouveaux concepts apparus depuis 2006, afin d'indexer les ressources documentaires.
LORIS est équipé d'un module de gestion de thésaurus, mais il n'a pas été utilisé pour produire l'édition papier de 2006.

3 Base de données sur les Produits pédagogiques multimédias

3.1 Présentation de la Base Produits pédagogiques multimédias – Constats

Nom de la base et/ou Type d'objet	Type de base	Service gestionnaire	Utilisation des données Publication sur le Web
Produits pédagogiques multimédias pour la formation des adultes	Application développée par eMotive	Documentation Utilisateur : 1 documentaliste	• **Base des Produits pédagogiques multimédias** pour la formation des adultes <http://www.centre-inffo.fr/La-base-de-donnees-de-produits.html> encore hébergée sur le site Web de l'ancien organisme ALGORA <http://ressourcesv2.e-motive.com/frontblocks/produits/>. Fiches descriptives de produits ou ressources Web, et pour certaines d'entre elles, accompagnées de fiche(s) d'évaluation. • Une **sélection thématique des nouveautés recensées** est également proposée environ tous les deux mois : <http://www.centre-inffo.fr/Selection-thematique-des,2683.html>. • Des **dossiers documentaires** au format PDF valorisent les produits recensés par thèmes : **Etat des lieux des ressources pédagogiques** <http://www.centre-inffo.fr/Etat-des-lieux-des-ressources.html> ; **Fichiers doc express** <http://www.centre-inffo.fr/Fichiers-express.html>. Les références de la BDD Produits pédagogiques multimédias sont également consultables via : • La **BCRP – Base coopérative d'information sur les ressources pédagogiques**, qui constitue un métamoteur de recherche <http://www.centre-inffo.fr/BCRP-Base-Cooperative-d.html> sur quatre bases de données réalisées par différentes institutions (partenariat Centre INFFO – Agrosup Dijon/EDUTER, les CNR – Centres Nationaux de Ressources de l'Éducation Nationale, et le CARIF de la région Lorraine INFFOLOR). • Et, pour certaines données, consultables sur le **site educauto.org** du Centre national de ressources pour la formation automobile (partenariat entre l'Education nationale et l'ANFA – Association nationale pour la formation automobile), qui propose un formulaire de recherche <http://www.educauto.org/OutilsMultimedia/multimed.php>. Sont notamment proposées des fiches descriptives des outils multimédias spécifiques à la formation professionnelle automobile.

3.1.1 Présentation sur le site Web du Centre INFFO

Le Centre INFFO met à la disposition des acteurs de la formation, une base de données de produits pédagogiques multimédias <http://www.centre-inffo.fr/La-base-de-donnees-de-produits.html>. Cette base dont la création remonte à 1994, était auparavant produite et gérée par l'association ALGORA, dont une partie des activités d'information sur la FOAD – Formation ouverte et à distance a été transférée au Centre INFFO en janvier 2007 au moment de sa dissolution.

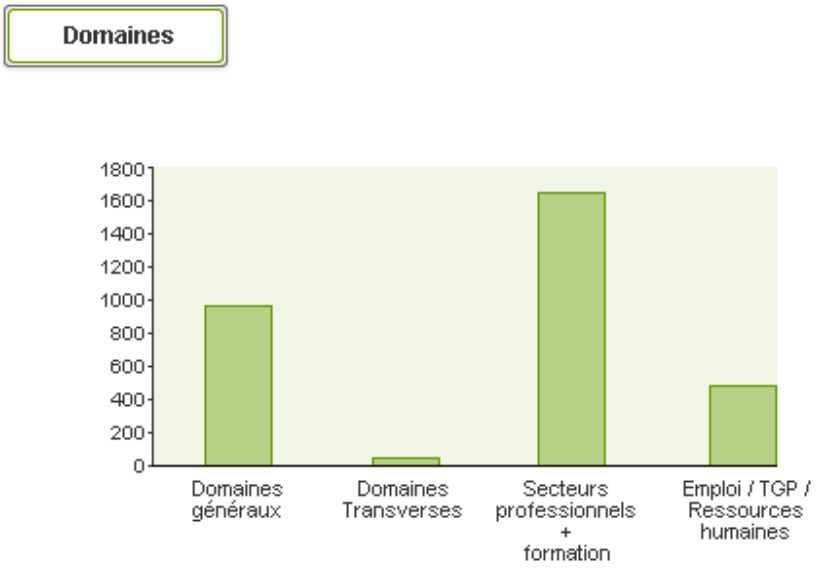
Cette base se compose de CD-Rom, DVD-Rom, DVD-Vidéo, VHS. Elle référence également des sites proposant des contenus de formation en ligne. Elle cible un public adulte uniquement. Elle vise à apporter à tout utilisateur potentiel, impliqué dans la Formation Ouverte et à Distance, des informations sur les produits pédagogiques, actuellement présents sur le marché (pour les produits hors ligne) et pouvant être utilisés dans des dispositifs de formation.

Environ 2600 références sont présentées : 30 % des produits couvrent les domaines généraux et 70 % les domaines professionnels ou ayant trait à l'emploi. Ce qui fait la spécificité de cette base est son caractère généraliste – la majorité des domaines ou secteurs d'activité est représentée – et la possibilité, pour les utilisateurs, de consulter directement des évaluations de produits en ligne. 230 évaluations sont répertoriées concernant 114 produits. Un même produit peut faire l'objet de plusieurs évaluations, apportant ainsi une approche différenciée en fonction des situations d'apprentissage et du public impliqué.[60]

3.1.2 Domaines

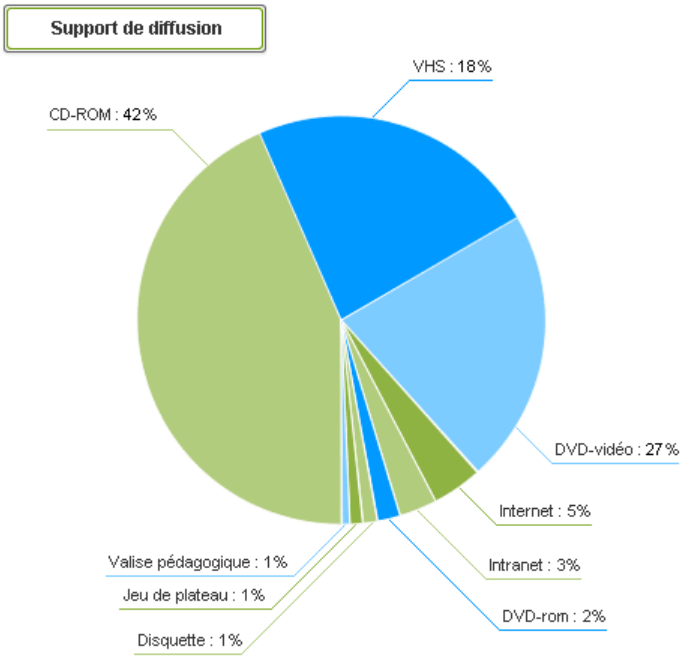
Les différents domaines ou secteurs d'activité représentés se répartissent en :

- **Domaines généraux : 994 produits (31 %)**
  Français, mathématiques, langues et FLE - Français Langue Etrangère, Illettrisme, sciences humaines et sociales, sciences de la vie, sciences de la terre, physique, chimie, dictionnaires, encyclopédies
- **Secteurs professionnels + Formation : 1634 produits (51 %)**
  Secteurs professionnels : automobile, agriculture, BTP – Bâtiment-Travaux Publics, mécanique, management, comptabilité, entreprise, vente, bureautique, etc.
  Formation : outils du formateur, apprentissage, alternance, évaluation des connaissances, formation de formateurs, etc.
- **Emploi / Développement personnel / Ressources humaines : 500 produits (16 %)**
  Emploi : informations sur les métiers, politiques de l'emploi
  Développement personnel : recherche emploi TRE – Techniques de Recherche d'Emploi, orientation professionnelle, insertion, etc.
  Ressources humaines : recrutement, entretien d'évaluation, etc.
- **Remise à niveau / apprentissage de base (Domaines transverses) : 49 produits (2 %)**



3.1.3 Supports de diffusion

Support de diffusion	Nombre de produits	Pourcentage
CD-Rom	1302	42 %
DVD-vidéo	859	27 %
VHS	571	18 %
Internet	162	5 %
Intranet	85	3 %
DVD-Rom	67	2 %
Jeux de plateau	33	1 %
Disquette	31	1 %
Valise pédagogique	16	1 %



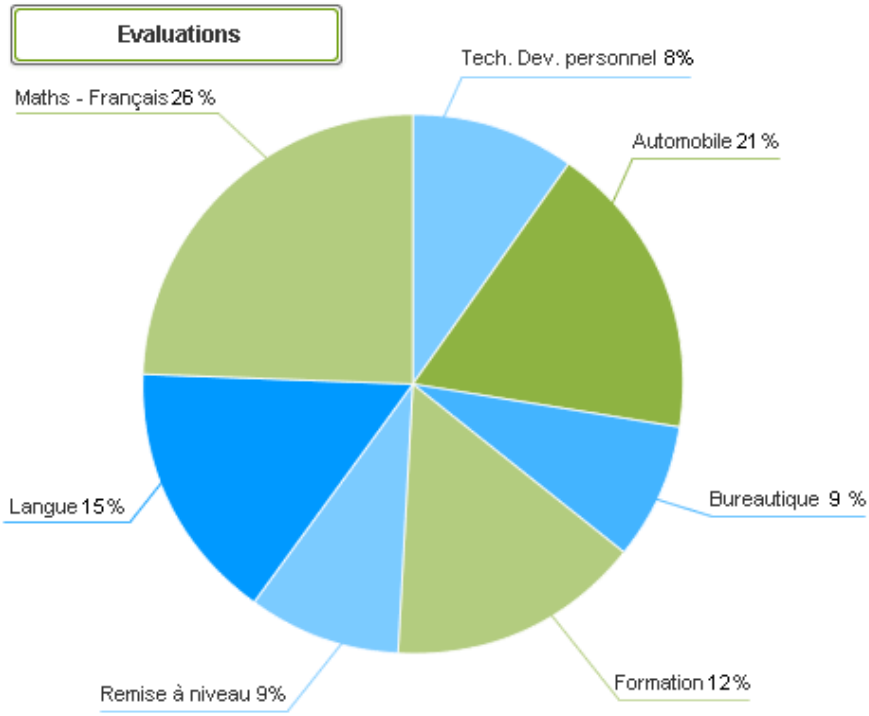
3.1.4 Evaluations

241 évaluations – 115 produits évalués

La base de données de produits pédagogiques multimédias comptabilise actuellement 231 évaluations concernant 115 produits. Un même produit peut faire l'objet de plusieurs évaluations, apportant ainsi une approche différenciée en fonction des situations d'apprentissage et du public impliqué. Les produits le plus souvent évalués concernent le

[60] Données extraites de la page <http://www.centre-inffo.fr/La-base-de-donnees-de-produits.html> au 30/06/2009.

français, les mathématiques, la formation, le secteur automobile et les langues. Ces évaluations sont, actuellement, essentiellement le fait de formateurs émanant soit du réseau APP - Ateliers de Pédagogie Personnalisée, soit du réseau ANFA - Association Nationale pour la Formation Automobile.



3.1.5 Consultations

Un système de statistiques permet de recueillir des informations précises sur le type de produit consulté ou recherché et d'orienter ainsi le référencement en fonction des besoins et des demandes des publics utilisateurs.

Depuis 2007, on remarque une consultation significative des produits ayant trait :
- Pour les domaines généraux, à :

- l'apprentissage de base, et du français en particulier
- les outils de remédiation

- Pour les domaines d'ordre professionnel :

- la prévention des risques professionnels
- les secteurs d'activité relatifs à la mécanique, au BTP, aux métiers du service et à la santé.
- les produits d'information sur les métiers

Top	Statistiques des produits les plus consultés – année 2008	Total
01	Schémaplic Zi – Edit. : Groupe Fitec	286
02	CCDMD – Amélioration du français (site)	157
03	Nanosciences Nanotechnologies – Edit. : CNRS Images	156
04	Maintenir l'agriculture de montagne : l'exemple du Val d'Abondance – Edit. : Educagri éditions	153
05	Concordia, sciences australes – Edit. : CNRS Images	143
06	A la découverte de Jean de La Fontaine (site)	142
07	Accueil sécurité : Carrière – Edit. : 2J Media/Process Image	141
08	L'air et autres gaz – Edit. : Pierron Education	139
09	Israël – Palestine, l'emprise des images – Edit. : INA/Editions Chrysis	134
10	Le récif corallien – Edit. : IRD Audiovisuel /CDP de Nouvelle-Calédonie	132
11	Donner à chacun sa chance – Edit. : Palmis Production	126
12	Les Hôpitaux d'excellence : Choix pour naître – Edit. : Les écrans du social	115
13	Les fondamentaux en comptabilité 2 – Edit. : Aunege/Canal U (site)	114
14	Objectif Chine. Entre fièvres et vaccins. Le cas Mérial – Edit. : R&O Multimedia	112
15	Scènes d'un chantier ordinaire – Edit. : Inrs	109
16	Spectrum – Les messages de la lumière – Edit. : Pierron Education	108
17	Buffon, le penseur de la nature – Edit. : CNDP	108
18	Chercher pour comprendre... un entretien avec Emile Meurice – Edit. : CNASM	108
19	Jeu, set et maths – Edit. : CRDP du Centre Académie d'Orléans-Tours	104
20	Boulanger : une formation, un métier – Edit. : CRDP de Versailles	102

NB. Pour une bonne interprétation des données relatives aux statistiques de consultation, il est nécessaire de tenir compte de la date d'édition du produit et de sa date de référencement dans la base de données.

3.1.6 Constats et réflexions

La BDD Produits pédagogiques multimédias est maintenue et hébergée par l'entreprise e-Motive <http://www.e-motive.com/>. La connaissance fine de sa structure nécessiterait donc de passer par cet organisme puisqu'il est en charge de son administration.

Les 2500 références comprennent des ressources disponibles en ligne – *on line* – ainsi que des ressources *off line*, par exemple des éditions vidéos. Les objets documentaires décrits ne sont pas en tant que tels acquis, et ne correspondent donc pas à un fonds documentaire disponible à Centre INFFO.

Cette base de références nécessite une veille continue ainsi qu'une mise à jour très régulière car les produits pédagogiques bougent beaucoup. Chaque fiche produit doit être révisée au minimum une fois par an, mais cela peut

aller jusqu'à trois fois selon les produits. Il est en effet important que le produit recensé dans la base demeure accessible et que l'utilisateur connaisse son prix.

Comme développement pouvant être envisagé :
de la même manière que ce qui est fait pour les Produits intéressant l'ANFA – Association Nationale pour la Formation Automobile sur son site <http://www.educauto.org/OutilsMultimedia/multimed.php>, valoriser les Produits d'autres domaines en proposant une démarche similaire à des organismes sectoriels ou des branches professionnelles, par exemple. Le principe : les notices Produits concernées font l'objet d'un « tag » au moment de la saisie, permettant d'être recherchées via un formulaire directement depuis le site Web de l'organisme partenaire.

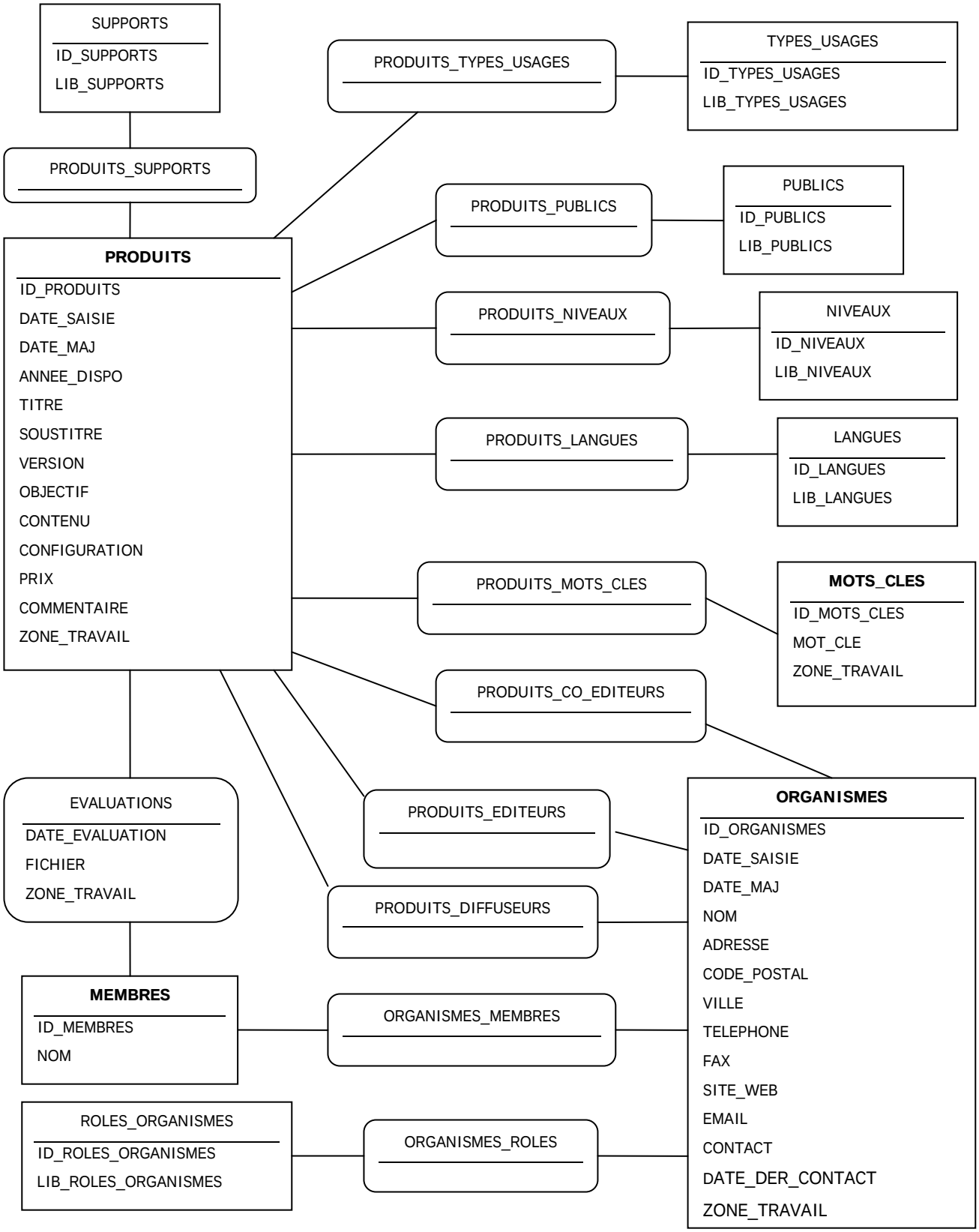
3.2 Structure de la base Produits pédagogiques multimédias : tables, modèle

Tables principales

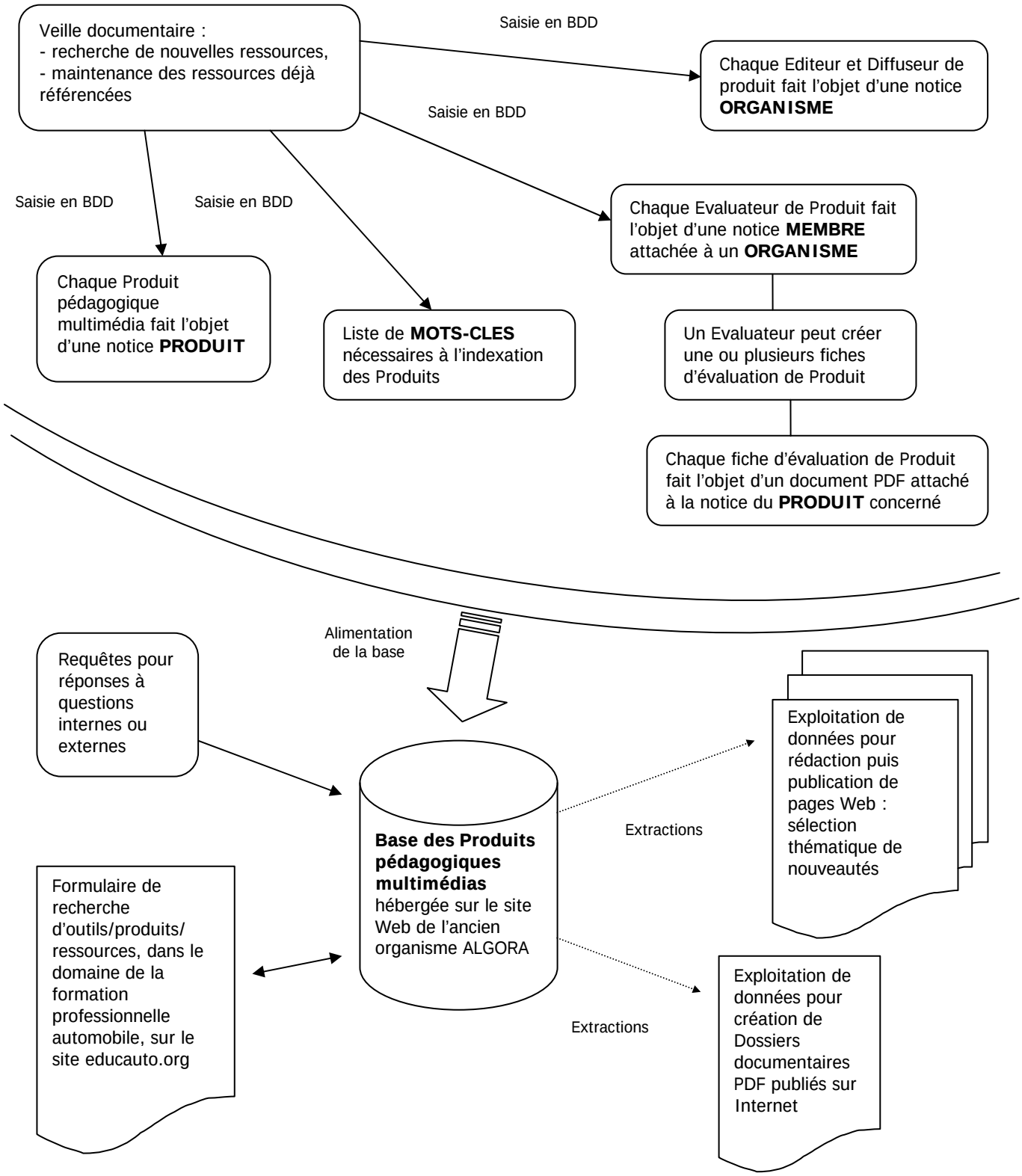
Noms des Tables	***Remarques***
PRODUITS	- Si un PRODUIT fait l'objet d'une évaluation, cette dernière consiste en un document rempli par un évaluateur, attaché en format PDF à la notice descriptive du PRODUIT concerné.
ORGANISMES	Comprend les notices descriptives d'éditeurs et de distributeurs.
MEMBRES	Table utilisée pour gérer des fiches « évaluateur » liées à des ORGANISMES
MOTS_CLES	Liste de termes appelés « mots-clés » proposant jusqu'à trois niveaux de hiérarchisation. Cette liste est ouverte en fonction des besoins d'indexation des produits et ressources Web recensés.

Modèle des données simplifié

Le schéma qui suit, a été réalisé à partir des données recueillies par entretien auprès de la personne gestionnaire de la base, ainsi que des cahiers des charges produits au moment de la création de la base en 2003. Il mentionne les tables et leurs champs.


SUPPORTS
ID_SUPPORTS
LIB_SUPPORTS
PRODUITS_TYPES_USAGES
TYPES_USAGES
ID_TYPES_USAGES
LIB_TYPES_USAGES
PRODUITS_SUPPORTS
PRODUITS_PUBLICS
PUBLICS
ID_PUBLICS
LIB_PUBLICS
PRODUITS
ID_PRODUITS
DATE_SAISIE
DATE_MAJ
ANNEE_DISPO
TITRE
SOUSTITRE
VERSION
OBJECTIF
CONTENU
CONFIGURATION
PRIX
COMMENTAIRE
ZONE_TRAVAIL
PRODUITS_NIVEAUX
NIVEAUX
ID_NIVEAUX
LIB_NIVEAUX
PRODUITS_LANGUES
LANGUES
ID_LANGUES
LIB_LANGUES
PRODUITS_MOTS_CLES
MOTS_CLES
ID_MOTS_CLES
MOT_CLE
ZONE_TRAVAIL
PRODUITS_CO_EDITEURS
EVALUATIONS
DATE_EVALUATION
FICHIER
ZONE_TRAVAIL
PRODUITS_EDITEURS
ORGANISMES
ID_ORGANISMES
DATE_SAISIE
DATE_MAJ
NOM
ADRESSE
CODE_POSTAL
VILLE
TELEPHONE
FAX
SITE_WEB
EMAIL
CONTACT
DATE_DER_CONTACT
ZONE_TRAVAIL
PRODUITS_DIFFUSEURS
MEMBRES
ID_MEMBRES
NOM
ORGANISMES_MEMBRES
ROLES_ORGANISMES
ID_ROLES_ORGANISMES
LIB_ROLES_ORGANISMES
ORGANISMES_ROLES

3.3 Workflow du traitement documentaire des produits pédagogiques multimédias



→ Une seule documentaliste alimente la BDD et en exploite les données.

NB : En cas de besoin de supprimer une fiche PRODUIT, nécessité préalable de supprimer le(s) document(s) PDF correspondant à la (aux) fiche(s) d'évaluation du Produit concerné, si elle(s) existe(nt).

4 Application NOEL – Nouvelle Offre en Ligne

4.1 Présentation de NOEL application de gestion de l'offre de formation

Nom de la base Type d'objet	Type de base	Service gestionnaire	Utilisation des données Publication sur le Web
Application NOEL – Nouvelle Offre en Ligne	Base MySQL	Documentation Utilisateurs : 2 documentalistes	• **Organismes de formation** proposant des formations dans le domaine des métiers de la formation et de l'insertion professionnelle <http://www.centre-inffo.fr/spip.php?page=organismes-de-formation>. • **Titres et Diplômes** dans le champ des métiers de la formation des adultes <http://www.centre-inffo.fr/spip.php?page=stages-metiers-formation>. • **Stages agréés par l'Etat** au titre de la rémunération au plan national <http://www.centre-inffo.fr/spip.php?page=stages-agrees>.

NOEL – Nouvelle Offre en Ligne constitue une application maison développée en PHP, remplaçant une ancienne application DORIS utilisée jusqu'en 2008.

Depuis la décentralisation des compétences en matière de formation professionnelle, les Régions ont la mission d'informer sur l'offre de formation. Cette mission est confiée aux CARIF – Centres d'Animation et de Ressources de l'Information sur la Formation, cofinancés par l'Etat et les Régions.

Au Centre INFFO, le périmètre d'investigation sur l'offre de formation en France – organismes et stages (actions de formation continue) – s'est donc réduit ces dix dernières années. Les données collectées ne concernent désormais que les métiers de la formation et de l'insertion professionnelle, les « stages agréés » ainsi que certains secteurs professionnels pour le besoin d'enquêtes ponctuelles effectuées sur commande.

L'**application NOEL** est en fait surtout composée de deux « fichiers »*****, qui seuls intéressent directement la production :

- le fichier **ORGANISMES** : contient des fiches descriptives d'**Organismes de formation**. Tous les enregistrements n'ont plus une utilité directe, mais sont encore susceptibles d'en avoir une. Seules sont mises à jour les fiches descriptives d'organismes (que l'on peut estimer aux alentours de 1000) qui proposent des « Stages agréés » ou des formations menant aux métiers de la formation et de l'insertion.
- le fichier **STAGES** : comprenant les fiches descriptives de deux types de **Stages de formation**. Les stages agréés par l'Etat au titre de la rémunération au plan national ; les formations qui préparent aux multiples métiers de la formation des adultes et de l'insertion professionnelle (de courte ou longue durée menant à des certificats ou diplômes).

*****NB : Un fichier TDH « Titres et diplômes homologués » perdure, que l'on peut qualifier de « mort », car l'ensemble des données ont été transférées à la Commission nationale de la certification professionnelle (CNCP) lors de la création du Répertoire national des certifications professionnelles (RNCP) en 2003-2004. Le fichier TDH est conservé car il est encore relié à des organismes du fichier ORGANISMES.

4.2 Structure des fichiers STAGES et ORGANISMES, constitutifs de NOEL

Le tableau ci-dessous propose un portrait de la structure des fichiers STAGES et ORGANISMES, listant les tables principales et leurs relations. Les nombreux champs disponibles dans chacune des tables, ne sont pas rappelés, mais sont en conformité avec le **standard LHÉO – Langage Harmonisé d'Échange d'informations sur l'Offre de formation**[61] <http://lheo.org/> [40]. Ces éléments sont extraits de la base MySQL au moyen des outils d'administration mis en place sur le site Intranet par le développeur Web ayant écrit l'application.

Liste des tables au 15/10/2009

Noms des tables	Commentaires	Nb.	A pour références	Est référencée par
off_decideurs	Décideur	135	off_familles_decideurs (RESTRICT)	**off_stage_decideur_matches**
off_departements	Département	109	off_regroupements_geographiques (RESTRICT)	**off_departements_limitrophes_matches**
			off_regions (RESTRICT)	off_departements_limitrophes_matches
			off_statuts_departement (RESTRICT)	**off_organismes**
				off_stages
off_departements_limitrophes_matches	Correspondance département / département limitrophe	481	off_departements (RESTRICT)	
			off_departements (CASCADE)	
off_exige_souhaite	Exigence	2		**off_stages**
off_familles_decideurs	Famille de décideurs	5		**off_decideurs**
off_niveaux	Niveau *(d'admission ou de sortie)*	9		**off_stages**
off_normes_labels	Norme ou label	99	off_organismes_certificateurs (RESTRICT)	**off_organisme_norme_label_matches**
			off_types_norme (RESTRICT)	
off_organisme_diplome_matches	Correspondance organisme / formacode	153	fcd_formacode (RESTRICT)	
			off_organismes (CASCADE)	
off_organisme_domaine_matches	Correspondance organisme / formacode Permet le filtre sur le domaine « professionnalisation acteurs de la formation	3622	fcd_formacode (RESTRICT)	
			off_organismes (CASCADE)	
off_organisme_moyen_pedagogique_matches	Correspondance organisme / formacode	51	fcd_formacode (RESTRICT)	
			off_organismes (CASCADE)	
off_organisme_norme_label_matches	Correspondance organisme / norme	1	off_normes_labels (RESTRICT)	
			off_organismes (CASCADE)	
off_organisme_public_matches	Correspondance organisme / formacode	132	fcd_formacode (RESTRICT)	
			off_organismes (CASCADE)	
off_organisme_ref_ci_matches	Correspondance organisme / référence Centre INFFO	4592	off_organismes (CASCADE)	
			off_references_ci_organismes (RESTRICT)	
off_organismes	Organisme (formateur ou gestionnaire)	2891	off_departements (RESTRICT)	*dir_structures*

[61] Le Centre INFFO fait partie du Comité de Pilotage de ce projet : cf. chapitre 1.2.3.1 de la Quatrième partie de ce mémoire.

Noms des tables	Commentaires	Nb.	A pour références	Est référencée par
			off_statuts (RESTRICT)	off_organisme_diplome_matches
			dir_structures (SET NULL)	off_organisme_domaine_matches
				off_organisme_moyen_pedagogique_matches
				off_organisme_norme_label_matches
				off_organisme_public_matches
				off_organisme_ref_ci_matches
				off_stage_organisme_formateur_matches
				off_stage_organisme_gestionnaire_matches
off_organismes_certificateurs	Organisme certificateur	18		**off_normes_labels**
off_references_ci_organismes	Référence Centre INFFO (organismes) *Liste de références servant à l'affichage Web. Permet la synchronisation avec la base sur serveur externe.*	15		**off_organisme_ref_ci_matches**
off_references_ci_stages	Référence Centre INFFO (stages) *Liste de références servant à l'affichage Web. Permet la synchronisation avec la base sur serveur externe.*	10		**off_stage_ref_ci_matches**
off_regions	Région	33		**off_departements**
off_regroupements_geographiques	Regroupement géographique	8		**off_departements**
off_sessions	Session	79	off_stages (CASCADE)	
off_stage_decideur_matches	Correspondance stage / décideur *Utilisé pour les stages agréés.*	1345	off_decideurs (RESTRICT) off_stages (CASCADE)	
off_stage_diplome_matches	Diplôme *(décerné ou requis)* Correspondance stage / formacode	992	fcd_formacode (RESTRICT) off_stages (CASCADE)	
off_stage_domaine_matches	Correspondance stage / formacode	2801	fcd_formacode (RESTRICT) off_stages (CASCADE)	
off_stage_modalite_pedagogique_matches	Correspondance stage / formacode	1	fcd_formacode (RESTRICT) off_stages (CASCADE)	
off_stage_organisme_formateur_matches	Correspondance stage / organisme	2109	off_organismes (RESTRICT) off_stages (CASCADE)	
off_stage_organisme_gestionnaire_matches	Correspondance stage / organisme	935	off_organismes (RESTRICT) off_stages (CASCADE)	
off_stage_public_matches	Correspondance stage / formacode	1659	fcd_formacode (RESTRICT) off_stages (CASCADE)	
off_stage_ref_ci_matches	Correspondance stage / référence Centre INFFO	1738	off_references_ci_stages (RESTRICT) off_stages (CASCADE)	
off_stage_rythme_matches	Correspondance stage / formacode	1126	fcd_formacode (RESTRICT) off_stages (CASCADE)	
off_stage_type_sanction_matches	Correspondance stage / formacode	6	fcd_formacode (RESTRICT) off_stages (CASCADE)	
off_stages	Stage	2390	off_departements (RESTRICT)	off_sessions
			off_niveaux (RESTRICT)	off_stage_decideur_matches
			off_exige_souhaite (RESTRICT)	off_stage_diplome_matches
			fcd_formacode (RESTRICT)	off_stage_domaine_matches
			off_unites_duree (RESTRICT)	off_stage_modalite_pedagogique_matches

Noms des tables	Commentaires	Nb.	A pour références	Est référencée par
				off_stage_organisme_formateur_matches
				off_stage_organisme_gestionnaire_matches
				off_stage_public_matches
				off_stage_ref_ci_matches
				off_stage_rythme_matches
				off_stage_type_sanction_matches
off_statuts	Statut	63		**off_organismes**
off_statuts_departement	Statut de département	5		**off_departements**
off_types_norme	Type de norme	6		**off_normes_labels**
off_unites_duree	Unité de durée	7		**off_stages**

Le sous-chapitre qui suit propose une modélisation des données d'après la structure extraite de la base MySQL.

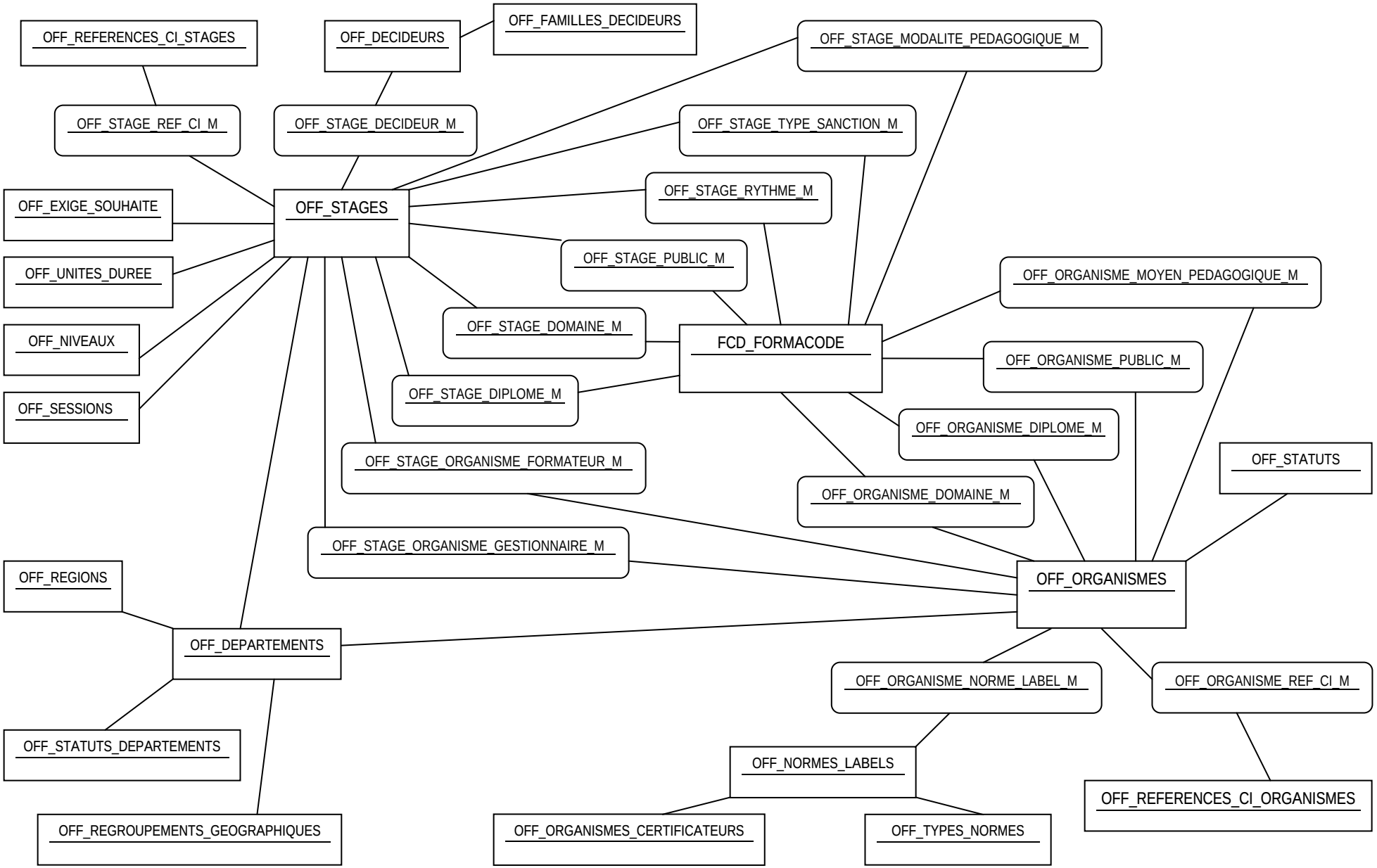
Il apparaît que les tables OFF_STAGES et OFF_ORGANISMES sont reliées à la table FCD_FORMACODE.

« Formacode » est le nom du « Thésaurus de l'offre de formation », coproduit et publié par le Centre INFFO, qui fait également l'objet d'une application MySQL décrite au chapitre 5.

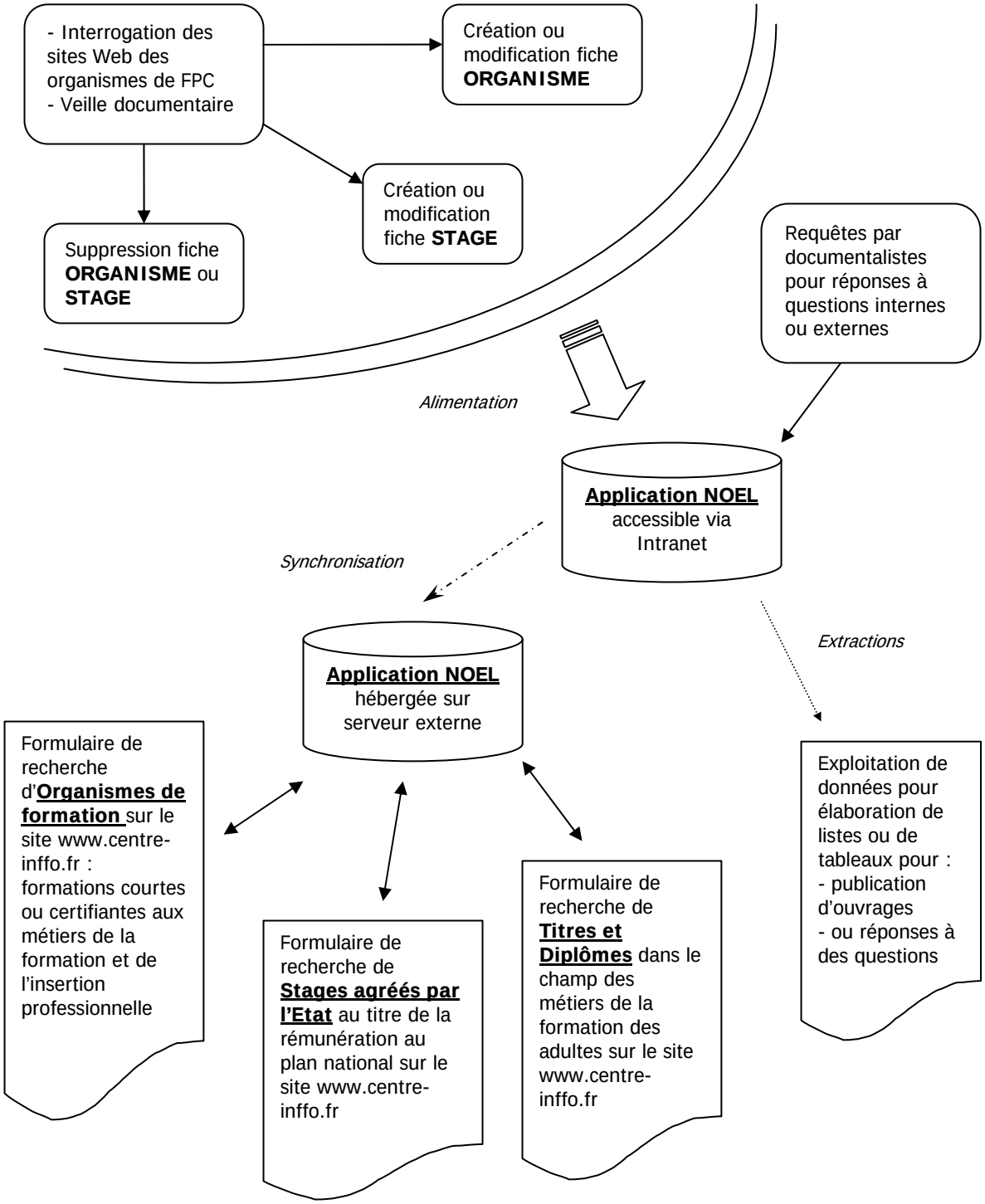
Pour plus de lisibilité : - les noms des tables sont reproduits en majuscules

- les noms des tables de jonction ne se terminent pas par « matches » mais par M.

4.3 Modèle des données de NOEL



4.4 Workflow du traitement de l'offre de formation avec NOEL



5 Application de gestion du Thésaurus de l'offre de formation : Formacode

5.1 Présentation du Thésaurus Formacode – Constats

Nom de la base Type d'objet	Type de base	Service gestionnaire	Utilisation des données Publication sur le Web
Gestion du **thésaurus Formacode de l'offre de formation**	base MySQL	Documentation Utilisateurs : 3 documentalistes	• Publication en version papier du **Formacode® 9.0** (édition 2008) ; • Publication en version informatique **Formacodex V9.0** (édition 2008) • **Base Formacode** consultable en ligne <http://www.centre-inffo.fr/Le-Formacode-R-thesaurus-de-l.html>.

Le **Formacode** constitue un vocabulaire de référence pour toute personne travaillant sur l'offre de formation. Intitulé « Thésaurus des domaines de formation » dans ses éditions antérieures, le **Formacode 9ème édition**, publiée en 2008, constitue maintenant le **« Thésaurus de l'offre de formation ».**

Conçu pour caractériser et indexer finement l'offre de formation, les métiers et les certifications, cet outil documentaire propose des termes normalisés, classés et codifiés. Les termes répertoriés relatifs à des domaines de formation sont mis en correspondance avec les entrées de deux autres nomenclatures : le **ROME – Répertoire Opérationnel des Métiers et des Emplois** publié par l'ANPE – Agence Nationale pour l'Emploi (devenu Pôle Emploi en décembre 2008), et la **NSF – Nomenclature des Spécialités de Formation**, créée par le CNIS – Conseil National de l'Information Statistique. Le thésaurus Formacode permet ainsi l'articulation entre une **offre de formation**, un **métier** et un type de **certification**.

Deux produits sont distingués pour la diffusion :

- un ouvrage en version « papier » Formacode® 9.0, en deux tomes ;
- le module d'intégration informatique Formacodex V9.0 pour intégrer le thésaurus Formacode dans un logiciel de gestion documentaire. Il facilite la gestion de bases de données sur l'offre de formation et la mise à disposition de ces données sur les sites Internet.

Le Formacode est utilisé par le Centre INFFO pour indexer des fiches descriptives d'organismes ou de stages de formation professionnelle. Mais il peut être également utilisé par les entreprises, organismes de formation, gestionnaires, financeurs, informateurs sur la formation professionnelle, conseillers d'orientation, pour de multiples usages :

- identifier, classer et exploiter l'offre de formation et les certifications ;
- explorer les bases de données sur l'offre de formation accessibles sur Internet ;
- mesurer par domaine l'investissement formation de l'entreprise ;
- gérer les dossiers des stagiaires et des parcours de formation des salariés ;
- faire le lien entre la formation et les métiers.

Composition du thésaurus : 14 domaines, 64 champs sémantiques et 3065 descripteurs, organisés et présentés sous la forme de :

- Liste de champs sémantiques
- Listes hiérarchiques et alphabétiques des descripteurs
- Listes annexes :

  - certifications/diplômes/titres ;

  - pays ; publics ;

  - mots-outils : plusieurs rôles sont distingués : aires d'influence ; certifications qualité, chartes, labels ; financeurs ; fonctions de l'organisme ; modalités d'accueil ; modalités d'admission ; modalités pédagogiques ; moyens pédagogiques ; objectif général de formation ; rythmes ; types de validation ;
- Index permuté ;
- Correspondance ROME-Formacode.

La précision de Formacode lui vaut d'être la **nomenclature de référence** dans les documents officiels traitant de l'offre de formation et d'alimenter les outils et systèmes d'information développés par le Service Public de l'Emploi et de la Formation :

- le **référentiel LHÉO** - Langage Harmonisé d'Échange d'informations sur l'Offre de formation[62] <http://lheo.org/> [40],
- les bases de données des **CARIF** – Centres d'Animation et de Ressources de l'Information sur la Formation <http://www.intercariforef.org/formations/reseau-carif.html> sur l'offre de formation initiale et continue,
- le système d'information **OFAA - Offre de Formation ANPE-Assédic**[63],
- les outils des **Missions locales**,
- le **portail national Orientation-Formation** <http://www.orientation-formation.fr/>.

---

[62] Cf. chapitre 1.2.3.1 de la Quatrième partie de ce mémoire

[63] L'ANPE – Agence Nationale pour l'Emploi et le réseau des Assédic ont fusionné en décembre 2008 pour donner Pôle emploi <http://www.pole-emploi.org/>.

5.2 Structure du fichier FORMACODE

Le tableau ci-dessous propose un portrait de la structure du fichier FORMACODE, listant les tables principales, les relations qu'elles ont entre elles, ainsi que les relations établies avec des tables des fichiers STAGES et ORGANISMES décrites dans le précédent chapitre 5.

Les nombreux champs disponibles dans chacune des tables du fichier FORMACODE ne sont pas rappelés. Ces éléments sont extraits de la base MySQL au moyen des outils d'administration mis en place sur le site Intranet par le développeur Web ayant écrit l'application.

Liste des tables au 15/10/2009

Nom	Commentaire	Nb.	A pour références	Est référencée par
fcd_associated_term_matches	Correspondances de termes associés (TA)	8831	fcd_formacode (CASCADE)	
fcd_certifier_matches	Correspondance formacode / certificateur	150	fcd_certifiers (RESTRICT) fcd_formacode (CASCADE)	
fcd_certifiers	Certificateur	9		**fcd_certifier_matches**
fcd_deleted_formacodes	Corbeille des formacodes	0	fcd_editions (CASCADE)	
fcd_diploma_level_matches	Correspondance entre les diplômes (formacodes particuliers)	556	fcd_diploma_levels (CASCADE) fcd_formacode (CASCADE)	
fcd_diploma_levels	Niveaux de diplôme	7		fcd_diploma_level_matches
fcd_editions	Édition du Formacode	5		fcd_deleted_formacodes fcd_exports fcd_formacode fcd_permuted_index
fcd_empty_words	Mots vides	20		
fcd_export_options	Option d'export	4		**fcd_exports**
fcd_export_order_columns	Colonne ordonnable pour l'export	3		**fcd_exports**
fcd_exports	Export de formacodes	0	fcd_editions (CASCADE) fcd_export_options (RESTRICT) fcd_export_order_columns (RESTRICT)	
fcd_formacode	Formacode (toute édition confondue)	14974	fcd_editions (CASCADE)	fcd_associated_term_matches fcd_certifier_matches fcd_diploma_level_matches fcd_meta_domain_matches *fcd_modifications* fcd_nsf_matches fcd_onisep_domain_matches fcd_pseudo_synonyms fcd_role_matches fcd_rome_v2_matches fcd_rome_v3_matches **fcd_synonyms** **off_organisme_diplome_matches** **off_organisme_domaine_matches** **off_organisme_moyen_pedagogique_matches**

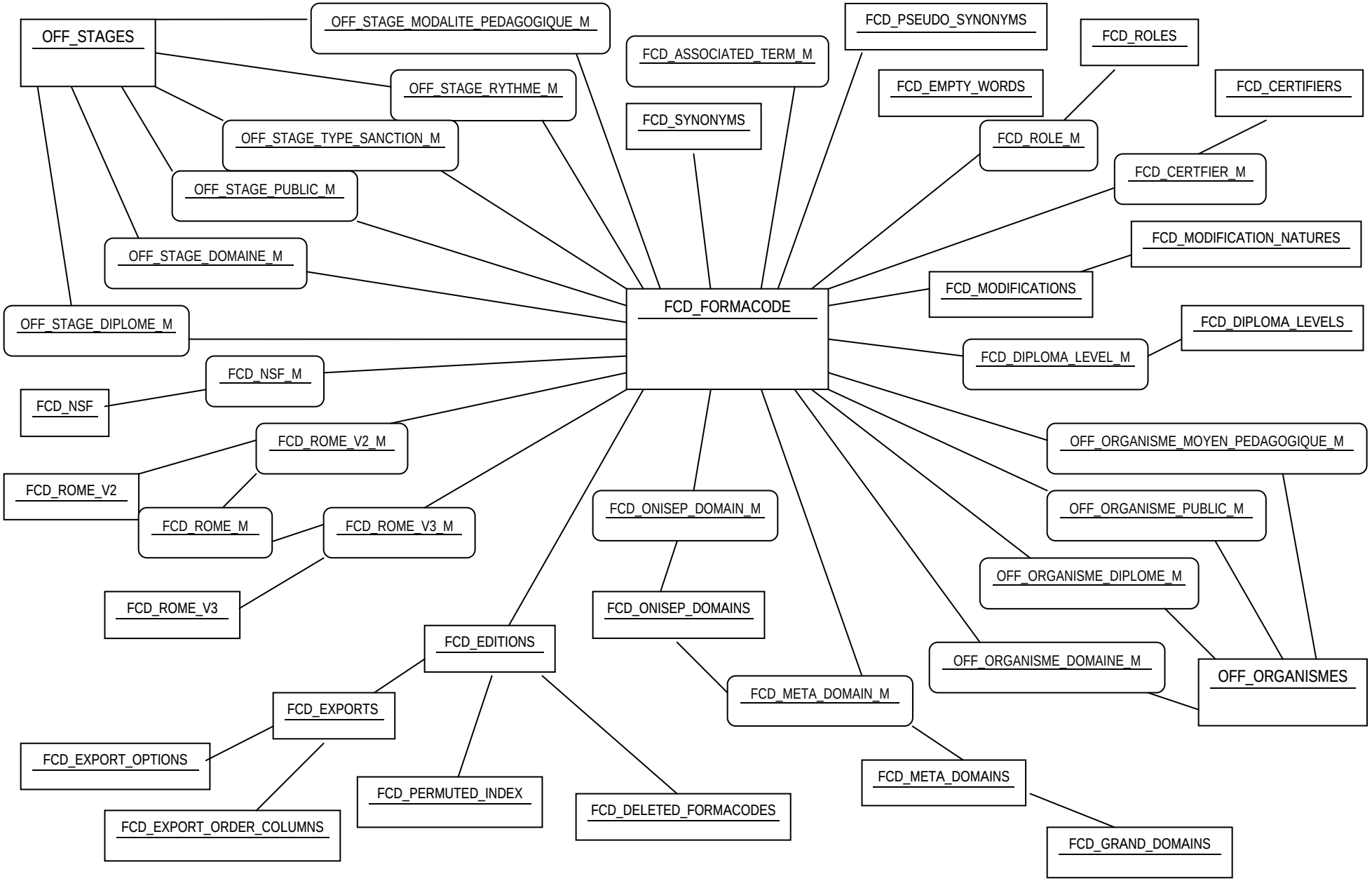
				off_organisme_public_matches
				off_stage_diplome_matches
				off_stage_domaine_matches
				off_stage_modalite_pedagogique_matches
				off_stage_public_matches
				off_stage_rythme_matches
				off_stage_type_sanction_matches
				off_stages
fcd_grand_domains	Grands domaines de formation de l'ONISEP (de la v5 d'IFM - Interface Formation Métiers)	14		**fcd_meta_domains**
fcd_html_formacode	Formacode formaté en HTML, pour le site. *(Apparemment ne sert plus ?)*	2752		
fcd_meta_domain_matches	Correspondance Formacode / Méta-domaine IFM - Interface Formation Métiers	5046	fcd_formacode (CASCADE) fcd_meta_domains (CASCADE)	
fcd_meta_domains	Méta-domaines de formation de l'ONISEP (de la v5 d'IFM - Interface Formation Métiers)	140	fcd_grand_domains (RESTRICT)	fcd_meta_domain_matches **fcd_onisep_domains**
fcd_modification_natures	Nature de modification	24		**fcd_modifications**
fcd_modifications	Modification réalisée sur un formacode	167	fcd_formacode (SET NULL) fcd_modification_natures (RESTRICT)	
fcd_nsf	Nomenclature des spécialités de formation (NSF)	93		fcd_nsf_matches
fcd_nsf_matches	Correspondances Formacode / NSF	12379	fcd_formacode (CASCADE) fcd_nsf (CASCADE)	
fcd_onisep_domain_matches	Correspondance Formacode / domaine ONISEP	2	fcd_formacode (CASCADE) fcd_onisep_domains (RESTRICT)	
fcd_onisep_domains	Domaine ONISEP	1301	fcd_meta_domains (RESTRICT)	**fcd_onisep_domain_matches**
fcd_permuted_index	Index permuté	15314	fcd_editions (CASCADE)	
fcd_pseudo_synonyms	Pseudo-synonymes	3348	fcd_formacode (CASCADE)	
fcd_role_matches	Correspondances entre le Formacode et les rôles	10166	fcd_formacode (CASCADE) fcd_roles (RESTRICT)	
fcd_roles	Rôle que peut avoir un terme du Formacode	18		**fcd_role_matches**
fcd_rome_matches	Correspondance ROME V2 / ROME V3	541	fcd_rome_v2 (CASCADE) fcd_rome_v3 (CASCADE)	
fcd_rome_v2	Métiers de la 2e version du ROME	466		fcd_rome_matches fcd_rome_v2_matches
fcd_rome_v2_matches	Correspondances Formacode / ROME V2	9318	fcd_formacode (CASCADE) fcd_rome_v2 (CASCADE)	
fcd_rome_v3	Métiers de la 3e version du ROME	531		fcd_rome_matches fcd_rome_v3_matches
fcd_rome_v3_matches	Correspondances Formacode / ROME V3	6182	fcd_formacode (CASCADE) fcd_rome_v3 (CASCADE)	
fcd_synonyms	Synonymes	22938	fcd_formacode (RESTRICT)	

Le sous-chapitre qui suit propose une modélisation des données d'après la structure extraite de la base MySQL.

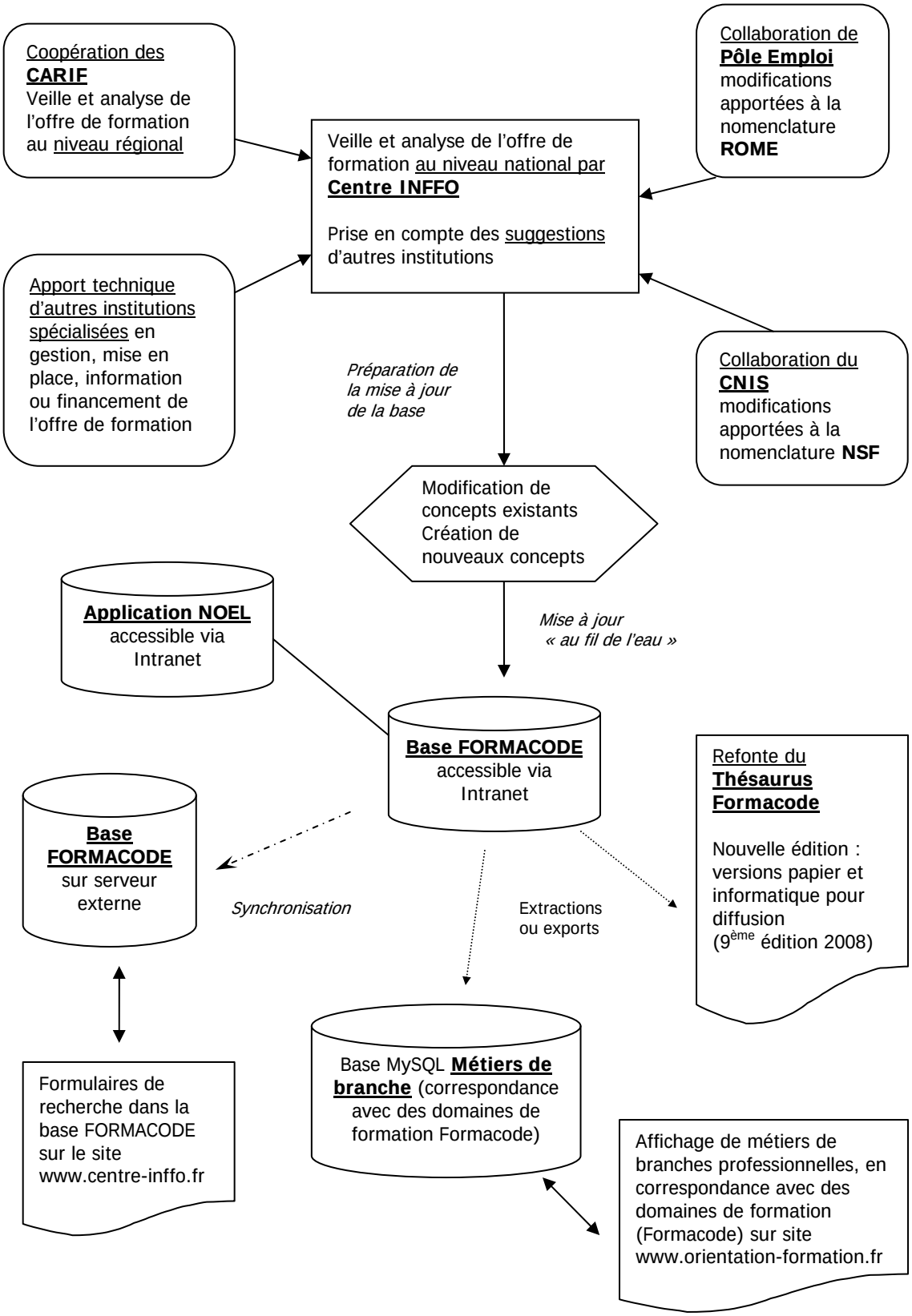
Pour plus de lisibilité : - les noms des tables sont reproduits en majuscules

- les noms des tables de jonction ne se terminent pas par « matches » mais par M.

5.3 Modèle des données du fichier Formacode



5.4 Workflow relatif au Formacode



6 Base de données FAR – Formation et Apprentissage en Régions

Nom de la base Type d'objet	Type de base	Service gestionnaire	Utilisation des données Publication sur le Web
Formation et Apprentissage en Régions (FAR)	Base MySQL	Mission Régions et Europe Utilisatrices et gestionnaires : 1 documentaliste 1 chargée d'études documentaires	• Recensement des aides régionales en vigueur en matière de formation professionnelle et d'apprentissage, ainsi que des contrats de projets et d'autres outils de programmation gérés par les Conseils régionaux. La **base FAR** est consultable en ligne : <http://www.centre-inffo.fr/spip.php?page=baseregions>.

6.1 Présentation de la base FAR

La base FAR - Formation et Apprentissage en Régions est réalisée par le Centre INFFO en collaboration avec les Conseils régionaux et les CARIF – Centres d'animation et de Ressources de l'Information sur la Formation. L'objectif principal de FAR est de présenter les **politiques régionales en matière de formation professionnelle et d'apprentissage**, de faciliter la lisibilité transversale des dispositifs en vigueur.

Elle recense donc

• les **aides et mesures** qui sont le reflet de la **mise en œuvre de la politique des Conseils Régionaux** dans les domaines suivants :

- formation professionnelle continue
- apprentissage
- enseignement professionnel
- enseignement supérieur
- insertion professionnelle

• ainsi que les **outils de programmation et de partenariats** (dispositifs techniques) :

- contrats de projets Etat-Région
- plans régionaux de développement des formations (PRDF)
- contrats d'objectifs territoriaux (COT)

- contrats d'objectifs et de moyens Apprentissage (COM)
- conventions de partenariats Région-partenaires sociaux.

FAR été créée en 2005 sous LORIS, puis les données ont migré sous une application maison développée sous MySQL en 2007, dont l'interface de saisie est accessible depuis l'Intranet.

La base est publiée sur le site internet du Centre INFFO <http://www.centre-inffo.fr/spip.php?page=baseregions>. Deux types de recherche sont proposés :

• une recherche libre par formulaire

• une recherche avancée au moyen de deux choix possibles de « **documents** » par cases à cocher :

- **aides** : puis possibilité de choisir un(e) ou plusieurs Région(s), Thème(s) et/ou Bénéficiaire(s)
- **outils** : puis possibilité de choisir un(e) ou plusieurs Région(s) et/ou outil(s).

Les résultats de recherche proposent une liste de « titres » d'aides et/ou outils, classés par Région, avec lien hypertexte menant directement vers la page du Conseil Régional concerné, décrivant une « aide » ou un « outil ».

Certains « **documents** », notion qui, dans FAR, regroupe les **aides** et les **outils**, sont des fichiers PDF hébergés sur le serveur du site Internet du Centre INFFO.

6.2 Constats sur FAR

• **Sur le contenu** :

- certains **outils de programmation et de partenariats** ne sont pas encore répertoriés dans FAR : les conventions de partenariat Etat-Région, et les conventions de partenariat Région-OPCA (Organismes Paritaires Collecteurs Agréés) ou Régions-FONGECIF (Fonds de Gestion du Congé Individuel de Formation), les conventions de création de CFA – Centres de Formation d'Apprentis, etc.

• **Sur les relations avec d'autres bases** du Centre INFFO :

- seuls quelques liens sont établis avec le **site Droit de la formation** (pour les PRDF).
- des liens pourraient être effectués entre FAR et la base **Juriweb** qui gère les données des « Fiches pratiques de la formation continue ». C'est-à-dire une liaison entre des dispositifs juridiques nationaux décrits dans Juriweb, et les dispositifs régionaux qui en sont à l'origine, répertoriés dans FAR. Ou une liaison entre des dispositifs régionaux décrits dans Juriweb, et les aides correspondantes répertoriées dans FAR (ex : pour les CTP - Contrats de Transition Professionnelle, pour le RSA - Revenu de Solidarité Active, etc.).

- une liaison pourrait être établie entre FAR et les **textes conventionnels** référencés sous **LORIS**, ou l'**état des négociations de branche** (données gérées dans une base FileMakerPro pour diffusion sur le site officiel Centre INFFO et le site Droit de la formation).

• **Sur l'exploitation des données** :

La base FAR est seulement publiée sur le Web. Aucune autre exploitation n'est effectuée, au moyen par exemple d'extractions ou d'exports. Des liens hypertextes prédéfinissant des requêtes dynamiques dans la base FAR, pourraient être insérés sur des pages du site Droit de la Formation, ou depuis des dossiers (documentaires) ou articles en ligne par exemple. Tout comme il existe de tels liens dans la base Juriweb qui gère « Les Fiches pratiques de la formation continue ».

6.3 Structure de FAR

Les champs disponibles dans chacune des tables de FAR ne sont pas rappelés. Ces éléments sont extraits de la base MySQL au moyen des outils d'administration mis en place sur le site Intranet par le développeur Web ayant développé l'application.

Liste des tables au 20/10/2009

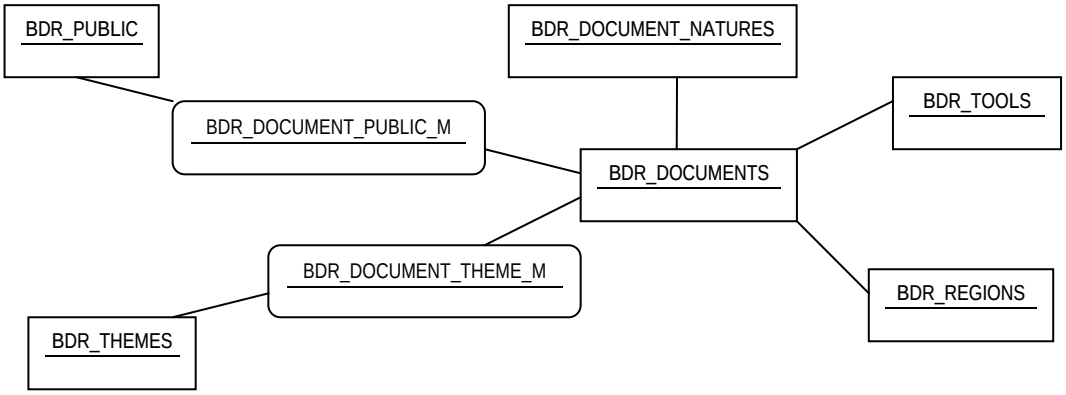
Nom	Commentaire	Nb.	A pour références	Est référencée par
bdr_document_natures	Nature de « document » : aide, dispositif, outil ou convention	4		**bdr_documents**
bdr_document_public_matches	Correspondance entre un document et un public	546	bdr_documents (CASCADE) bdr_public (CASCADE)	
bdr_document_theme_matches	Correspondance entre un document et un thème	609	bdr_documents (CASCADE) bdr_themes (CASCADE)	
bdr_documents	« Documents » : aides ou outils.	513	bdr_document_natures (RESTRICT) bdr_regions (RESTRICT) bdr_tools (RESTRICT)	bdr_document_public_matches bdr_document_theme_matches
bdr_public	Public	25		bdr_document_public_matches
bdr_regions	Région	26		**bdr_documents**
bdr_themes	Thème	39		bdr_document_theme_matches
bdr_tools	Types d'outils de programmation et de partenariats	5		**bdr_documents**

6.4 Modèle des données de FAR

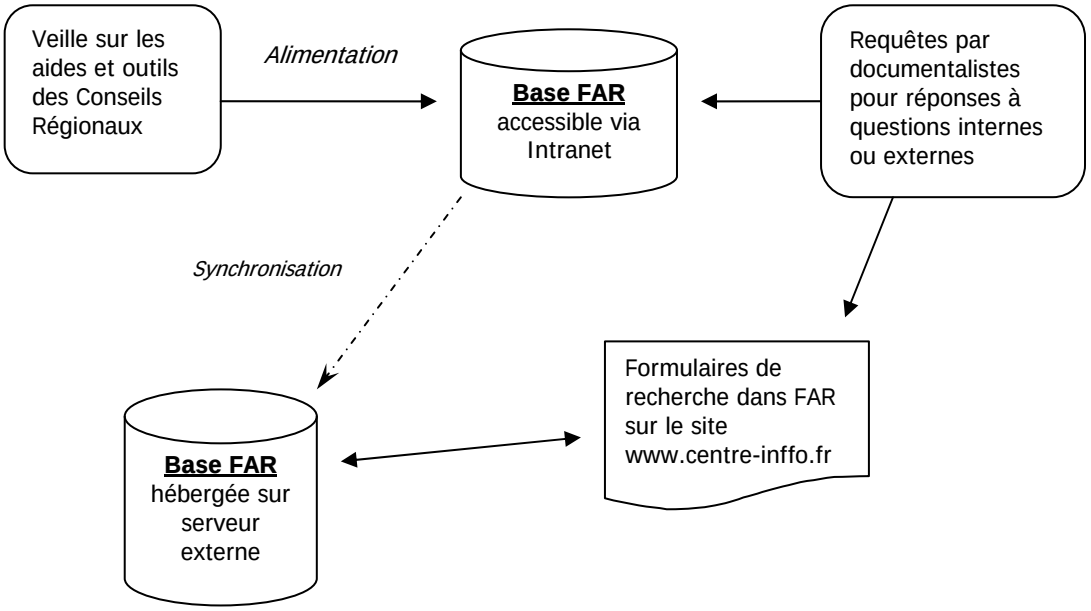
Le graphe suivant propose une modélisation des données d'après la structure extraite de la base MySQL.

Pour plus de lisibilité :

- les noms des tables sont reproduits en majuscules
- les noms des tables de jonction ne se terminent pas par « matches » mais par M.



6.5 Workflow du traitement des ressources de FAR



7 Base Annuaire des structures et contacts

Nom de la base et/ou Type d'objet	Type de base	Service gestionnaire	Utilisation des données Publication sur le Web
Annuaire des structures et contacts : anciennement appelé « Adresses utiles »	Base MySQL	Direction Juridique Observatoire Utilisatrice 1 Assistante de production de BDD	• **Annuaire des acteurs de la formation** <http://www.centre-inffo.fr/spip.php?page=annuaire&recherche_libre=Annuaire> mais dont les données sont aussi utilisées pour : • le site **Droit de la formation** <http://www.droit-de-la-formation.fr/spip.php?page=adresses> • et le **Portail Orientation-Formation** <http://www.orientation-formation.fr/>.

7.1 Présentation de la base Annuaire des structures et contacts

Il s'agit d'un annuaire d'organismes et institutions du domaine de la FPC, anciennement appelé « Adresses utiles », sous la responsabilité d'une Assistance de production de base de données.

Les données ont migré en 2008 depuis une base sous logiciel DORIS, abandonnée au profit d'une base MySQL développée en interne.

L'interface utilisateur est accessible via l'Intranet.

Ses données ont plusieurs destinations :

- l'**Annuaire des acteurs de la formation** publié sur le **site du Centre INFFO** <http://www.centre-inffo.fr/spip.php?page=annuaire&recherche_libre=Annuaire> ;

• la base de données interne **Juriweb** (cf. annexe 7 de ce mémoire) :

- pour la production de la version imprimée du guide juridique « **Les Fiches pratiques de la formation continue** » <http://www.centre-inffo.fr/-Les-Fiches-pratiques-de-la-.html> (listes de coordonnées d'organismes en annexes) ;

- pour les adresses utiles (OPCA, organismes agréés, DRTEFP…) consultables sur le site **Droit de la formation** <http://www.droit-de-la-formation.fr/spip.php?page=adresses>.

• le **Portail Orientation-Formation** <http://www.orientation-formation.fr/>.

• des **usages internes** exclusivement : informations stockées dans une table « Contacts », utilisées pour un suivi des partenariats, pour un suivi des prestations orales, pour constituer un annuaire des intervenants pour des prestations de formation (NB : les noms des tables liées correspondantes de ces **autres bases internes** sont exclues du tableau relatif à la structure de l'annuaire, reproduit ci-après).

Constat : il n'y a toutefois pas de lien avec le fichier commercial du Centre INFFO.

• la **base LORIS**, pour sa partie **textes conventionnels** : adresses des signataires des accords, adresses des OPCA – Organismes Paritaires Collecteurs Agréés, destinées à être exportées pour compléter la base des accords de branche « CPNFP » (négociation collective en matière de formation professionnelle) (cf. chapitre 2 de la Troisième partie de ce mémoire).

7.2 Structure de la base Annuaire des structures et contacts

Les champs disponibles dans chacune des tables de cette base ne sont pas rappelés. Ces éléments sont extraits de la base MySQL au moyen des outils d'administration mis en place sur le site Intranet par le développeur Web ayant développé l'application.

Liste des tables au 20/10/2009

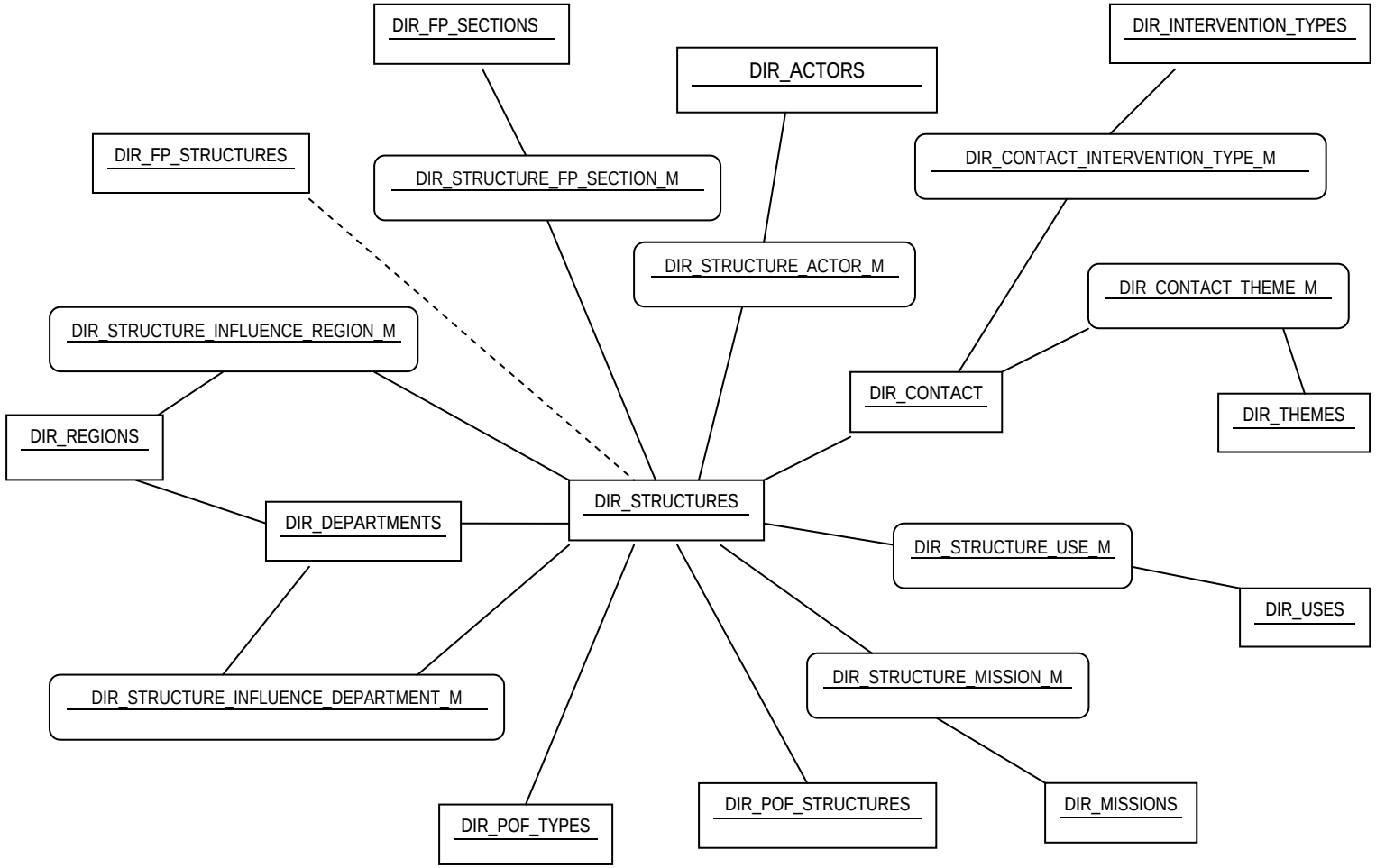
Nom	Commentaire	Nb.	A pour références	Est référencée par
dir_actors	Acteur	18		**dir_structure_actor_matches**
dir_contact_intervention_type_matches	Correspondance contact / type d'intervention	57	dir_contacts (CASCADE) dir_intervention_types (RESTRICT)	
dir_contact_theme_matches	Correspondance contact / thème	63	dir_contacts (CASCADE) dir_themes (RESTRICT)	

dir_contacts	Contacts du Centre INFFO (utilisation interne)	60	dir_structures (RESTRICT)	dir_contact_intervention_type_matches dir_contact_theme_matches
dir_departments	Département	105	dir_regions (RESTRICT)	dir_structure_influence_department_matches **dir_structures**
dir_fp_sections	Sections des FP - Fiches Pratiques de la formation continue	33		**dir_structure_fp_section_matches**
dir_fp_structures	Structure destinée à être affichée sur le site des FP	278		**dir_structures ?**
dir_intervention_types	Type d'intervention	5		**dir_contact_intervention_type_matches**
dir_missions	Mission	31		**dir_structure_mission_matches**
dir_pof_structures	Structure destinée à être affichée sur le POF - Portail Orientation-Formation	1398		**dir_structures**
dir_pof_types	Type de structure selon le POF	11		**dir_structures**
dir_regions	Région	31		**dir_departments** dir_structure_influence_region_matches
dir_structure_actor_matches	Correspondance structure / acteur	5884	dir_actors (RESTRICT) dir_structures (CASCADE)	
dir_structure_fp_section_matches	Correspondance structure / section Fiches pratiques	281	dir_fp_sections (RESTRICT) dir_structures (CASCADE)	
dir_structure_influence_department_matches	Correspondance structure / département d'influence	9861	dir_departments (CASCADE) dir_structures (CASCADE)	
dir_structure_influence_region_matches	Correspondance structure / région d'influence	1262	dir_regions (CASCADE) dir_structures (CASCADE)	
dir_structure_mission_matches	Correspondance structure / mission	7942	dir_missions (RESTRICT) dir_structures (CASCADE)	
dir_structure_use_matches	Correspondance structure / utilisation	9095	dir_structures (CASCADE) dir_uses (RESTRICT)	
dir_structures	Structure (ex-adresses utiles)	5812	dir_departments (RESTRICT) dir_pof_types (RESTRICT)	**dir_contacts** dir_structure_actor_matches dir_structure_fp_section_matches dir_structure_influence_department_matches dir_structure_influence_region_matches dir_structure_mission_matches dir_structure_use_matches
dir_themes	Thème	2265		**dir_contact_theme_matches**
dir_uses	Utilisation	5		**dir_structure_use_matches**

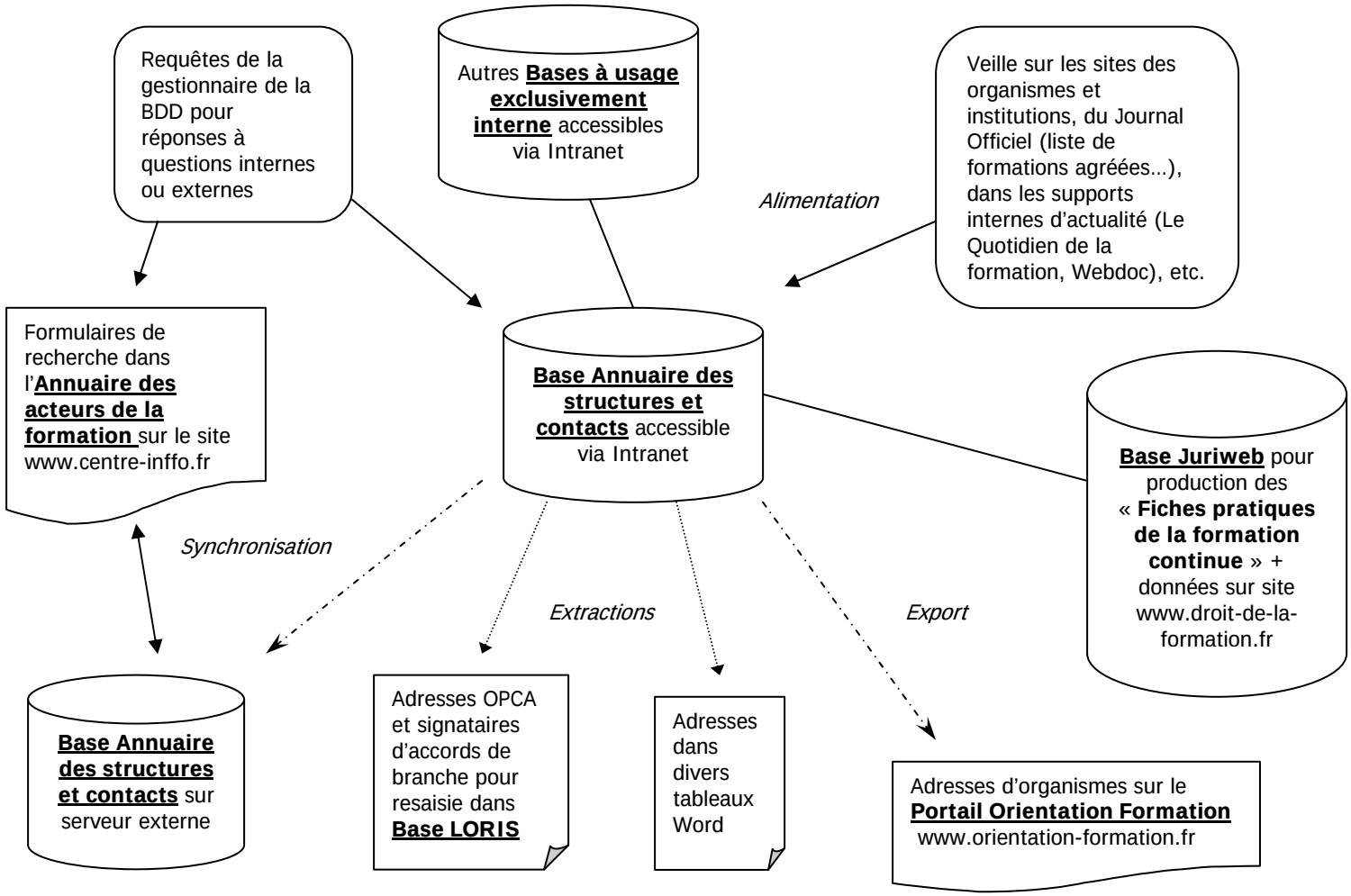
7.3 Modèle des données de la base Annuaire des structures et contacts

Le graphe qui suit, propose une modélisation des données d'après la structure extraite de la base MySQL.

Pour plus de lisibilité : - les noms des tables sont reproduits en majuscules
- les noms des tables de jonction ne se terminent pas par « matches » mais par M.


DIR_FP_SECTIONS
DIR_ACTORS
DIR_INTERVENTION_TYPES
DIR_FP_STRUCTURES
DIR_STRUCTURE_FP_SECTION_M
DIR_CONTACT_INTERVENTION_TYPE_M
DIR_STRUCTURE_ACTOR_M
DIR_CONTACT_THEME_M
DIR_STRUCTURE_INFLUENCE_REGION_M
DIR_CONTACT
DIR_THEMES
DIR_REGIONS
DIR_STRUCTURES
DIR_DEPARTMENTS
DIR_STRUCTURE_USE_M
DIR_USES
DIR_STRUCTURE_MISSION_M
DIR_STRUCTURE_INFLUENCE_DEPARTMENT_M
DIR_POF_TYPES
DIR_POF_STRUCTURES
DIR_MISSIONS

7.4 Workflow relatif à la base Annuaire des structures et contacts



Quatrième partie

Pour une évolution des modèles de données : optimiser la description des ressources et leur exploitation

1 Normes, standards et schémas de métadonnées : de nouvelles modélisations à envisager pour la description des ressources gérées au Centre INFFO

La troisième partie de ce mémoire a permis d'illustrer la démarche d'enquête menée sur quelques bases hétérogènes, produites et gérées par Centre INFFO, et d'en livrer le produit au moyen d'une présentation, de tableaux, de diagrammes, de schémas.
Deux autres bases ont également fait l'objet d'un début d'investigation. Les résultats en sont rassemblés en annexes 6 et 7, complétant le panorama.
L'ensemble des bases enquêtées pour les besoins de l'étude offre un bon aperçu des différentes ressources types gérées, ainsi que des structures actuelles de données, variées.

L'utilisation à venir de la technologie XML, faciliterait l'échange automatisé de contenu entre ces sous-systèmes d'information hétérogènes, en d'autres termes l'interopérabilité. Le chapitre 13 de la Première partie de ce mémoire, nous a permis de décliner les avantages et usages possibles de XML. Résumons les en quelques mots qui font l'éloge de ce langage : [32]

- lisibilité et simplicité (texte)
- autodescriptif et extensible, car utilisable dans tous les domaines d'application
- langage universel, non propriétaire
- balisé, donc structuré et hiérarchisé (structure arborescente)
- peut être validé (conforme par rapport à la DTD ou schéma, à laquelle ou auquel il est associé)
- garantit la séparation du contenu et de la forme.
- langage pivot : XML permet de puiser des données dans différents systèmes existants et d'en unifier la présentation.

A l'appui des concepts clés présentés en première partie, décrivant notamment la galaxie XML, nous pouvons dès lors décliner <u>plusieurs schémas de métadonnées</u> adéquats, développés depuis plusieurs années et utilisés dans des métiers spécifiques, évolutifs, disponibles librement, qui offriraient une modélisation selon la norme XML. Car comme l'avance Sylvie Dalbin [13], « les applications au plus près des utilisateurs finals des ressources doivent bien souvent articuler plusieurs schémas voire plusieurs répertoires de métadonnées ».

Et au sein d'une « structure d'ensemble sémantiquement cohérente », il faut prendre en compte les spécificités des différentes catégories de publics, et de ressources, mais aussi veiller à ne pas « réduire les potentialités d'exploitation » des réservoirs de métadonnées. Il faut donc garder à l'esprit qu'il ne faut pas réduire d'emblée le périmètre des éléments de données, c'est-à-dire en les enfermant dans un format trop spécifique au moment où l'on effectue ses choix. De façon à faciliter une articulation entre les schémas, et une adaptation possible aux besoins qui se trouveraient en périphérie.

Ci-après sont donc proposés des Schémas qui nous semblent recouvrir les données spécialisées gérées dans les bases enquêtées, les métadonnées utiles au repérage des ressources et à leur réexploitation ultérieure (recomposition, ré-éditorialisation).

1.1 Norme Dublin Core Metadata

Le **Dublin Core**, aussi appelé **DC**, est un jeu de métadonnées générique créé en 1995[64], devenu la norme ISO 15836[65] depuis 2003. Il est maintenu par la DCMI – *The Dublin Core Metadata Initiative* <http://www.dublincore.org/>[66]. [51]

L'objectif initial était d'avoir un ensemble de métadonnées utilisable par tous pour décrire de façon simple mais non ambiguë des ressources en ligne et améliorer la recherche sur leur Web.

Le **Dublin Core simple,** aussi appelé **Dublin Core non qualifié,** se compose d'un jeu de 15 éléments, chacun de ces éléments étant facultatif et répétable[67]. [52]

Élément (en français)	Element (en anglais)	Commentaire
1. Titre	Title	Titre principal du document
2. Créateur	Creator	Nom de la personne, de l'organisation ou du service à l'origine de la rédaction du document
3. Sujet ou mots clés	Subject	Mots-clefs, phrases de résumé, ou codes de classement

[64] L'initiative de métadonnées du Dublin Core (DCMI) a été définie à Dublin (Ohio, USA) en 1995.

[65] Sous la référence ISO 15836:2009 «Information et documentation - L'ensemble des éléments de métadonnées Dublin Core» <http://www.iso.org/iso/fr/catalogue_detail.htm?csnumber=52142>.

[66] Une liste de discussion pour les initiatives de métadonnées francophones autour du Dublin Core est disponible auprès de l' INIST - INstitut de l'Information Scientifique et Technique [64]

[67] Consultable en anglais à l'adresse <http://www.dublincore.org/documents/dces/>, une traduction en français, réalisée par Anne-Marie Vercoustre, de l'INRIA – Institut National de Recherche en Informatique et en Automatique, est consultable à l'adresse <http://www-rocq.inria.fr/~vercoust/METADATA/DC-fr.1.1.html>.

4. Description	Description	Résumé, table des matières, ou texte libre. Raffinements : table des matières (TableOfContents), résumé (Abstract)
5. Éditeur	Publisher	Nom de la personne, de l'organisation ou du service à l'origine de la publication du document
6. Contributeur	Contributor	Nom d'une personne, d'une organisation ou d'un service qui contribue ou a contribué à l'élaboration du document. Chaque contributeur fait l'objet d'un élément Contributor séparé
7. Date	Date	Date d'un évènement dans le cycle de vie du document
8. Type de ressource	Type	Genre du contenu
9. Format	Format	Type MIME, ou format physique du document
10. Identifiant de la ressource	Identifier	Identificateur non ambigu : il est recommandé d'utiliser un système de référencement précis, afin que l'identifiant soit unique au sein du site, par exemple les URI ou les numéros ISBN. Raffinement : Is Available At
11. Source	Source	Ressource dont dérive le document : le document peut découler en totalité ou en partie de la ressource en question. Il est recommandé d'utiliser une dénomination formelle des ressources, par exemple leur URI
12. Langue	Language	
13. Relation	Relation	Lien avec d'autres ressources. De nombreux raffinements permettent d'établir des liens précis, par exemple de version, de chapitres, de standard, etc.
14. Couverture	Coverage	Couverture spatiale (point géographique, pays, régions, noms de lieux) ou temporelle
15. Droits	Rights	Droits de propriété intellectuelle, Copyright, droits de propriété divers

La signification de certains éléments peut être précisée à l'aide de **raffinements (*refinements*)** ou qualificatifs pour affiner la description des données, restreindre la signification d'éléments, mais sans les changer fondamentalement. Une cinquantaine d'attributs, dont l'utilisation est facultative, ont ainsi été validés par la DCMI. Lorsque ces raffinements sont utilisés, on parle de **Dublin Core qualifié** (ou étendu). [48] [65]

Le DC est un format ouvert. Son utilisation est libre et la documentation concernant son utilisation est disponible sur Internet. Il est interopérable et compatible avec les principaux standards du Web tels que HTML, XML, RDF... Par exemple, les métadonnées du DC permettent de donner des informations à propos de pages Web XHTML et peuvent y être intégrées. [51] [57]

Devenu un **noyau de métadonnées de référence**, DC est de plus en plus utilisé, notamment par les bibliothèques numériques pour exposer leurs données en « OAI ». **OAI-PMH – *Open Archives Initiative – Protocol for Metadata Harvesting***[68], qui consiste en un protocole d'exposition et de collecte automatique de métadonnées utilisé pour les archives ouvertes. OAI-PMH requiert le Dublin Core non qualifié comme format minimal de données pour assurer une interopérabilité technique minimale[69]. [22] [44] [55] [65]

En vue de l'interopérabilité de ses bases de données, le Centre INFFO pourrait prévoir que pour chacune d'entre elles, le format de métadonnées choisi – notamment en cas de respect d'une DTD ou d'un Schéma XML – suive la norme Dublin Core, le format minimal DC et ses 15 champs. Cela favoriserait les passerelles entre les structures de données de chaque base.

XML Schema permet d'implémenter le modèle de métadonnées **Dublin Core en XML** <http://dublincore.org/schemas/xmls/>. Les noms d'éléments sont indiqués de la façon suivante :

```
<xs :element name= »title » type= »elementType » />
<xs:element name="creator" type="elementType" />
<xs:element name="subject" type="elementType" />
<xs:element name="description" type="elementType" />
<xs:element name="publisher" type="elementType" />
<xs:element name="contributor" type="elementType" />
<xs:element name="date" type="elementType" />
<xs:element name="type" type="elementType" />
<xs:element name="format" type="elementType" />
<xs:element name="identifier" type="elementType" />
<xs:element name="source" type="elementType" />
<xs:element name="language" type="elementType" />
```

---

[68] Site de l'Open Archives Initiative <http://www.openarchives.org/>.

[69] Mais il est vivement conseillé de décrire ses métadonnées sous le format, encodable en XML, le plus riche possible. Plus les métadonnées seront précises, plus les fonctionnalités de recherche développées seront fines. [55]

```
<xs:element name="relation" type="elementType" />
<xs:element name="coverage" type="elementType" />
<xs:element name="rights" type="elementType" />
```

La DCMI recommande le **Schéma XML DC Simple** version 2002-12-12 et le **Schéma XML DC qualifié** version 2008-02-11. [51]

Le jeu de métadonnées DC fait aujourd'hui l'objet d'un large consensus, utilisé dans de multiples applications, en version simple ou version qualifiée, et très souvent dans des profils d'applications spécifiques. Mais « il présente aussi des limites, notamment sur la gestion de collections de ressources, les métadonnées techniques, administratives et de préservation ». [65]

D'autres jeux de métadonnées lui sont complémentaires avec des objectifs plus spécifiques, orientés vers la description du contenu (ex : LOM), orientés métiers (ex : MODS, BiblioML, EAD, ONIX), ou encore liés à des formats de documents (ex : TEI, IPTC, MPEG). [65]

1.2 Schémas de description – Modèles applicatifs – Profils d'application basés sur XML : des schémas de métadonnées ouverts autorisant une exploitation variée des ressources

1.2.1 Schéma XML DC simple ou qualifié pour les « Fiches pratiques de la formation continue »

La base de données Juriweb nécessite encore une modélisation approfondie de sa structure actuelle. Une fois établie cette modélisation, une étude pourra alors être menée pour connaître plus précisément les besoins en métadonnées pour la description et la valorisation des « objets » gérés.

Nous préconisons d'emblée de pouvoir respecter au minimum un Schéma XML DC simple ou un Schéma XML DC qualifié[70] si des besoins de raffinements sont révélés. L'une ou l'autre solution assureraient le respect des champs minimaux prévus par la norme Dublin Core, et l'interopérabilité avec les Schémas retenus pour les autres ressources du Centre.

En vue d'une exploitation complémentaire, selon un Schéma XML, le découpage en unités permettrait d'imaginer pouvoir attribuer des identifiants à chaque portion de contenu de la ressource électronique que constituent « Les Fiches pratiques », au moyen par exemple d'un numéro d'identification (cf. par exemple des systèmes non étudiés dans ce mémoire comme

[70] Cf. <http://dublincore.org/schemas/xmls/>.

URI – *Uniform Resource Identifier* ou encore DOI – *Digital Object Identifier*...). Ainsi, par exemple pour chaque chapitre ou par ensemble de fiches, les métadonnées associées à chaque unité via son identifiant préciseraient les droits qui y seraient attachés (auteur, éditeur, accès, utilisation), et en faciliterait la représentation, la diffusion et la commercialisation séparée.

1.2.2 Norme LOM pour la description de ressources pédagogiques

1.2.2.1 LOM – Learning Object Metadata

Le **LOM – *Learning Object Metadata*** fournit un schéma de métadonnées recommandé par l'IEEE – *Institute of Electrical and Electronics Engineers*[71] au plan mondial pour les objets d'apprentissage ou objets pédagogiques quels qu'ils soient, numériques ou non.

Est considérée comme ressource pédagogique toute entité (numérique ou non) utilisée dans un processus d'enseignement, de formation ou d'apprentissage. Il peut s'agir de livres, de films, de scénarios pédagogiques, de sites Web, de logiciels, etc.

Le standard LOM a été conçu en 2002 et est enregistré sous la référence 1484.12.1-2002. Il est construit au dessus du Dublin Core de base et le complète par des extensions propres au domaine éducatif, notamment sur le contexte d'usage. Il est techniquement complexe, proposant au total une cinquantaine de champs, organisés en neuf rubriques, mais nombre d'entre eux sont facultatifs. [22]

Catégories du LOM :

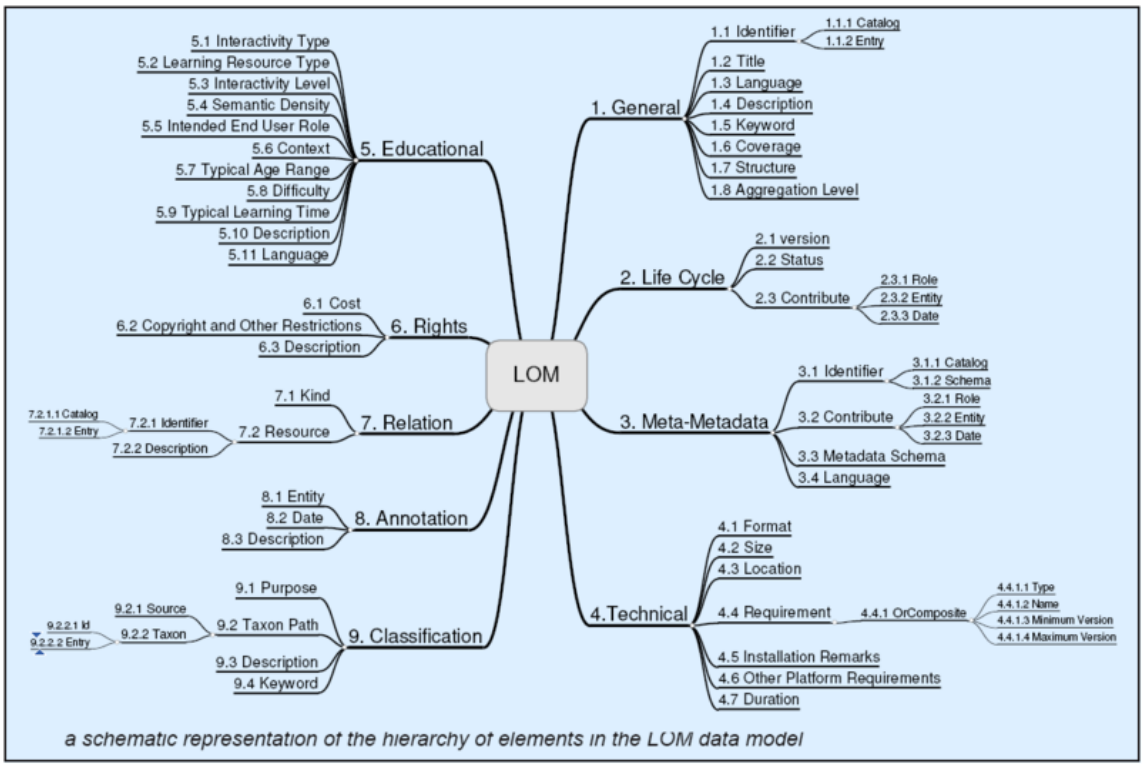
Catégories		Description
1. General	Généralités	Informations générales de description de la ressource, caractéristiques qui sont indépendantes du contexte d'utilisation (identifiant, titre, langue, description, etc.).
2. Life Cycle	Cycle de vie	Etat actuel de la ressource, les éléments sur le cycle de vie. Identifie les auteurs et contributeurs et leurs actions sur la ressource durant son évolution (version, entités contributives, etc.).
3. Meta-Metadata	Méta-métadonnées	Données détaillant la fiche descriptive elle-même plutôt que la ressource, c'est-à-dire les informations concernant l'enregistrement des métadonnées elles-mêmes (quels intervenants, dans quelles fonctions, dans quelles langues...).
4. Technical	Technique	Informations sur la nature, les caractéristiques

[71] Voir les informations relatives au groupe de travail correspondant sur le site de l'IEEE <http://ltsc.ieee.org/wg12/>, ainsi que le standard <http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf>.

		techniques de la ressource (format, taille, localisation, exigences techniques, durée).
5. Educational	Pédagogie	Informations relatives aux exploitations pédagogiques de la ressource (type de ressource, public / rôle de l'utilisateur, contexte d'utilisation, niveau, âge de l'apprenant).
6. Rights	Droits	Informations relatives à la propriété intellectuelle, qui spécifient les conditions d'utilisation de la ressource (coûts, droits d'auteur, etc.).
7. Relation	Relations	Contextualisation de la ressource : décrit, s'il y a lieu, la relation entre la ressource et d'autres ressources (comprend, est élément de, est en relation avec, etc.).
8. Annotation	Annotations	Espace dédié aux annotations, notamment les commentaires sur l'utilisation pédagogique de la ressource.
9. Classification	Classifications	Espace dédié aux informations de classification : vocabulaires contrôlés, listes d'autorités, thésaurus, etc.

Sources : d'après [22] et <http://fr.wikipedia.org/wiki/Learning_Object_Metadata> ou <http://fr.wikipedia.org/wiki/LOMFR>[72].

Schéma heuristique du LOM



a schematic representation of the hierarchy of elements in the LOM data model

[72] Voir également une fiche détaillée proposée par l'Université Paris V René Descartes <http://wiki.univ-paris5.fr/wiki/LOM>.

Source : <http://meta.wikimedia.org/wiki/File:LOM.PNG> {{GFDL}}

NB : Une version dynamique de la carte heuristique du schéma LOM, en anglais, est consultable à la page <http://dicos.ens-lyon.fr/metadata/lomall/lom/>.

Une version statique, en français, est aussi proposée à l'adresse <http://crdp.ac-bordeaux.fr/documentalistes/formation/m%C3%A9tadonnees_fichiers/LOM.pdf>. [16]

1.2.2.2 LOM-FR, profil d'application français

A partir de LOM, la communauté éducative française a décliné un profil d'application français, **LOM-FR**, qui a été normalisé par l'AFNOR – Association Française de Normalisation en décembre 2006 sous la référence NF Z76-040.

Le site collaboratif de l'implémentation de l'initiative LOM-FR <http://www.lom-fr.fr/> propose les résultats du travail du groupe AFNOR-SDTICE mettant en place la norme française LOM-FR. [41]

Tout comme le LOM, LOM-FR est destiné à la description d'objets/ressources pédagogiques. Ses objectifs sont les suivants :

- permettre une meilleure visibilité de l'offre d'objets pédagogiques, numériques ou non ;
- disposer de normes à respecter pour garantir la pérennité, l'interopérabilité, et une certaine ouverture des ressources numériques produites, en lien avec les plates-formes de distribution et de consultation de ces ressources.

Il concerne l'ensemble de la communauté éducative : communauté éducative française (enseignements primaire, secondaire, supérieur[73]) ; communauté de la formation (formation professionnelle, formation continue) ; documentalistes et bibliothécaires ; éditeurs, entreprises.

Sur le positionnement des éléments de LOM-FR au regard de DC et de LOM, voir le dossier en ligne proposé par le site EducNet <http://www.educnet.education.fr/dossier/metadata/ressources-educatives-numeriques/lomfr-learning-object-metadata>, qui propose également un schéma heuristique du LOM-FR <http://www.educnet.education.fr/data/lomfr.gif>. [62]

Selon Gérard Puimatto [22], « l'exploitation complète du LOM-FR est souvent juge lourde et contraignante ». Mais tout comme le LOM, la norme prévoit que de nombreux champs sont facultatifs, et les champs obligatoires recouvrent à peu près les données DC. De même il remarque que la « diffusion exclusivement commerciale » de LOM-FR « limite de façon considérable son accès et sa maîtrise par les professionnels concernés ».

---

[73] Cf. Une illustration de l'utilisation de LOM-FR à l'INRP - Institut National de Recherche Pédagogique et à l'ENS - Ecole Normale Supérieure de Lyon [68]. Et le portail Pr@tic de l'ENS-Lyon : <http://pratic.ens-lyon.fr/>.

Les **outils de création et d'édition de métadonnées** sont souvent intégrés aux plates-formes destinées à l'exploitation des métadonnées. Mais il existe des outils en ligne et hors ligne permettant de s'exercer pour les générer : [16] [22]

- Une boîte à outils proposée par le **SCÉRÉN-CNDP** (Services Culture Editions Ressources pour l'Education Nationale – Centre National de Documentation Pédagogique) <http://www.sceren.fr/standards/metadonnees/> offre en ligne deux formulaires d'éditions de métadonnées DC simples ou complètes au format HTML [novembre 2002].
- Le CNDP propose également l'outil en ligne **Educameta** <http://www.educameta.cndp.fr/>, permettant d'éditer et de convertir des métadonnées aux formats DC (en HTML ou XML), LOM et LOM-FR [mars 2007].
- **LOMPAD** <http://helios.licef.ca:8080/LomPad/en/index.htm> est un éditeur de métadonnées d'objets d'apprentissage, libre et téléchargeable, développé par la téléuniversité du Québec. Il permet d'éditer des métadonnées au format LOM.
- **LOMFRPAD** <http://adullact.net/projects/lomfrpad/> est un outil dérivé de LOMPAD, destiné à éditer des métadonnées LOM-FR. Il a été développé pour les catalogues Corrélyce[74] <http://correlyce.regionpaca.fr/correlyce/public.html> et Courdecol <http://www.courdecol13.fr/courdecol13/public.html>.
- **ORI-OAI**, Outil de Référencement et d'Indexation pour un réseau de portails OAI-PMH, <http://www.ori-oai.org/display/ORIOAI/ORI-OAI.ORG> propose un éditeur de métadonnées LOM-FR diffusé en logiciel libre.

1.2.2.3 Sup-LOMFR

Les ministères de l'Education nationale et de l'Enseignement supérieur et de la Recherche, dans le cadre de leur soutien à l'édition numérique pour l'éducation, ont souhaité, à des fins d'harmonisation, une adaptation de cette norme LOM-FR, appelée **Sup-LOMFR.** Il s'agit d'une spécialisation de LOM-FR, adaptée aux besoins et contraintes de l'enseignement supérieur. Son avantage est de faciliter le travail de description des ressources pédagogiques grâce à des critères spécifiques tels que « étude de cas », ou niveau d'étude sous la forme « BAC+… », etc. Le site Sup-LOMFR propose un guide d'utilisation en ligne <http://www.sup.lomfr.fr/>[75] [42]

---

[74] Corrélyce propose un guide de catalogage LOM-FR <http://www.correlyce.fr/actus/article86.html>.

[75] Voir également la plaquette de présentation disponible sur le site <http://www.sup.lomfr.fr/images/4/4e/Plaquette_presentation_suplomfr.pdf>.

1.2.3 Standards pour la description de l'offre de formation

1.2.3.1 LHÉO – Langage Harmonisé d'Échange d'informations sur l'Offre de formation

LHÉO – Langage Harmonisé d'Échange d'informations sur l'Offre de formation [40] [61] est un projet qui fut lancé dès 2001 pour assurer une meilleure lisibilité des offres de formation par les utilisateurs, ainsi qu'une fluidification des échanges d'informations sur l'offre entre les acteurs français de la formation professionnelle.

Langage de référence commun de description de l'**information sur l'offre de formation**, LHÉO est décrit par son site officiel <http://lheo.org/> [40] comme un « langage minimal permettant d'exprimer de façon cohérente une action de formation qui peut ensuite être lue, diffusée, classée et décrite par le plus grand nombre ». Il est placé sous licence Creative Commons. Il se compose de schémas (grammaires XML), d'un glossaire et de tables de références représentés sous la forme de fichiers XML. Le site précise qu'il est « accompagné par des outils (scripts, exemples) qui facilitent sa mise en œuvre dans les systèmes informatiques existants et minimisent les coûts d'intégration ».

Structure de LHÉO : en 3 cercles : [53]

- informations obligatoires en collecte et en diffusion ;
- informations facultatives en diffusion ;
- éléments d'extensions, pour les échanges particuliers d'institution à institution.

La dernière version (1.3.1) du standard LHÉO date d'avril 2009.

Ce référentiel est respecté par l'application NOEL – Nouvelle Offre en Ligne du Centre INFFO, qui fait d'ailleurs partie du Comité de pilotage de ce projet.

Le **Comité de pilotage** du projet LHÉO [40] est composé de représentants de nombreuses institutions et organismes du domaine de la formation professionnelle : le CNFPTLV - Comité National de la Formation Professionnelle Tout au Long de la Vie ; le Centre INFFO ; la DGEFP – Délégation Générale à l'Emploi et à la Formation Professionnelle ; le GIP CARIF Ile-de-France ; l'ONISEP – Office National d'Information sur les Enseignements et les Professions ; l'AFPA – Association Nationale pour le Formation Professionnelle des Adultes ; Pôle emploi ; la DARES – Direction de l'Animation de la Recherche, des Etudes et des Statistiques ; le Ministère de

l'Éducation nationale ; la DRTEFP Ile-de-France – Direction Régionale du Travail, de l'Emploi et de la Formation Professionnelle ; la FFP - Fédération de la Formation Professionnelle.

1.2.3.2 CDM – Course Description Metadata et CDMFR

CDM - *Course Description Metadata* est un standard d'échange des données sur l'offre de formation, en présentiel et à distance, entre les établissements d'enseignement supérieur européen. Initié par l'Université ouverte de Norvège, il a été défini dans le cadre du système européen de transfert et d'accumulation de crédits ECTS – *European Credits Transfer System*. Il est ainsi en cohérence avec le processus de Bologne[76], qui, visant une harmonisation de l'enseignement supérieur en Europe, a conduit à la mise en place du système LMD – Licence-Master-Doctorat. Ce système facilite ainsi la visibilité et la reconnaissance au niveau européen des formations supérieures, en permettant d'afficher les crédits ECTS proposés par les établissements. CDM a été adopté sous la forme de profils nationaux par plusieurs pays européens.

Structure de CDM, Schéma XML ayant 4 éléments constitutifs principaux : [53] [61]

- OrgUnit : l'unité organisationnelle, établissement d'enseignement responsable de l'offre ;
- Program : programme d'étude, ensemble de cours ou modules conduisant à l'obtention de diplôme, titre ou qualification, ou permettant la préparation d'un examen ou concours ;
- Course : UE - unité d'enseignement, cours ou ensemble de cours, ou EC – éléments constitutifs regroupant la description des cours et portant des crédits du système ECTS (validation) ;
- Person : les personnes impliquées dans cet enseignement.

CDMFR constitue la déclinaison française de CDM. Il s'agit d'une norme expérimentale en cours de rédaction. Le site collaboratif CDMFR <http://cdm-fr.fr/> a été créé pour accompagner la réflexion et la mise en place de CDMFR. [39]
Sous l'égide de la SDTICE - Sous-Direction des Technologies de l'Information et de la Communication pour l'Education des Ministère de l'Education Nationale et Ministère de l'Enseignement supérieur et de la Recherche, CDMFR vise à fournir aux établissements

[76] Cf. les sites européens
<http://europa.eu/legislation_summaries/education_training_youth/general_framework/c11088_fr.htm>

d'enseignement supérieur français un dispositif technologique permettant d'afficher une offre de formation[77] (initiale ou continue) de manière structurée et adaptable, de la partager à l'échelon national et à l'échelon régional, et de la proposer dans des dispositifs externes (portail étudiant, organismes d'orientation).

Sous forme de schéma XML, CDMFR permet la définition de métadonnées, spécifie la structure et la sémantique des concepts clés nécessaires à la description de programmes d'étude, d'unités d'enseignement, de cours de tout niveau, des établissements et des personnes impliquées[78]. [61]

Une carte heuristique dynamique du Schéma CDMFR est proposée en ligne <http://cdm-fr.fr/ressources/ressources/cartes-heuristiques-cdm-fr/cartes-cdmf>.

1.2.3.3 MLO - Metadata for Learning Opportunities

La norme européenne en cours d'élaboration, **MLO - *Metadata for Learning Opportunities*** (en français : métadonnées sur les opportunités d'apprendre) a pour ambition d'harmoniser les différentes spécifications en cours en Europe pour la description et l'échange d'informations sur les cours, le eLearning et les possibilités d'apprentissage. Ce modèle européen est valable quelles qu'en soient les caractéristiques (formelle, non formelle, qualifiante ou non...).

Le projet de norme est porté au niveau du CEN - Comité européen de normalisation par le comité technique 353 (*Information and Communication Technologies for Learning, Education and Training*) [79]. [53]

En l'état actuel, MLO distingue[80] : [61]

- le fournisseur (*Learning Opportunity Provider*),
- la spécification d'une opportunité d'apprendre (*Learning Opportunity Specification*),
- l'instance d'opportunité d'apprendre (*Learning Opportunity Instance*).

<http://ec.europa.eu/education/higher-education/doc1290_en.htm> et français <http://www.europe-education-formation.fr/bologne.php>.

[77] Voir une démonstration d'export CDM-fr au format XML pour quelques diplômes sur le site de l'Université des Sciences et Technologies de Lille 1 <http://formations.univ-lille1.fr/cdm/>.

[78] Pour des exemples d'utilisation, voir également : <http://www.educnet.education.fr/dossier/metadata/ressources-educatives-numeriques/cdmfr>.

[79] Cf. la note d'alerte de Bernard Blandin du FFFOD – Forum Français pour les Formations Ouvertes et à Distance, datée septembre 2008, à la page <http://fffod.over-blog.com/article-24024849.html>.

[80] Pour plus de détail, consulter le document en anglais <http://mjukis.blogg.skolverket.se/files/2008/10/mlo-ad-v5.pdf> ou la traduction en français qu'en

De même, MLO doit comporter deux parties :

- **MLO-AD** (*MLO-Advertising)* : décrivant les informations minimales à donner pour présenter une opportunité d'apprendre (=offre de formation) pour faciliter les échanges d'information sur le marché européen,
- **MLO-DESC** (*MLO-Description)* : décrivant d'une manière détaillée les opportunités d'apprendre.

La DGEFP – Délégation Générale à l'Emploi et à la Formation Professionnelle pilote un groupe de travail interministériel en charge de définir des tables de correspondance entre les deux standards français **LHÉO** et **CDMFR**, et **MLO**. Cette correspondance est « une condition essentielle pour accroître la visibilité de l'offre française au niveau européen et favoriser sa compétitivité dans le cadre du marché européen ». [53]

1.2.4 Modèle pour les références bibliographiques : BiblioML

1.2.4.1 Les inconvénients et lacunes des formats MARC à l'heure du Web

Depuis leur apparition en 1965, les formats MARC, dont UNIMARC, ont peu évolué. Ils s'intéressent au même type de contenu spécialisé – l'information bibliographique calquée sur la norme de catalogage ISBD, avec peu de données de gestion. La structuration est demeurée la même : à deux niveaux. De nouvelles zones ou sous-zones sont apparues pour répondre aux besoins des utilisateurs, les bibliothèques, marquant une tendance à structurer plus finement.
« Le Web, les moteurs de recherche et l'arrivée massive de documentation électronique ont radicalement transformé les usages en termes de recherche documentaire. » [49]
Mais aujourd'hui les catalogues traditionnels des bibliothèques manquent de visibilité sur le Web, souffrent de la comparaison avec une certaine offre commerciale (Google, Amazon, etc.), et ne sont pas assez tournés vers l'usager (absence de recherche fédérée ou de Web services).
Ce qui manque aux standards MARC :

- une structuration des données avec davantage de niveaux hiérarchiques,
- une visibilité sur le Web et une interopérabilité des données, alors que MARC et les normes qui lui sont associées ne sont pas des standards du Web (ISO2709, Z3950),
- les données de gestion, non prévues avec MARC sont pourtant indispensables pour le traitement de la documentation numérique : droits, archivage, conservation.

---

propose le CREFOR - Centre de Ressources pour la Formation de Haute-Normandie <http://www.crefor-hn.fr/drupal/system/files/users/documentalistes/MLO_Fr.pdf>.

Selon Thierry Clavel de l'ENSSIB – Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques, «MARC contribue à isoler les catalogues », et ne peut les faire évoluer. [49]

1.2.4.2 Les successeurs de MARC : enrichir les liens entre éléments, faciliter le partage et l'échanges de données, accroître la visibilité de fonds bibliographiques

Pour succéder à MARC, des outils XML sont apparus. **MARC-XML** est un Schéma créé en 2000 et maintenu par la Bibliothèque du Congrès permettant de convertir des notices MARC 21 en XML.

De même, la récente norme **MARCXchange** (ISO 25577[81]) permet de convertir en XML les notices de différents formats MARC, dont UNIMARC. [49]

Par exemple, utilisant ce principe, le SUDOC – Système Universitaire de Documentation <http://www.sudoc.abes.fr/> fournit les notices de son catalogue à Google pour Google Scholar. Il s'agit en fait d'un habillage en XML de MARC, avec ses zones, ses sous-zones et ses indicateurs. [71]

L'indexation du SUDOC par Google est un élément de la stratégie globale de l'ABES, Agence Bibliographique de l'Enseignement Supérieur, pour assurer une plus grande visibilité aux ressources des bibliothèques universitaires françaises sur le Web. Cette stratégie porte à la fois sur l'amélioration du SUDOC Web et sur l'ouverture des données du SUDOC sur le Web. [43]

Dans le même ordre d'idée, l'INIST – Institut de l'Information Scientifique et Technique a fait évoluer ses notices bibliographiques d'un format UNIMARC-texte à un format UNIMARC-XML afin d'exploiter les possibilités d'arborescence et de liens entre éléments. Et un format d'échange simplifié, **EXODIC**[82], profil d'application Dublin Core, est utilisé en sortie vers des serveurs traditionnels ou vers Google et Google Scholar, et en entrée dans le cadre de partenariats avec des laboratoires ou des éditeurs de revues pour alimenter ses bases de données. [65]

---

[81] Voir sur le site de *The Library of Congress* <http://www.loc.gov/standards/iso25577/>.
[82] EXODIC - *Exchange object description for INIST citations* <http://exodic.inist.fr/>.

1.2.4.3 Focus sur les avantages du format BiblioML

BiblioML – *Bibliographic Markup Language* a été élaboré en 1999 par la Mission de Recherche et de Technologie du Ministère de la Culture français, qui souhaitait mettre en place un outil commun de gestion des références bibliographiques de ses applications XML (dossiers d'inventaire, dossiers d'artistes, rapports de fouilles, etc.). [38] [45]

Il s'agit d'une **DTD XML**[83] permettant : [49]

- de convertir au format XML des données bibliographiques et d'autorités initialement saisies au format UNIMARC,
- ou de créer directement en XML des notices complètes conformes à l'ISBD.

Le site officiel de BiblioML <http://90plan.ovh.net/~adnx/biblioml/doku.php> [38] précise que cette application XML est en fait composée de deux formats XML pour l'échange d'enregistrements bibliographiques et d'autorités entre applications, **BiblioML** et **AuthoritiesML** basés sur les formats bibliographique et d'autorités UNIMARC. Les champs et sous-champs de BiblioML correspondent à ceux d'UNIMARC. Les étiquettes et codes de sous-zones sont remplacés par des mots. BiblioML et AuthoritiesML ont « la même granularité que le format UNIMARC d'origine c'est à dire les mêmes éléments de base. Mais le marquage est utilisé pour indiquer de manière lisible, non seulement la structure de l'enregistrement bibliographique, mais aussi la sémantique précise de chaque élément. »

Exemple d'utilisation : la Bibliographie nationale française <http://bibliographienationale.bnf.fr/> qui rassemble les notices bibliographiques des documents édités ou diffusés en France, et reçus par la BNF – Bibliothèque Nationale de France au titre du dépôt légal.

1.2.4.4 Avenir du catalogue de la bibliothèque du Centre INFFO

Adopter **BiblioML pour le catalogue de la bibliothèque** permettrait de :

- changer de format pour s'ouvrir au Web et offrir davantage de services aux usagers, tout en gardant la richesse descriptive proposée par UNIMARC.
- garder sa fonction première liée à la mission du centre de documentation d'identification, de recensement, et de localisation des documents.
- maintenir sa qualité, gage de pérennité du fonds documentaire.

---

[83] La version 0.6 alpha 1 de **BiblioML** constitue la première étape de la nouvelle version expérimentale. Elle est fournie sous forme d'un **XML Schema**.

1.2.5 Modèle pour les actualités et articles journalistiques : NewsML

NewsML est une spécification XML développée par l'IPTC – *International Press Telecommunications Council* <http://www.newsml.org/>[84] pour la transmission et l'échange des informations d'actualités.

Il existe deux types :

NewsML 1.x : La première version 1.0 a été publiée en 2000, et l'actuelle version 1.2 date d'octobre 2003.

NewsML-G2 : qui se présente comme un jeu de métadonnées enrichi. NewsML-G2 2.0 a été publié au printemps 2008, et la version actuelle est NewsML-G2 2.1.

NewsML est en plein essor. Il est d'ores et déjà utilisé par les agences de presse (AFP – Agence France-Presse[85] <http://www.afp.com/>, et Reuters) pour la transmission des dépêches et l'automatisation des fils d'agence. Il est conçu pour l'échange des textes, graphiques, photos, séquences audio, vidéo et animations. [67]

L'intérêt de **NewsML** pourrait être à étudier pour les articles publiés par Centre INFFO dans ses revues « Le Quotidien de la formation », « INFFO Flash » devenu en 2009 « L'Inffo formation », ou directement sur le nouveau site Web « Actualité de la formation »). De même, NewsML pourrait intéresser les informations variées actuellement diffusées chaque jour en interne via l'outil SPIP (appelé « Webdoc » sur l'Intranet) par le Centre de documentation, sachant que certaines de ces informations sont amenées à être reprises sur des sites Web de Centre INFFO[86].

---

[84] ou <http://www.iptc.org/cms/site/single.html?channel=CH0087&document=CMS1206527546450>.

[85] Un document en anglais « NewsML for dummies » de Laurent Meur, publié en 2001, se présente comme une introduction à NewsML et son usage à l'AFP <http://xml.coverpages.org/NewsMLForDummies.pdf>.

[86] C'est le cas actuellement pour le signalement de textes publiés au « Journal officiel Lois et Décrets » <http://www.droit-de-la-formation.fr/spip.php?page=textesofficiels-rubrique>.

1.2.6 SKOS - Simple Knowledge Organization System, langage de représentation pour les thésaurus

SKOS - *Simple Knowledge Organization System* est un langage de représentation des thésaurus (avec leurs relations) et autres vocabulaires contrôlés et structurés, comme les index, les plans de classement, les taxonomies. [50] [70]

Le principal objectif de SKOS est de permettre la publication facile de vocabulaires structurés pour leur utilisation dans le cadre du Web sémantique. Construit sur la base du langage RDF, SKOS peut être encodé en RDF/XML[87]. Il a le statut officiel de recommandation du W3C depuis août 2009[88].

Le Centre INFFO produisant deux langages structurés, le **Thésaurus de la formation** et le **Thésaurus de l'offre de formation Formacode**, et gérant d'autres listes de descripteurs contrôlés, comme **ThesJuris**, pourrait être intéressé par le Schéma SKOS en XML.

---

[87] RDF/XML est une syntaxe XML particulière pour RDF.

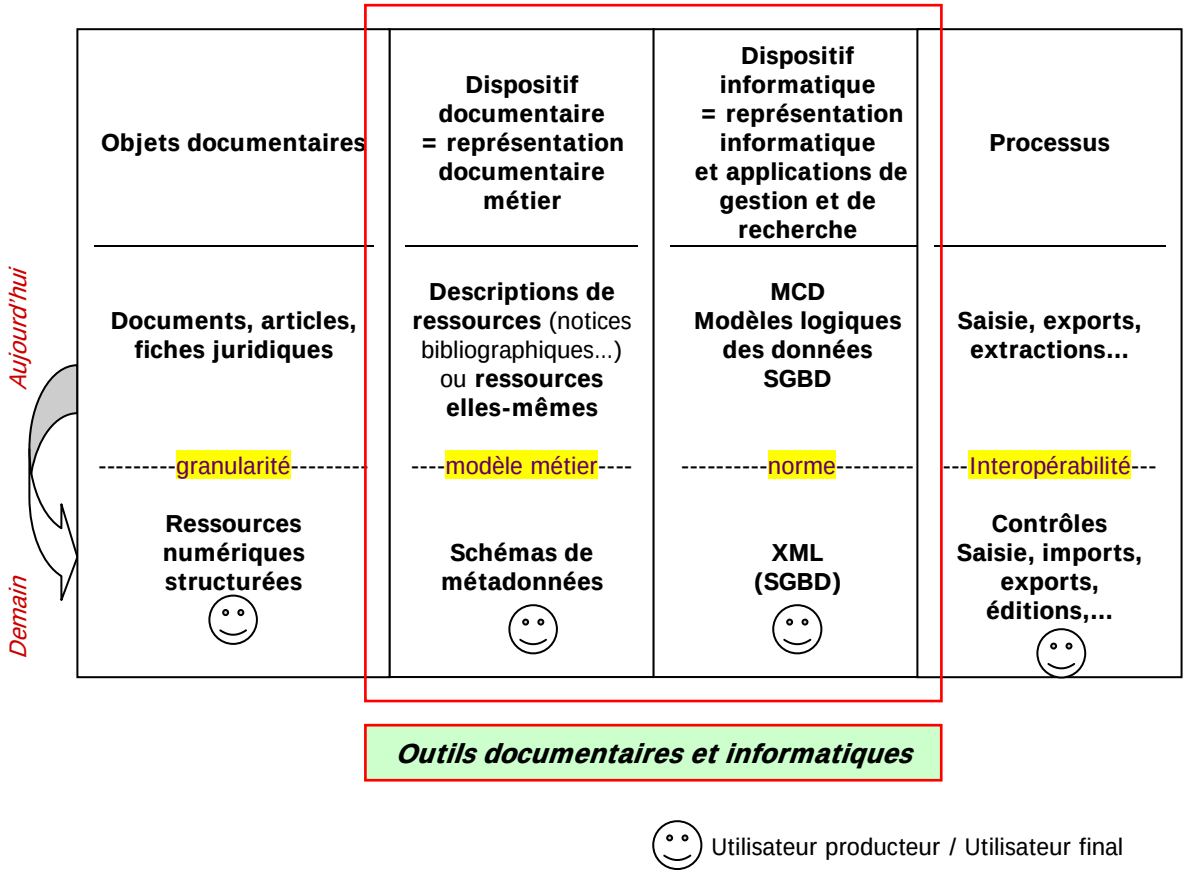
[88] Voir le site du W3C : <http://www.w3.org/2004/02/skos/> et <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>.

2 Des métadonnées pour la représentation et l'accès aux informations gérées dans les bases : perspectives, réflexions et questions

2.1 Perspectives pour le Système d'information documentaire (SID) : de la situation actuelle à la situation future

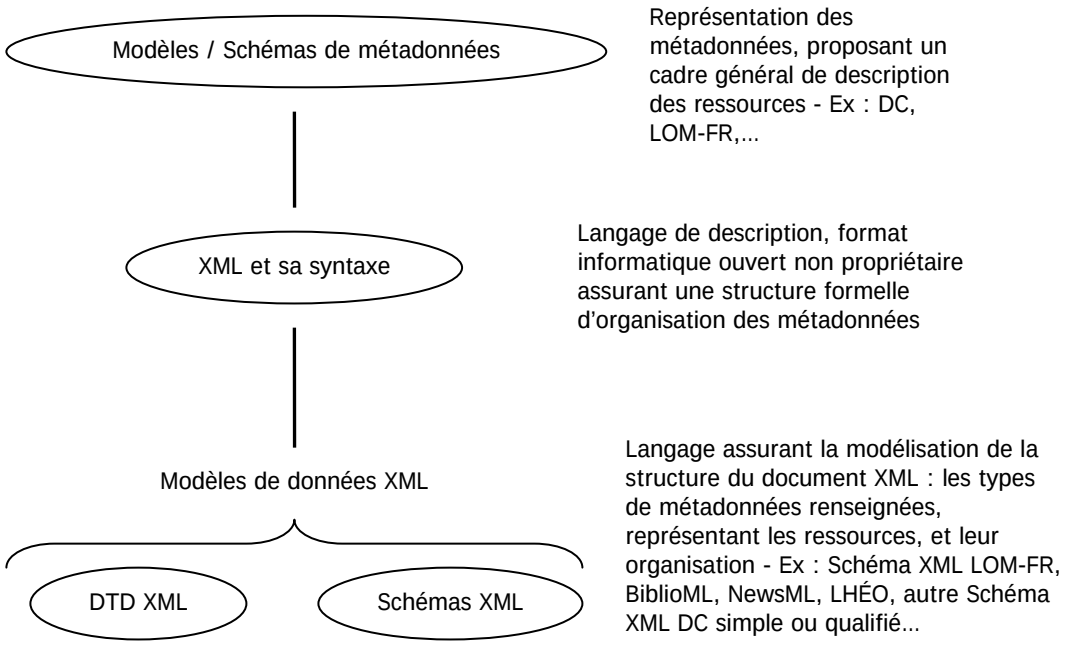
Des standards et normes variés pourraient être utilisés pour représenter les objets actuellement gérés dans des bases hétérogènes. Ils ont été détaillés tout au long du chapitre précédent. Ils concerneraient : **les descriptions de ressources** : BiblioML, LOM-FR, LHÉO ; ou bien **les ressources elles-mêmes** : NewsML, Schéma XML DC simple (ou qualifié) pour les données de la base Juriweb (« Les Fiches pratiques de la formation continue » Cf. Annexe 7).

2.1.1 Etats du Système d'information du Centre INFFO, aujourd'hui et demain



Le schéma qui précède s'inspire d'une présentation de Sylvie Dalbin [12]. Il vise à synthétiser et ordonner les thèmes qui sont en jeu, jusqu'ici abordés. Il fait ressortir les objectifs visés et les différents niveaux d'exigence, qui sont autant d'enjeux.

Concrètement, « à l'écran », les « vues » que l'utilisateur aurait sur les données gérées par le système, seraient des **descriptions de ressources** (une notice bibliographique, une notice descriptive de ressource pédagogique) ou les **ressources elles-mêmes** (une fiche juridique, un article d'actualité). Tandis que « derrière l'écran », ces vues seraient assurées par des **métadonnées structurées**, c'est-à-dire des éléments de données, regroupés et hiérarchisés selon des **modèles ou schémas de métadonnées** (DC, LOM-FR...), qui seraient eux-mêmes implémentés en XML selon des **Schémas XML** appropriés pour la représentation de chaque type de ressource gérée (BiblioML, NewsML, Schéma XML LOM-FR, LHÉO, Schéma XML DC simple ou qualifié,...).



2.1.2 Concilier plusieurs jeux de métadonnées

Plusieurs jeux de métadonnées (cf. chapitre 1 de cette Quatrième partie du mémoire) ont été sélectionnés car ils reflètent la **diversité** des ressources gérées dans les bases et peuvent en exprimer toute la **richesse**. Ces modèles métiers devraient être **évolutifs**, car amenés à être complétés pour les besoins internes, mais aussi, une fois mis en place, devant continuer de

respecter des standards et normes, eux-mêmes en évolution, pour garantir une interopérabilité avec des partenaires externes.

Mais il faut noter qu'au-delà de la **diversité des modèles métier**, la complexité du système résiderait dans les **relations** que ces modèles auraient entre eux, c'est-à-dire l'**articulation** qui pourrait être établie entre les DTD et les Schémas XML correspondants.

Les **nombreux standards de métadonnées disponibles**, utiles dans les différents contextes de création et de gestion de documents numériques, posent en effet des problèmes théoriques et pratiques pour faire jouer l'interopérabilité et l'échange de données. Sur le plan technique le passage d'un jeu de métadonnées à un autre, en utilisant les mécanismes de transformation XML, est tout à fait possible et de plus en plus d'outils sont proposés à cet effet sur Internet[89]. En revanche, sur le plan conceptuel, la question est plus complexe car ce passage implique des choix sur l'établissement d'équivalences entre des éléments spécifiques et des éléments plus génériques, ou encore sur les valeurs des éléments si les pratiques diffèrent. [65]

Un **jeu « transversal »** comme le **Dublin Core** a explicitement cet objectif d'interopérabilité et d'échange, c'est pourquoi nous le proposons comme **passerelle** entre les modèles de données des différentes ressources gérées à Centre INFFO.

2.1.3 Processus de production, modèles de données, passerelles et convergence

Du côté des **processus de production**, ils seraient donc bien entendu impactés. Il est attendu une **interopérabilité** accrue entre les éléments du système d'information, un **accès** amélioré aux ressources, pour une **exploitation** moins limitée qu'actuellement, plus variée et même démultipliée car favorisée par la(les) **passerelle(s)** autorisé(es) entre les modèles de données. L'**articulation des schémas** pourrait se situer notamment dans les éléments minimaux du jeu Dublin Core (DC simple) figurant dans tous les Schémas et DTD qui auraient été adoptés à dessein.

---

[89] A titre d'illustrations, l'OCLC - *Online Computer Library Center* développe des outils assurant une passerelle entre différents jeux de métadonnées : par exemple entre MARC et Dublin Core, MARC-XML et DC-XML, MARC et ONIX, etc. Voir <http://www.oclc.org/research/activities/schematrans/default.htm> et <http://www.oclc.org/research/activities/past/orprojects/schematrans/schematrans.htm>.
D'autres organismes peuvent aussi être cités, tels la *Library of Congress* dans le domaine archives-bibliothèques et ses nombreux standards pour les ressources numériques <http://www.loc.gov/standards/>, proposant des *« crosswalks »*, tableaux de concordance. Ou encore, dans le domaine des archives AJLSM <http://www.ajlsm.com/> et son logiciel PLEADE ou Archivists' Toolkit <http://archiviststoolkit.org/>.

Dans une réflexion sur la concordance avec le Dublin Core (DC), Sylvie Dalbin [50] alerte toutefois sur les difficultés que constitue une **mise en correspondance de jeux de métadonnées**. Car « il ne s'agit pas d'**aligner** strictement une donnée d'un jeu à une métadonnée d'un autre jeu, mais de rechercher la meilleure **correspondance** »[90], l'objectif étant qu'il ne faut pas perdre d'informations, de métadonnées qui pourraient être utiles ultérieurement.

La **convergence des modèles documentaires** serait aussi conditionnée par l'**harmonisation** de certaines **listes d'autorité**[91], qui pourraient être communes ou respecter si elles sont distinctes, un certain formalisme dans le nommage. Par exemple, la façon de nommer les organismes (auteurs, éditeurs, organismes d'information), les personnes physiques (auteurs), les types de supports, les types de données, etc.

Les **langages documentaires** utilisés pour l'indexation sont aussi un facteur déterminant. Par l'interopérabilité qu'ils pourraient présenter entre eux, ou la valorisation de ceux qui sont le plus formalisés, à savoir les thésaurus.

• Ainsi, le **Thésaurus de la formation** pourrait être le langage documentaire commun utilisé pour la description des contenus de ressources bibliographiques (LORIS), mais aussi des produits pédagogiques multimédias. Quant aux articles (Le Quotidien), actualités quotidiennes (Webdoc), descriptions d'aides et d'outils régionaux (FAR), les termes nécessaires à leur indexation, en nombre bien moindre, pourraient néanmoins être mis en correspondance, ou harmonisés si nécessaire avec les descripteurs du Thésaurus de la formation.

• En revanche, pour ce qui concerne **ThesJuris**, la liste de descripteurs et leurs synonymes utilisés pour l'indexation des « Fiches pratiques de la formation continue » dans la base Juriweb, la question est plus délicate. Il existe effectivement des différences sémantiques entre certains de ces termes et d'autres, homonymes, présents dans le Thésaurus de la formation. Il s'agirait donc plutôt de réaliser un travail de recherche des meilleures correspondances possibles.

• Le **Thésaurus de l'offre de formation Formacode**, quant à lui, propose non seulement une liste fournie de descripteurs pour les domaines de formation, mais aussi une grande variété de termes pour toute offre de formation dans ses listes annexes. Il pourrait ne plus seulement être utilisé pour les Bases Organismes de formation et Stages, mais aussi pour plusieurs métadonnées de la Base Produits pédagogiques multimédias (domaines de formation, publics, niveaux de formation, modalités pédagogiques...). Il faudrait d'ailleurs se repencher sur

---

[90] Sur le rapprochement de Schémas et l'interopérabilité, lire [46] et [47].

[91] Concernant l'évolution des listes d'autorité utilisées comme « lexiques contrôlés », Sylvie Dalbin parle de « réservoirs de vocabulaires désambiguïsés »... [12]

l'interopérabilité entre le Formacode et le Thésaurus de la formation, utilisé pour les ressources bibliographiques sous LORIS, pour certains descripteurs qui recouvrent les mêmes notions.

Sur la délicate question de l'**interopérabilité entre systèmes d'indexation**, l'auteure d'un mémoire de Master, Marie Peterlongo [68], s'est penchée sur la correspondance entre des taxonomies disciplinaires utilisées par des enseignants et la Classification décimale de Dewey, pour l'indexation de ressources pédagogiques dans le cadre d'une utilisation du profil d'application LOM-FR. Citant plusieurs auteurs, elle rapporte qu'une **harmonisation de langages documentaires** est « soumise à des conditions draconiennes de similarité sémantique et structurelle ». Pour organiser l'interopérabilité et tendre vers une comptabilité entre différents langages documentaires, plusieurs solutions sont possibles :

- la mise en équivalence ou correspondance (*mapping*) ;
- l'intégration ou la fusion des langages (*merging*) ;
- l'utilisation d'un langage faisant transition ou de commutation (*switching language*) ;
- le développement d'un langage de type méta, super ou macro-thésaurus.

2.1.4 Architecture du SID : distinguer production et recherche

Un point de réflexion questionne **l'architecture du système**. Selon Sylvie Dalbin, il faut distinguer la **production** de données, de la **recherche** sur ces données. [12]
Dans le même ordre d'idée, Bertrand Calenge aussi conseille que « les bibliothèques distinguent dans leurs cahiers des charges la gestion bibliographique ou numérique de leurs ressources et catalogues, et la question des outils de requêtes sur ces ressources ». [3]
Une exploitation directe par un moteur de recherche des données XML peut être envisagée.
Et cette recherche pourrait même être élargie à d'autres ressources existantes à Centre INFFO. Une **recherche fédérée** sur d'autres ressources sera à envisager, au moins celles produites **en interne**, informations publiées sur les sites Internet du Centre INFFO (portail du Centre, site Actualité, site Droit, etc.) ou sur son Intranet. Ces informations sont actuellement au format HTML ou sous la forme d'objets attachés (aux formats .doc .pdf .ppt). C'est le cas de bibliographies, dossiers documentaires, supports pédagogiques, études, rapports… Pour toutes ces autres ressources, qui sont autant d'objets documentaires, l'on devrait s'interroger sur leur gestion, leur traitement, leur structure et les métadonnées qui les accompagnent… toujours pour ces mêmes questions d'**interopérabilité et d'économie** relatives à leur exploitation.

2.2 Hypothèses au regard de la valorisation des données

2.2.1 Une collection de ressources complémentaires et enrichies à penser dans son contexte

Une collection de ressources est à penser dans son contexte : elle s'inscrit dans un environnement complexe, économique, social et technique notamment, et pour mieux prendre en compte les usages, les publics utilisateurs et les métiers qui interviennent dans sa constitution et sa gestion, elle doit être considérée plus comme un processus que comme un état. [3]

L'application de la méthode dite de « **l'arbre des causes** » permettrait en cela de mieux visualiser la « richesse » des causalités et interactions ayant mené à un résultat, un constat. Comme illustration, voici quelques exemples d'application formulés sous forme de questions :

- Pourquoi cet effort documentaire particulier ne produit-il pas le résultat espéré ?
- Qu'a-t-il fallu pour que tel document v répondant à une question w posée, soit retrouvé ?
- Qu'a-t-il fallu pour que telle donnée x de telle base y émerge au cours d'une requête ?
- Qu'-a-t-il fallu pour que la forme de l'information trouvée soit adaptée à sa réutilisation ?

Les décisions prises en matière de structure de bases de données, de modélisation, d'enrichissement des métadonnées, de rapprochement de modèles et schémas, procèderaient de cette démarche intellectuelle, permettant de correctement définir les conditions préalables à réunir en amont, visant à prendre en compte les besoins, les contraintes et les processus, menant à la constitution d'une collection de ressources complémentaires et enrichies (référencement, description, localisation, mise en forme des données,...).

Des **axes d'évolution** qui sont donc à considérer pour **valoriser** et **mieux exploiter** les ressources numériques :

- approfondir la **réexploitation** des données : produits, déclinés sous des formats divers, ressources électroniques autonomes ou complémentaires de produits, *on-line* ou *off-line*, etc. ;
- jouer sur la **complémentarité** des types d'objets documentaires gérés dans les bases ;
- et plus globalement, positionner davantage le SID sur le traitement et la gestion des **ressources numériques**, afin d'en élargir l'exploitation : celles **produites en interne** et non gérées dans des bases, et celles **disponibles en externe**, exploitables, et pour certaines déjà repérées et signalées dans des produits documentaires, mais elles aussi non répertoriées en BDD. Cette remarque rejoint celle de la recherche fédérée citée un peu plus haut.

2.2.2 De l'intérêt d'enrichir les métadonnées des ressources numériques internes

Nous l'avons détaillé, les métadonnées, informations structurées, décrivent, expliquent, localisent, et ainsi facilitent la découverte, l'utilisation et la gestion des ressources informationnelles. [22]

Se pencher sur les métadonnées, c'est faire ressortir les nombreux **intérêts** qu'elles présentent, notamment pour : l'extension des applications et l'amélioration de l'interopérabilité ; la diversification des usages des jeux de métadonnées ; la possibilité de structurer plus finement le contenu des documents numériques produits nativement tout en restituant leur sens ; l'exploitation plus précise des contenus, tant des bases que des documents électroniques ; la modélisation d'ensembles de documents (cf. schémas XML) ; l'optimisation de la visibilité de sites Web.

La création de **métadonnées de qualité** a un coût qui s'ajoute à la maintenance des ressources, mais elle y participe grandement. Il apparaît donc intéressant d'automatiser le processus au maximum, dans les différentes étapes correspondant au cycle de vie des ressources. Mais aussi, « manuellement » [18], à la création de la ressource, des éléments descriptifs peuvent être apportés par l'auteur, et il est « souhaitable d'intégrer une vision de documentaliste complémentaire dans le *workflow* » [65]. Pratiquement, il n'est pas aisé de déterminer toutes les métadonnées qui seront utiles à d'autres demain, mais la standardisation des valeurs et pratiques (d'indexation par exemple), ont pour objectif la valorisation à la recherche, la diffusion, et même l'archivage et la migration. Sur ce dernier point, le choix de XML est un élément de réponse au maintien de fichiers de métadonnées pérennes et accessibles, et les métadonnées, objets numériques intégrés aux ressources ou gérés dans des bases séparées, peuvent être appelées à migrer, par exemple lors de l'évolution de standards. Elles doivent donc faire l'objet de **modèles documentés**. [65]

Parlant de documents numériques, les professionnels de l'**infodoc** pour reprendre l'expression de Sylvie Dalbin, peuvent toujours se centrer sur le traitement du document une fois produit et sur les métadonnées descriptives associées : bibliographiques, mais aussi métadonnées représentant le document (contenu technique, structure, informations de gestion…). Mais Sylvie Dalbin [50] met en exergue les avantages importants également à intégrer des « éléments liés en amont aux fonctions de production et de gestion de la production ».

« Pour qu'il puisse y avoir recyclage et réexploitation, la donnée doit être présente et administrée ». Le but visé est de disposer de « **réservoirs de métadonnées**[92] » à partir desquels seuls sont utilisées pour certaines exploitations ou affichages, les métadonnées qui s'avèrent pertinentes, les autres métadonnées restant à tout moment « disponibles » pour d'autres exploitations, d'autres usages.

2.2.3 Les données, leur sens, leur filiation, leur complémentarité : application à Centre INFFO

Les données sont souvent considérées comme des informations ou des connaissances. Il faut donc pouvoir fournir les données dans un **cadre structurel** « qui permette de restituer un sens à un ensemble d'entre elles. Ce qui est important, c'est l'éclairage mutuel, le lien qui existe entre données et structure, et qui s'établit grâce aux métadonnées ». [10]
Que ce soit pour les ressources gérées en plein texte dans des bases (articles journalistiques, fiches juridiques…), ou les documents numériques produits nativement (dont une partie peut résulter de l'exploitation des données de certaines bases, par exemple pour la production de monographie, synthèse documentaire, annuaire…), les éléments associés contenus dans les **métadonnées**, véhiculés en même temps que ces contenus informatifs devront être enrichis. Ainsi que nous l'avons avancé plus haut, certains seront alimentés automatiquement par l'application, d'autres relèveront plus d'une alimentation manuelle notamment pour ce qui concerne l'analyse du contenu et leur qualification à l'aide de mots-clés. Si nous prenons le cas de l'indexation de l'ensemble des « Fiches pratiques de la formation continue » dans la base Juriweb, pour améliorer leur accès sur le site Droit de la formation : nous avons bien là la recherche d'une valorisation de contenus informationnels, cœur de métier du Centre INFFO, pour apporter une réponse plus pertinente à la recherche de ces fiches sur le site Web Droit de la formation <http://www.droit-de-la-formation.fr/> que la simple « recherche libre ».

Pour chaque type de donnée ou de ressource gérée à Centre INFFO, il est nécessaire d'évaluer **l'exploitation** que l'on désire en faire, en lien avec d'autres, complémentaires, le traitement et l'interprétation de chacune étant assurés par l'éclairage qui sera apporté par les métadonnées, sur le contexte de leur création et le sens dans lequel elles ont été élaborées.
Les **trois catégories de métadonnées** (décrites au chapitre 4.2 de la Première partie du mémoire) revêtent chacune leur importance. Si les métadonnées descriptives apparaissent les

---

[92] Une fois constitués ces réservoirs « riches » de métadonnées, une étape supplémentaire consiste aussi à mettre en œuvre des outils à même de les exploiter : moteurs, outils de catégorisation ou classification, outils de visualisation graphique,… et interfaces d'accès.

plus intéressantes pour les exploitations documentaires, les métadonnées administratives sont importantes alors qu'elles assurent la gestion des droits d'usage et d'accès aux ressources.
Quant aux métadonnées structurelles elles intéressent directement la **granularité**. Le système d'information se retrouve à gérer des données à la granularité variable : un ensemble de ressources, une ressource individuelle ou encore une partie d'une ressource. Dans le cas de l'accès à l'information par fragments, les métadonnées structurelles permettent de reconstituer la filiation entre les données elles-mêmes. [10] [22]
Se repencher sur les modèles de données des bases, serait ainsi l'occasion de réfléchir de nouveau au **degré de granularité** voulu, à ce que l'on considérera comme unité documentaire au sens classique du terme. Si l'on prend l'exemple de la base bibliographique : rester au niveau de granularité le plus large dans le cas de la description d'un site Web dans son ensemble, choisir un niveau de granularité plus fin pour la description d'articles composant un dossier dans une revue[93].
À moins qu'ils ne soient déjà déterminés par le standard de métadonnées adopté, les choix en matière d'**unité documentaire**, que l'on pourrait traduire par **unité opérationnelle**[94] dans le cas de ressources en ligne, seraient à effectuer en fonction des **usages** que l'on attend de ces objets numériques, qui ont chacun une unité sémantique. [22]

Au moment de la modélisation, **définir les métadonnées** à renseigner. Il ne faut pas sous-estimer ce qui doit être fait en amont de la production de ressources électroniques [10], ne pas se limiter à reproduire une indexation de type bibliographique. Cette réflexion doit s'intégrer à la production éditoriale elle-même, c'est-à-dire être intégrée dès la conception d'un document ou d'une ressource électronique.

2.2.4 Choisir des Schémas XML et suivre des normes et recommandations

Plusieurs Schémas XML correspondraient à la variété des ressources gérées par le Centre INFFO. Les avantages d'XML ont été longuement détaillés (au chapitre 13 de la Première partie du mémoire). Le suivi de ces recommandations techniques et normes assurerait notamment

---

[93] Sachant qu'il est toujours possible de maintenir les liens qui existent entre des données de niveaux de granularité différents : dans le format UNIMARC le champ « contient » permet de lister les articles de revues contenus dans un dossier ; dans le format Dublin Core qualifié également, l'élément *relation* et ses qualificatifs *IsPartOf* et *HasPartOf*, permettent de décrire les liens.
[94] Au sens de l'accès via les moteurs de recherche qui « institue les objets rendus accessibles en unités opérationnelles, qui s'établissent en général à l'échelle d'un fichier désigné par une URL. » [22]

compatibilité, interopérabilité, échange d'informations entre plates-formes hétérogènes. S'appuyer sur les **normes** et **standards** existants, permettra ainsi d'exploiter au mieux les données gérées pour produire d'autres ressources électroniques, mais aussi inversement de réexploiter les documents numériques produits.

Au-delà, les **normes** présentent « une dimension économique et stratégique ». Elles peuvent apporter une aide précieuse, « non seulement en interne, mais aussi vis-à-vis de l'extérieur, dans les négociations qui peuvent être menées avec une tutelle, un financeur ou « à l'occasion de marchés publics ». [66]

Nous nous sommes dans ce mémoire principalement focalisés sur les normes établies par l'ISO ou l'AFNOR, ainsi que sur des recommandations du W3C. Il faut toutefois noter que d'autres organismes produisent des standards dans des domaines complémentaires, et notamment d'**autres modèles de métadonnées** pour certaines ressources : l'IFLA – *International Federation of Library Associations* <http://www.ifla.org/> et la *Library of Congress* <http://www.loc.gov/standards/> (domaine bibliographique), l'IPTC – *International Press Telecommunications Council* <http://www.iptc.org/> (déjà cité pour NewsML, mais l'IPTC gère d'autres standards notamment pour la photo), l'*IMS Global Learning Consortium* <http://www.imsglobal.org/> (domaine éducation), l'IETF - *Internet Engineering Task Force* <http://www.ietf.org/> (normes techniques de l'Internet). De telles sources pourraient apporter des réponses pour des **besoins et usages futurs**, pour considérer d'autres objets documentaires.

2.3 Des solutions techniques à explorer pour développer le système et les services : propositions et remarques

2.3.1 Adopter la technologie XML : avantages et contraintes

« L'implémentation de métadonnées XML est un choix raisonné pour l'**interopérabilité** et la **valeur ajoutée**, à court et long termes ». [65]

XML assure de pouvoir proposer des vues différentes d'une même source : sur papier, en PDF, au format XHTML. De diffuser les informations produites par Centre INFFO sur différents supports, donc par différents canaux, mais également pour des terminaux variés.

Il garantit l'échange normalisé de contenus entre applications (BDD), ou plus simplement l'interopérabilité [10] (ex : intitulés et adresses d'organismes, références de textes officiels ou de documents sources, etc.).

Et l'usage d'XML n'empêche pas pour autant de conserver des bases « SQL » (MySQL, PostgreSQL). Il joue le rôle de « connecteur », au cœur de plateformes Web[95]. [32]

Néanmoins, malgré les qualités de XML, et des technologies Web en général – ouverture, interopérabilité technique, normalisation – il ne faut pas perdre de vue les difficultés de conception et de mise en œuvre qui sont engendrées lors d'un projet de représentation des métadonnées modélisées XML. Des **contraintes** sont dues principalement aux technologies, aux spécifications normatives, ainsi qu'aux ressources humaines à mobiliser et aux ressources financières à allouer.

Avantages et contraintes de XML sont recensés au chapitre 13 de la Première partie de ce mémoire.

2.3.2 XML et bases de données

La question du stockage des documents XML et le type de bases de données à implanter dépendent de l'usage des documents en question. Elle ne fait pas l'objet de ce mémoire et relève des choix opérés par la Direction du Système Informatique (prenant en compte aussi l'état des évolutions technologiques au moment où ces décisions sont prises). Les auteurs S. Calabretto et T. Accary-Barbier [27] en présentent une introduction technique.
Nous pouvons aussi rapporter ce qu'en dit C. Morel-Pair [65]. L'environnement des bases de données propose deux grands types de produits pour XML :

- Des SGBD compatibles XML qui permettent d'importer des données XML dans les différentes tables et de les manipuler.
- Des bases de données en XML natives[96] qui reposent sur le traitement des documents natifs. Elles sont appropriées pour des « documents complexes et variables en structure et contenus, dans lesquels les éléments ont un lien sémantique et structurel ».

---

[95] Cf. à titre d'illustrations les plateformes Web du CRHST <http://www.crhst.cnrs.fr/> et du CN2SV <http://www.cn2sv.fr/>. [32]
[96] Voir le site de Ronald Bourret « XML and databases » <http://www.rpbourret.com/index.htm>.

2.3.3 Des dispositifs pour la manipulation des données du Web

Le passage au Web 2.0 a renforcé le principe selon lequel les données doivent être séparées de leurs représentations. Les données du Web se divisent en deux grandes catégories : les bases de données (BDD) et les documents XML.

Les BDD sont utilisées pour le fonctionnement des sites Web : la gestion de sites de e-commerce ; la gestion d'applications du Web 2.0 tels que les wikis ; les *webmails*.

Dans le même temps, les documents XML permettent la gestion des flux d'informations (RSS), la description d'interfaces riches (RIA – *Rich Internet Application*) et la définition de pages Web (XHTML). Le nombre de formats de documents XML a grandement évolué.

Tous ces constats amènent à mettre en œuvre à la fois des méthodes et des outils dédiés à la **Manipulation des Données du Web (MDW)** qui peut être divisée en deux domaines d'application :

- la **translation de schémas** qui a pour but l'homogénéisation d'un ensemble de données sources hétérogènes : elle comprend la **translation de données** (ou échange de données) et l'**intégration de données** ;
- la **transformation de schémas**, lorsque la sémantique du/des schémas sources diffère du schéma cible, qui comprend la **création de vues**, et dans certains cas l'**intégration de données** également.

Actuellement, les approches majeures possibles pour cette manipulation peuvent être classées en trois groupes de techniques et langages, présentant chacun des avantages, des inconvénients, mais aussi des complémentarités :

- celles basées sur les **instances**, c'est-à-dire les données elles-mêmes (par ex. à l'aide d'XSLT et XQuery) ;
- celles basées sur les **schémas de données** ;
- et enfin celles dirigées par les **modèles** qui empruntent les principes de l'Ingénierie Dirigée par les Modèles (IDM).

Toutes ces approches ont pour buts de permettre l'**interopérabilité entre des sources** de données (des documents XML ou des BDD), l'**intégration des données**, ainsi que la **création de vues**. Ces « vues » peuvent être appréhendées comme des outils permettant à des utilisateurs de voir des données de différents points de vue. [26]

2.3.4 Intérêts de la finesse de structuration : pour l'exploitation des données et l'optimisation des résultats de recherche

Les structures et les formats des catalogues et des autres bases de données du Centre INFFO sont multiples. La tendance est même à l'enrichissement en métadonnées. Ces outils et leurs structures sont conditionnés par la diversité des usages attendus ou potentiels, des produits et services en sortie, et par la diversité des publics visés et de leurs besoins. Et comme l'on ne peut prévoir tous les usages possibles, s'assurer de pourvoir accéder aux métadonnées existantes dans les contenus numériques ou y ajouter les métadonnées qui pourraient s'avérer pertinentes, garantirait la finesse de description et d'indexation, à même de répondre aux besoins et usages de demain que l'on ne connaît pas encore.

Pour ce qui concerne l'interrogation des contenus par formulaires de recherche, l'étude des formats de données, de leur complémentarité, et l'utilisation à venir de Schémas de la technologie XML, pourrait faciliter l'usage d'**outils capables d'interroger simultanément** cette « diversité de bases et catalogues », permettant ainsi aux « usagers de s'y retrouver ». [3] Chaque type de base ou catalogue peut toujours disposer de sa structure d'interrogation, ayant son intérêt propre, mais les particularités de chaque interface sont surtout précieuses pour l'utilisateur qui en maîtrise la démarche intellectuelle, voire le vocabulaire d'interrogation pour ce qui concerne le catalogue de la bibliothèque. Parallèlement à ces modes de recherche dans chaque base séparément, il serait souhaitable de proposer une **interface d'interrogation unique**, simplifiée, sur plusieurs bases. Les usagers sont habitués à cette modalité d'interrogation sur Internet avec un moteur de recherche usuel. Sans préjuger des « processus complexes et invisibles » générés en arrière-plan pour produire des réponses pertinentes, et sans étudier plus avant les solutions techniques possibles, dont ce mémoire n'est pas l'objet, il est possible d'esquisser deux grandes pistes qui mériteraient une investigation poussée :

- Un **moteur de recherche fédérée** assurant une requête multibases en serait une. Il est d'un grand intérêt de permettre à un moteur de recherche de se reposer sur la fiabilité de la structuration de l'information, qui notamment améliore les classements de résultats par pertinence. Les moteurs ne sont pas seulement dédiés à l'interrogation de documents et informations non structurés, ils exploitent également la structuration des données. [84] [85] [90]

- Un **métacatalogue** ou **entrepôt commun de métadonnées structurées**[97] en serait une autre. Il serait alimenté par des données exportées depuis différentes bases, converties au format XML selon, pour chaque base, un Schéma approprié mais respectant pour chacun le format de base minimal Dublin Core. [3]

2.3.5 Avenir d'XML : le Web sémantique, en cours de développement

Le Web 2.0 fait appel à de nombreux outils sociaux (blogs, wikis, réseaux sociaux), propose des fonctionnalités comme la syndication de contenu (avec les flux RSS), met à disposition des outils collaboratifs, à usage privé ou professionnel. Malgré la variété des usages proposés, cet ensemble de technologies présente des limites, la principale consistant en l'immensité des flots de documents et d'informations générés. [54]

L'évolution qui se profile pour le Web, est le **Web sémantique** ou **Web 3.0** [27], basé sur des **données** et non plus des **documents** comme aujourd'hui. Le développement du Web sémantique s'appuie sur les métadonnées, à la fois pour augmenter la qualité des recherches, mais aussi offrir de **nouveaux services**. [10] Donner une signification aux informations manipulées, tel est son objet. Il existe déjà en partie sur le Web 2.0 [69], notamment avec les informations qui peuvent être marquées (« taguées ») pour mieux être retrouvées ensuite. Des attributs spécifiques HTML sont utilisés (class, rev, rel, title) : les **microformats**[98]. Ces métacontenus permettent aux moteurs de recherche de trouver des informations caractéristiques (cartes de visite, évènements, critiques, coordonnées géographiques) ayant du « sens », lues et comprises tant par les hommes que par les machines.
Sans remettre en cause le langage de description des pages Web, le **Web sémantique** vise à apporter des précisions pour qualifier l'information, par exemple en établissant des liens entre des données et des catégories, que les logiciels sauront exploiter. [54]
Cette notion d'ajout d'information peut être représentée par les **métadonnées**. Mais également par les **annotations**, situées au sein de ressources numériques et écrites au cours d'un processus d'annotation/lecture, par exemple une note critique ou explicative associée à un texte, une marque de lecture portée sur un document textuel ou une image. [14] [27]

---

[97] Illustration par un cas de gestion et de mise à disposition de contenus culturels au moyen d'un portail entrepôt : « Les portails entrepôts de données vont probablement se généraliser », par Laurent Bel, paru dans Archimag, n°220, décembre 2008 - janvier 2009, p.25.
[98] Voir [50] et les sites <http://microformats.org/> <http://fr.wikipedia.org/wiki/Microformat> <http://en.wikipedia.org/wiki/Microformat>.

Ce Web sémantique doit être considéré dans une logique d'utilisation du Web par des machines, et non seulement, comme actuellement encore pour une grande part, par l'homme. Ainsi toute donnée doit être accompagnée d'informations permettant à des machines de l'exploiter. La métadonnée est alors considérée comme « une donnée qui contient la sémantique, c'est-à-dire l'explication de la donnée à laquelle elle réfère et son contexte » [10].

En d'autres termes, le Web sémantique vise à organiser non pas des textes, mais des informations formalisées pour être traitées automatiquement.[99] [22]

Il s'appuie fortement sur des données déjà structurées et caractérisées comme les **référentiels**, les **langages documentaires** ou les **listes d'autorité**. « Quand les données sont fiables et bien structurées elles peuvent être exploitées pour créer de nouveaux outils ». Il est possible par exemple, par un travail d'alignement des langages documentaires, de faire dialoguer entre eux des vocabulaires hétérogènes (modélisés au moyen du modèle **SKOS - *Simple Knowledge Organization System***[100]) pour faciliter l'accès à de multiples collections[101]. [58]

Enfin, le Web sémantique s'appuie également sur d'autres outils et métalangages spécifiques. Ainsi, le langage **OWL – *Web Ontology Language*** permet de définir des domaines de connaissance[102].

Un cadre conceptuel intellectuel de description des ressources, tel **RDF – *Resource Description Framework***, peut aussi prendre le relais des formats XML[103]. [20] [27] [49] [65]

RDF est un modèle, géré par le consortium W3C[104], destiné à décrire de façon formelle les ressources du Web, en intégrant leurs métadonnées, de façon à en permettre le traitement

---

[99] Reprenant la prose de Gérard Puimatto [22], qui explique ces **principes du Web sémantique**, « les documents considérés comme des instances des informations qu'ils supportent, sont générés, traités, échangés par des logiciels qui permettent aux usagers de produire eux-mêmes, sans connaissance particulière, les données, de les agréger, de les publier et de les échanger en fonction de leurs relations sémantiques et règles d'inférence issues de l'intelligence artificielle ».

[100] Le modèle SKOS - *Simple Knowledge Organization System*, construit sur la base du RDF, est utilisé pour représenter les vocabulaires contrôlés et structurés, comme les thésaurus.

[101] Antoine Isaac lors d'une journée d'étude organisée en mars 2008, en fit la démonstration avec le projet STITCH - *Semantic Interoperability To access Cultural Heritage* où les relations sémantiques entre deux vocabulaires contrôlés sont exploitées pour les rendre interopérables et même susceptibles d'offrir un accès multilingue aux collections qu'ils référencent [58] ; et <http://www.bnf.fr/PAGES/infopro/journeespro/no-Afnor2008.htm>.

[102] **OWL** est un langage de définition pour les **ontologies**, définissant des classes et types de propriétés... [27]

[103] L'**initiative RDF** intéresse les producteurs de jeux de métadonnées. Ainsi **DC** ou **LOM** disposent par exemple d'un schéma RDF. [65]

automatisé. « Malgré une diffusion limitée du RDF liée à la complexité de son implémentation, les principes du Web sémantique sont toujours au premier plan des travaux documentaires », car « ce sont ceux qui permettent le mieux d'illustrer les objectifs à viser en matière de métadonnées ». [22] [37] [70]

Avoir une vision clarifiée des formats de ses bases de données, pour améliorer la structuration de ses contenus à l'aide de Schémas XML, et faire converger les métadonnées, c'est ainsi pour le Centre INFFO non seulement se doter de ressources davantage interopérables, mais aussi se préparer au Web sémantique en cours de développement. Et la fonction documentaire devrait tenir une place importante dans ce projet. Les **professionnels de l'Information et Documentation** ont en effet un rôle clé à jouer dans la mise en place du Web sémantique, car ils sont habitués à gérer des données structurées et des référentiels. Ils peuvent aussi jouer le « rôle d'architectes des données, intermédiaires entre l'expert fonctionnel et l'architecte technique, pour modéliser les données dans les systèmes d'information sur le Web ». [59]

2.3.6 Proposer des ressources documentaires sous forme d'archives ouvertes

[22] [65]

Optimisant la description de ses données, le Centre INFFO pourrait non seulement mettre en ligne plus aisément ses contenus développés, mais aussi, autoriser la circulation plus large des descriptions d'une partie de ses ressources, qui décrites au format minimal Dublin Core (non qualifié), pourraient alors être moissonnées au moyen du protocole OAI-PMH. Le Centre, s'il le souhaite, pourrait ainsi participer au mouvement des archives ouvertes qui ouvre l'accès à l'information et à la connaissance, tenant le rôle de fournisseur pour certains contenus qu'il choisirait de partager librement.

2.3.7 Se doter d'outils collectifs ou individuels pour la gestion de l'information non structurée

En complément des outils référentiels que constituent des bases documentaires ou même des bases de connaissance, pour **mieux gérer et exploiter l'information (non structurée)** produite en interne ou d'origine externe, d'**autres outils**, **individuels** – et pour certains plus couramment utilisés dans la sphère privée que professionnelle (forums de discussion, outils d'alerte et de veille…) ou **collectifs** peuvent être mis en œuvre. Parmi tous ces dispositifs

---

[104] La recommandation du W3C relative à RDF est consultable à l'adresse <http://www.w3.org/RDF/>.

complémentaires, en constant développement, on peut citer des : outils de collecte sur le Web, outils d'extraction d'information, moteurs de recherche, outils de filtrage et de classement, portails de veille et de fédération de contenus, etc. [7] [9]

2.3.8 Créer des services (documentaires) à valeur ajoutée

L'intégration de sources hétérogènes, finement décrites et indexées, au sein d'un service d'information peut contribuer à accroître la valeur du service[105]. Pour le formuler en d'autres termes à partir des ressources d'information internes d'une organisation, il est possible de créer une véritable **valeur ajoutée**[106]. [2]

Les passerelles établies par le biais de métadonnées communes (champs minimaux DC, langage documentaire, autorités…), standardisées ou normalisées (nommage, éléments explicitement balisés, modèles respectant des Schémas…), ou encore par l'usage de tables de correspondance entre modèles, facilitent la représentation et l'exploitation des informations.
S'en trouvent ainsi améliorés : l'accès aux informations par les utilisateurs humains ; la recherche par les systèmes de filtrage et d'identification de contenus ; les échanges d'information entre les machines ; l'agrégation et l'intégration de données auparavant dispersées, qui nécessitaient des requêtes séparées (par exemple, un article d'actualité portant sur une expérience, en lien avec une description de ressource pédagogique et une étude disponible sur le Web, répertoriée en base bibliographique sur le thème concerné par l'expérience en question).
Et la réutilisation de contenus en les agrégeant, autorise à ré-éditorialiser, à recomposer ou à construire de nouveaux produits éditoriaux (cas des sites collaboratifs, des éditeurs juridiques, etc.). Cette agrégation, de plus en plus automatisée, « s'appuie sur les données d'indexation thématique, sur les personnes et organisations citées, sur l'appréciation du contenu par les utilisateurs, sur les sources, les droits d'utilisation, les dates de validité, etc. ». [14]

Enfin, face à la profusion et la dispersion des ressources numériques sur le Web, l'usager confronté à « une surcharge informationnelle, une désorientation », est « demandeur de voies

---

[105] Des exemples de services d'information à valeur ajoutée externes dans le domaine de la chimie et des brevets sont cités (page 29) dans l'ouvrage : CHAUMIER Jacques, SUTTER Eric. Documentalistes, ajoutez de la valeur à vos services ! Paris, ADBS éditions, Octobre 2007. 63 p. L'essentiel sur… ISSN 1773-729X. ISBN 978-2-84365-095-6.
[106] Sur le thème de la valorisation de la fonction info-documentaire à travers des prestations à forte valeur ajoutée, lire également l'article d'Odile Quesnel et Eric Sutter «Documentalistes,

d'accès simplifiées, performantes et qualifiées ». Les services d'information documentaire se retrouvent confrontés à une **économie de l'attention**, qui devient centrale dans la stratégie d'une offre de contenus : pour capter l'attention de l'internaute et être retenu comme un de ses points d'accès au Web, « il faut désormais privilégier la simplicité des accès, la mutualisation et l'intégration des ressources ». [2] [4]

L'offre de contenus numériques sur le Web s'accompagne de **services complémentaires**, notamment documentaires, que Ghislaine Chartron [4] qualifie de **services joints** (pour leur grande majorité en accès libre pour les non-abonnés) : moteur de recherche, visualisation de sommaires, visualisation des archives, services de veille et d'alerte sur les nouveautés (profils documentaires, fils RSS, envois de sommaires), services d'actualités, etc.

2.4 Investigation menée sur le système d'information documentaire : problèmes soulevés et ajustements à apporter

2.4.1 Dresser le portrait du SID : mieux le connaître pour le faire évoluer

L'investigation menée, permet de lister quelques points d'attention, afin d'optimiser la gestion, le traitement et l'exploitation des données/ressources, de faciliter le suivi de l'existant pour les personnes qui ont en charge l'administration des bases, a fortiori surtout si elles ne sont pas à l'origine de leur création, et enfin améliorer si besoin l'activité des utilisateurs-producteurs desdites bases. Mais aussi plus globalement, pour le responsable du système d'information pouvoir établir une sorte de cartographie des éléments techniques du système, pour pouvoir mieux repérer et esquisser les améliorations à y apporter au regard des évolutions technologiques, des pratiques des utilisateurs-producteurs, des usages des utilisateurs finaux révélés également, des nouveaux besoins dans l'activité de l'entreprise et des contraintes de l'environnement.

2.4.2 Connaître le workflow

Afin d'optimiser le SID, de limiter les tensions dans le workflow et les ruptures de chaîne dans la production de données, il apparaît nécessaire et utile de schématiser les flux de travail autour de toute nouvelle base, ou d'avoir une vision à jour du processus concernant le traitement des données de toute base. Il nous semble un préalable pour pouvoir apporter des améliorations et

---

ajoutez de la valeur à vos services ! », paru dans Documentaliste – Sciences de l'information,

mieux accompagner les changements dans l'activité. Les flux de travail concernant une base sont, dans le temps, évolutifs et changeants, conditionnent ou sont conditionnés par les flux de travail d'autres systèmes, et par l'environnement et ses autres flux (économique, financier, force de travail).

Dans ce mémoire le schéma esquissé du workflow pour chaque base étudiée, volontairement synthétique, s'est essentiellement attaché aux flux de travail des documentalistes ou « producteurs » d'information. Pour pleinement suivre l'élaboration des données jusqu'à leur sortie du système, il faudrait compléter ce portrait en détaillant les flux de travail d'autres métiers intervenant : informaticiens, maquettistes PAO, webmasters...

Même si cela semble évident à déclarer, il est apparu tout au long de notre investigation des bases et des ressources gérées, que suivre le parcours des flux d'information dans le système était parfois fort complexe, et qu'une « cartographie » du fonctionnement de chaque sous-système aide à comprendre davantage le « tout », c'est-à-dire le Système d'information dans son ensemble, même si cet ensemble est en constante mutation (changement de logiciel, nouvelle technologie, nouvelle base, nouvelles ressources, nouveaux produits ou services en sortie, abandon de certaines activités, etc.).

2.4.3 Avoir des outils de suivi des bases et modèles

L'absence de documents d'information à jour relatifs à la constitution et aux fonctionnalités des bases (cahiers des charges, notices techniques, manuels/guides) et à leur structure (formats documentaires : tables, champs, listes) est parfois apparue, et semble dommageable.

Il nous apparaît fortement nécessaire d'assurer un suivi et de mémoriser les évolutions apportées à la structure de bases ; de classer, organiser et/ou assurer l'accessibilité d'un ensemble de documents techniques relatifs aux BDD ; de consigner ces caractéristiques au moyen par exemple de fiches descriptives de BDD, actualisées. Donc de documenter les bases et les modèles de données... et métadonnées.

2.4.4 Appréhender les bases et leur structure

Des difficultés variées pour interroger ces bases et en connaître la structure sont apparues. Parmi ces raisons diverses : un format propriétaire fermé (ex : LORIS et son « champ long » T_DORIS). De même, nous avons été confronté à une multiplicité de bases MySQL maison, ou plutôt de tables regroupées sous des vocables très génériques assurant à la fois la gestion de données diffusées exclusivement en interne sur Intranet, et la gestion de ressources pour des

---

2008, vol. 45, n°2, p. 14-15. ISSN 0012-4508.

bases diffusées sur le Web. L'interface d'administration de ces tables accessible sur l'Intranet a toutefois permis d'éclairer notre compréhension à la fois des relations entre ces tables, mais aussi du contenu même géré lorsqu'un commentaire était disponible. La difficulté demeure toujours, n'ayant pas participé à la modélisation, d'interpréter les formalismes établis par les tables, et par contrecoup d'avoir une bonne compréhension de la nature des données gérées, voire des concepts représentés et des relations établies. L'enquête menée auprès des utilisateurs permet à cet égard d'apporter son propre éclairage aux choix opérés.

En outre, à Centre INFFO, un très grand nombre de bases MySQL constitutives de contenus gérés au moyen du logiciel de gestion de contenu SPIP sont présentes et gèrent des contenus de sites. Pour certaines d'entre elles, leur nom est peu porteur de sens, elles comprennent un grand nombre de tables, et le tout ne favorise pas leur « reconnaissance ».

Parfois, enfin, il est apparu compliqué ou impossible de pourvoir accéder à la structure de la base : elle est administrée par un prestataire extérieur qui seul peut en décrire l'organisation, ou en connaît les clés d'accès. Par exemple, l'investigation menée sur une base ne peut se contenter du modèle des données proposé dans son cahier des charges technique, même si les informations récoltées auprès de l'utilisateur semblent en corroborer les éléments. Le modèle schématisé d'après les sources récoltées au cours de l'enquête, ne peut être véritablement validé, qu'après confirmation par l'administrateur que le cahier des charges rédigé a bien été respecté, et/ou qu'aucune modification ultérieure n'a été apportée à la structure décrite dans les documents descriptifs de la base. L'investigation menée auprès des utilisateurs d'une base de données permet d'avoir une démonstration de l'interface de saisie et de consultation de la base, à laquelle il accède, selon ses droits, et ses besoins. Mais cette « vue », définie par le concepteur de ladite base, se situe à un niveau externe : elle est destinée à faciliter la lecture, l'usage de la base et la sécurité de l'information, mais ne permet justement pas d'en connaître le fonctionnement interne. Car cette « vue » masque volontairement cette complexité. [6]

2.4.5 Sur la méthode d'analyse et d'enquête

Au-delà des difficultés techniques pour appréhender les applications, il faut noter la variété des apports fournis par l'enquête menée auprès des utilisateurs-producteurs ou administrateurs, certes longue à mener, mais indispensable. Sans prétendre à l'exhaustivité, voici quelques-uns des intérêts que procurent une telle démarche, qui apparaît complémentaire et riche d'enseignements : connaissance des métiers, recoupement d'informations, découvertes de « nouvelles » bases ou nouveaux fichiers, compréhension de la complémentarité des ressources

produites, relations entre les contenus, évaluation de la richesse des informations, compréhension de certains choix opérés/solutions choisies, révélation de dysfonctionnements dans le workflow ou dans l'organisation des tâches, relations entre les acteurs de l'entreprise, pratiques individuelles et rapport aux technologies employées, appréhension des évolutions technologiques par les « utilisateurs standards » contributeurs du SID, « lectures métier ».

2.4.6 Aspects méthodologiques de la conduite de projet

Si les technologies sont à disposition, pour favoriser la structuration de l'information et son exploitation, ou même pour gérer ou valoriser l'information non structurée toujours plus massive (ce point de l'information restant non structurée reste, au-delà de ce mémoire, à approfondir), leur intégration dans le contexte d'une entreprise ne se fait pas sans méthode. L'analyse du besoin et l'analyse du contexte de déploiement des technologies envisagées, doivent être « correctement menées », tout comme la conduite de l'ensemble du projet, qui n'est pas de court terme. Des interactions complexes entre les leviers techniques et les choix organisationnels sont en jeu, qui nécessitent de consacrer des ressources, dont des « ressources humaines ». [7]

2.4.7 Etre attentif au(x) facteur(s) humain(s) : un volet à approfondir

En prolongement de la réflexion précédente, dans ce mémoire nous nous sommes particulièrement focalisés sur les éléments techniques clés d'une amélioration de la structuration des informations, proposant une technologie à adopter, certes complexe et contraignante, mais aux nombreux avantages en termes de traitement et d'exploitation de l'information gérée. Nous avons privilégié une approche centrée sur les métadonnées des ressources, devenant elles-mêmes des objets numériques à part entière, et leurs usages. Nous n'avons en revanche pas approfondi les aspects relatifs aux ressources humaines. L'appréhension des technologies et des changements relatifs à l'organisation, aux processus ; le besoin d'accompagnement et de formation ; la place de l'individu dans le collectif de travail, son rôle dans l'articulation des métiers ; le changement du travail et les évolutions de l'activité en cours ou à venir intimement mêlés (changement technique, nouvelles productions, nouveaux services) et les nouvelles compétences et qualifications requises (pour la mise en œuvre du projet et son suivi). Ce « paramètre » ou « facteur » humain revêt une importance toute particulière lors d'une telle mutation technico-organisationnelle, qui est continue et alimentée de surcroît par de nombreux facteurs environnementaux.

Conclusion

Deux modèles types sont proposés par Alain Garnier pour décrire l'**organisation de l'information**. Le modèle « cathédrale », celui de grandes entreprises, disposant « d'un système d'information où tout est pensé », mais en retour pas très souple, voire rigide, non adapté à un monde sans cesse changeant, qui doit concilier sans cesse des exceptions. Et le modèle « grand bazar », où on laisse chacun apporter sa petite pierre dans un grand bazar qui finit tout de même par s'organiser », qui fonctionne par autorégulation spontanée. Ce modèle correspond aux entreprises de taille plus modeste. Le défi pour les entreprises est de trouver la bonne distance entre ces deux modèles. Pour faire face à la surabondance et à l'hétérogénéité de l'information, l'entreprise peut s'organiser et agir sur différents leviers. En formant les utilisateurs à tous ces enjeux, en mettant en place les outils dont on dispose aujourd'hui (moteurs de recherche, extraction de données…), en utilisant des systèmes d'accès les plus proches possibles des usages des utilisateurs (visualisation graphique…), etc. Mais s'infiltrer durablement dans les pratiques demande beaucoup de temps. [7] [9]
« Vouloir tout structurer débouche sur une mécanisation des processus intellectuels qui ne marche pas. En effet, les processus intellectuels à forte valeur ajoutée sont difficilement modélisables dans leur globalité ». Mais ce à quoi l'on peut procéder en revanche, c'est à « une structuration de micro-traitements bien définis et des automates qui aideront à structurer les documents ». [9]
Ainsi, la structuration de l'information garde toute son importance. Et une réflexion poussée sur les métadonnées associées aux objets documentaires, leur choix, leur enrichissement, leur organisation, apparaît incontournable.

Les métadonnées ont des impacts sur les systèmes d'information documentaire (SID), sur les outils documentaires et informatiques, de production, gestion et recherche. Métadonnées utilisées pour la représentation et l'accès à l'information : il faut « prendre acte des besoins et des pratiques des utilisateurs des SID », au sens des besoins et pratiques du métier concerné, déterminant le type de données voulu, la nature et le support de l'information, la granularité (pour la production d'information, la publication, la recherche documentaire, etc.). [12]

La notion de structuration de document est déjà abordée dans les outils bureautiques : principe des feuilles de style de Word par exemple, ou encore métadonnées inscrites dans la zone « Propriétés » de Word ou d'Acrobat Reader. La technologie XML propose une bien autre dimension. La recommandation XML et ses compléments, parfois appelée « galaxie » ou « nébuleuse » par certains acteurs et spécialistes de cette technologie, énonce des règles strictes de structuration des documents numériques, y compris des contenus, visant à faciliter leur exploitation de façon automatique par des machines. *In fine*, les documents numériques et

les données électroniques circulant sur un réseau, peuvent alors être étiquetées avec des balises sémantiques XML, spécifiques à des métiers, et spécifiques à des usages. La description normalisée de ces « informations étiquetées » en facilite tous les usages automatiques, en autorise l'exploitation et la réutilisation.

L'interopérabilité des données, les échanges entre bases ou systèmes d'information, et la compatibilité des éléments d'information gérés, peuvent être assurés par l'usage d'XML.

XML peut ainsi se retrouver au cœur d'applications internes, à la fois méthode et brique du système d'information. Mais aussi il peut se retrouver au centre de plateformes Web (jouant le rôle de « connecteur »), outil de portails, pour diffuser l'information, assurer la pérennité de données numériques, faire le lien avec les outils de l'édition numérique. [32]

Ou encore, si le format Dublin Core de base (non qualifié) est assuré, XML ouvre la possibilité future de traiter et rendre accessibles des archives dites ouvertes : un réservoir ou entrepôt de ces métadonnées XML donnant accès à des ressources numériques en ligne (au moyen du protocole OAI-PMH). [22] [44] [55]

La proposition de la technologie XML, déjà éprouvée, et de jeux et schémas de métadonnées au regard de cette première approche des bases de production du Centre INFFO et de leurs ressources, se veulent un préalable à une investigation techniquement plus poussée qui resterait à mener selon les décisions qui seront prises pour les orientations à venir du SID.

Il faut mettre l'accent sur la coopération et la **collaboration** indispensables entre différentes **professions** – documentalistes, bibliothécaires, informaticiens, développeurs Web, webmasters, « producteurs » des ressources… - pour organiser la gestion des données[107] et participer pleinement « en avançant » ensemble à la mutation des dispositifs, apporter toute la richesse de sa vision métier aux évolutions du système. Brigitte Guyot[108], à cet égard, évoque que les spécialistes de l'information documentaire, traditionnellement gestionnaires de fonds de documents, se repositionnent notamment « vers la production et l'assemblage de prestations à valeur ajoutée ». Gestionnaires de flux, veilleurs, « spécialistes des méthodes intellectuelles propres aux outils d'information », ils accompagnent et facilitent l'usage des outils, apportent leur conseil à l'organisation de l'information…, et leurs compétences autorisent une complémentarité avec celles des « informaticiens installateurs des systèmes ». Et précisément,

---

[107] A cet égard, le CN2SV – Centre National pour la Numérisation de Sources Visuelles <http://www.cn2sv.fr/> semble une bonne illustration du travail collaboratif nécessaire pour constituer une plateforme de préservation des ressources numériques, ainsi que constituer une activité de veille et d'expertise dans son domaine. [65] et [32]

pour l'objet sur lequel s'est focalisé notre travail dans ce mémoire, une confrontation de la compréhension réciproque des schémas à implanter sera nécessaire entre les informaticiens et les praticiens, gestionnaires de données, parmi lesquels de nombreux représentants de la fonction info-documentaire.

Pour monter notre projet, le suivre et progresser, nous proposons d'emblée d'adopter une vision **systémique** : « utiliser l'objectif pour avancer » [74]. Dans les situations de management, « toute décision est prise dans un contexte de complexité » : multiplicité des intervenants ; êtres vivants pas toujours prévisibles ; éléments contextuels non contrôlables, évolutifs voire inconnus… Et parmi ces derniers l'environnement extérieur, politique, économique, social et bien sûr technologique. Adopter une manière moins linéaire de concevoir les évènements dans le temps, mais de manière plus « circulaire », complexe, « ouvre à davantage de réactivité et d'inventivité ». Et pour nous aider à nous repérer dans cette complexité (de causes et d'effets) confrontés à un « problème », il nous faut avoir une visée, un objectif, et l'utiliser pour avancer.
Le **but visé** – aussi communément appelé **objectif, visée, attente** – doit aider à organiser « les éléments [du système] en interaction dynamique » qui vont concourir à l'atteindre. D'où la « nécessité de construire à partir d'une visée ». « Et c'est aussi la confrontation à ce but qui permet de fonctionner avec toute la rigueur nécessaire ».
Tout au long de ce mémoire, les éléments rapportés ont affiché et détaillé les résultats de notre investigation, et ce faisant, décliné trois éléments complémentaires, de manière à donner matière aux décideurs pour les stratégies à adopter, les politiques à mener, les choix techniques à opérer au regard du système d'information documentaire et de ses bases de données.
Le premier, l'exposé d'un **problème** – des bases de ressources documentaires aux structures hétérogènes, gérées par des applications diverses – en recensant ces bases, en dressant leur « portrait », tant de la forme que du contenu et en étudiant le workflow concernant les données gérées. Le deuxième élément, la formulation d'un **objectif** – viser l'interopérabilité des éléments du système d'information documentaire (SID) pour assurer une meilleure exploitation et, partant, repenser le traitement des objets documentaires disponibles. Et enfin, le troisième élément, la proposition d'une **solution** comme moyen d'atteindre le but défini précédemment en révélant les modèles de données, proposant des schémas de métadonnées et préconisant l'utilisation de la technologie XML. Cette ébauche de solution, composite et complexe, n'embrasse pas d'emblée tous les avantages et inconvénients, notamment parce qu'elle touche à un domaine en forte évolution, et amélioration, continue. De même, tout au long de la

---

[108] Dans son ouvrage : Dynamiques informationnelles dans les organisations. Paris, Lavoisier, 2006, 236 p. Collection finance gestion management. ISBN 2-7462-1294-3.

progression d'un tel projet, « la remise en contextes du changement souhaité » [74] obligera à réviser l'objectif, afin de considérer des paramètres nouveaux, notamment mais pas seulement techniques et informatiques, prendre en compte, intégrer des informations complémentaires ou supplémentaires, qui ne manqueront pas de survenir.

Bibliographie

Les références bibliographiques sont présentées selon les normes de référence :

- Z44-005. décembre 1987. Documentation. Références bibliographiques : contenu, forme et structure.
- NF ISO 690-2 Février 1998 Information et documentation. Références bibliographiques Documents électroniques, documents complets et parties de documents.

Les notices sont classées de façon thématique, puis à l'intérieur de chaque thème, s'il y a mention d'auteur(s), par ordre alphabétique du nom du premier auteur cité (auteurs personnes physiques ou personnes morales). Pour les références sans mention d'auteur, les notices sont classées en tête par ordre alphabétique du titre.

L'ensemble des notices est numéroté en continu. Ces numéros entre crochets correspondent à ceux figurant dans le corps du texte de ce mémoire.

Pour chaque référence de document électronique en ligne sur Internet ou référence de site Web, la date de consultation est précisée. Nous précisons aussi que dans le reste du mémoire, d'autres adresses Internet figurent tout au long des chapitres ainsi que dans des notes de bas de page. Dans ce cas précis les adresses sont valides à la date du 24 novembre 2009.

Plan de la bibliographie

1. **Système d'information documentaire : bases de données et gestion des ressources numériques**
2. **Les métadonnées : enjeux, utilisation, valorisation**
3. **De la norme XML aux standards et schémas de métadonnées XML**
   - **3.1. Norme XML et recommandations complémentaires – Modélisation XML**
   - **3.2. Jeux de métadonnées, standards, Schémas XML développés**
4. **Outils méthodologiques ; ouvrages usuels sur les enquêtes et entretiens ; terminologie**
5. **Pour en savoir plus**

La rubrique « Pour en savoir plus » propose des pistes de lecture complémentaires, reprenant des documents non cités dans ce mémoire. Ces références ne font pas l'objet d'un résumé.

Date d'arrêt des recherches bibliographiques : 10 octobre 2009.

1. Système d'information documentaire : bases de données et gestion des ressources numériques

[1] APIL – Association des Professionnels des Industries de la Langue, APROGED – Association des Professionnels pour l'Economie Numérique, CIGREF – Club Informatique des Grandes Entreprises Françaises. Valorisation de l'information non-structurée : Livre Blanc [en ligne]. Puteaux, APROGED, octobre 2007 [consulté le 12 novembre 2009]. 45 p. <http://www.aproged.org/index.php/Demarrer-telechargement/Publications/6-Livre-Blanc-Valorisation-de-l-information-non-structuree.html>

Ce rapport de trois grandes associations professionnelles du secteur des nouvelles technologies de l'information et de la communication, fait le point sur l'ensemble des technologies nécessaires à la gestion et à la valorisation de l'information non structurée. Une double perspective vise d'une part à sensibiliser les personnes concernées par une présentation globale des approches et des solutions disponibles. D'autre part à proposer des clés pour la compréhension de la gestion de l'information non-structurée et prendre en compte pour l'entreprise les enjeux qui lui sont liés. [Ce rapport fait l'objet d'un article d'un consultant, Gérard Dupoirier, publié par les Techniques de l'Ingénieur en mai 2009, également répertorié dans la présente bibliographie.]

[2] BIZE Céline, PASQUIER Karine. Services d'information documentaire : adaptations, innovations, nouveaux concepts. Documentaliste – Sciences de l'information, 2008, vol. 45, n°4, p. 21-22. ISSN 0012-4508

Compte-rendu d'une journée professionnelle, organisée le 25 septembre 2008, à Lausanne, par le département Information documentaire de la Haute Ecole de Gestion de Genève. Y étaient évoquées les voies que peuvent emprunter les services d'information et de documentation pour évoluer, et notamment produire et animer des « dispositifs informationnels ».

[3] CALENGE Bertrand. Bibliothèques et politiques documentaires à l'heure d'Internet. Paris, Electre-Editions du Cercle de la Librairie, 2008. 264 p. Collection Bibliothèques. ISBN 978-2-7654-0962-5

Conservateur général des Bibliothèques, Bertrand Calenge analyse la situation des bibliothèques dans une société marquée par de profonds changements, en matière de circulation de l'information et en matière de pratiques culturelles, où des usages nouveaux sont générés par les flux électroniques. Les collections des bibliothèques ne sont plus au cœur du savoir, les politiques documentaires perdent leur centralité. La mise en perspective des collections de ressources de bibliothèque au regard des technologies et pratiques du Web, et des documents

numériques, questionne notamment sur les besoins corrélés aux documents comme aux usages ; sur les formats de données ; sur les catalogues, leur couverture, leur structure, les modalités de leur interrogation et leur accessibilité.

[4] CHARTRON Ghislaine. L'offre documentaire numérique : repères et décryptage. Documentaliste – Sciences de l'information, 2008, vol. 45, n°2, p. 28-34. ISSN 0012-4508
L'espace documentaire électronique est façonné par l'économie de l'internet et les pratiques des usagers, souhaitant naviguer sans barrières dans les ressources. L'économie de l'attention devient centrale dans la stratégie d'une offre de contenus numériques. Les métadonnées, leur standardisation, leur transmission automatisée deviennent un enjeu face à la place des outils de recherche sur le Web.

[5] DUPOIRIER Gérard. Valorisation de l'information non-structurée. Techniques de l'Ingénieur, 10 mai 2009, H 7 710. 9 p. ISSN 1953-4957
Ce document souligne certains aspects du Livre blanc « Valorisation de l'information non-structurée » publié en octobre 2007, par trois grandes associations professionnelles du secteur des nouvelles technologies de l'information et de la communication, qui ont réuni leurs expertises : l'APIL – Association des Professionnels des Industries de la Langue, l'APROGED – Association des Professionnels pour l'Economie Numérique et le CIGREF – Club Informatique des Grandes Entreprises Françaises. Sont repris les chapitres qui définissent l'information non structurée, donnent une vision du marché et apportent des éléments de réponse en termes de méthode et d'outils. Trois cas de réalisation sont également présentés, esquissant des pistes de solutions opérationnelles pour des entreprises ou des administrations.

[6] LARROUSSE Nicolas. Création de bases de données. Paris, Pearson Education France, 2006. 193 p. Collection Synthex ISBN 2-7440-7154-4
Cet ouvrage constitue un véritable manuel décrivant toutes les étapes de création d'une base de données, de l'analyse à la réalisation, illustrées d'exemples et ponctuées d'exercices avec corrigés. L'auteur insiste particulièrement sur la modélisation du monde réel au moyen du modèle entité-association, le passage au mode relationnel et la mise en œuvre du système ainsi conçu à l'aide du langage SQL – *Structured Query Language*. Une étude de cas récapitulative permet d'appliquer les notions présentées.

[7] LEBIGRE Loïc. Un panorama des technologies de gestion de l'information non structurée. Documentaliste – Sciences de l'information, mai 2008, vol. 45, n°2, p. 74-75. ISSN 0012-4508

Note de lecture de l'ouvrage d'Alain Garnier « L'information non structurée dans l'entreprise : usages et outils », paru en 2007 (Editions Lavoisier/Hermès Science Publications, Collection management et informatique, 244 p., ISBN 978-2-7462-1605-1). L'information non structurée, de grande diversité dans la forme et le fond, recoupe en bonne part l'information textuelle. Des outils selon les usages, individuels ou collectifs, sont disponibles pour sa maîtrise et sa gestion.

[8] RAÏS Nadia. Note technique : informatique documentaire : techniques de gestion d'une base de données documentaire [en ligne]. [Saint-Cloud], [Médiadix], mars 2006, 25 p. [consulté le 6 septembre 2009] <http://mediadix.u-paris10.fr/du/du_cours05-06/opt_doc/nrsmanuelgestbdd.pdf>
Brochure décrivant la mise en place d'une base de données documentaire (fichiers, format documentaire, maquette), puis son alimentation, et enfin les outils de suivi (fiche technique descriptive de base, procédures de versement dans la base principale, cahiers d'extraction pour produits documentaires, guides d'utilisation et notices techniques, évaluation de la base).

[9] TEXIER Bruno. Alain Garnier : à ce jour, 99 % de l'info est non structurée. Archimag, décembre 2007-janvier 2008, n°210, p. 12-13. ISSN 0769-0975
Interview d'Alain Garnier au sujet de son ouvrage « L'information non structurée dans l'entreprise : usages et outils », paru en 2007 (Editions Lavoisier/Hermès Science Publications, Collection management et informatique, 244 p.). Le défi pour les entreprises est de trouver la bonne distance entre deux modèles pour décrire l'organisation de l'information. Le modèle « cathédrale », celui de grandes entreprises disposant d'un système d'information où tout est pensé, sans souplesse. Le modèle « grand bazar » qui fonctionne par autorégulation spontanée, correspondant aux entreprises de taille plus modeste. Cet auteur propose plusieurs pistes pour faire face à la surabondance et à l'hétérogénéité de l'information.

2. Les métadonnées : enjeux, utilisation, valorisation

[10] BAUDRY DE VAUX Marie, DALBIN Sylvie. Métadonnées et valorisation de l'information : journée d'étude ADBS-INTD. Documentaliste – Sciences de l'Information, avril 2006, vol. 43, n°2, p. 144-147. ISSN 0012-4508
Synthèse d'une journée d'étude organisée conjointement, le 4 avril 2006, par l'ADBS – Association des professionnels de l'information et de la documentation et l'INTD – Institut national des techniques de la documentation. Les nombreux intérêts que présentent les métadonnées et leurs usages y étaient expliqués, à l'appui d'illustrations variées. Gestion du cycle de vie des documents, production et diffusion de l'information sur différents supports, par

différents canaux, manipulation des données pour valoriser l'information et en améliorer l'exploitation, telles étaient les grandes lignes des échanges lors de cette manifestation.

[11] Comité français UNIMARC. Manuel UNIMARC : format bibliographique ; édition française en ligne [en ligne]. Paris, BNF, 19 février 2008. [consulté le 24 octobre 2009]. <http://www.bnf.fr/pages/infopro/normes/no-unimarcb.htm>
Ce document constitue une traduction française de « UNIMARC manual : bibliographic format » réalisé par l'IFLA – International Federation of Library Associations, Fédération internationale des associations de bibliothécaires et des bibliothèques. Il prend en compte le texte de la 2e édition anglaise de 1994 et les 5 mises à jour publiées en anglais de 1996 à 2005.

[12] DALBIN Sylvie. Métadonnées : ADBS [Association des professionnels de l'information et de la documentation] Groupe Sectoriel ATC – Aménagement Transport Construction, réunion du 24 mars 2009 [en ligne]. Paris, ADBS, 2009. [consulté le 7 novembre 2009]. 46 p. <http://www.adbs.fr/servlet/com.univ.collaboratif.utils.LectureFichiergw?CODE_FICHIER=1239983007642&ID_FICHE=6615>
Sylvie Dalbin d'ATD – Assistance et Techniques Documentaires rapporte dans ce compte-rendu ce que les métadonnées peuvent apporter dans la représentation et l'accès à l'information, ainsi que leur positionnement par rapport aux outils documentaires traditionnels. Une illustration est présentée : AdéLIe – Administration de Données locaLIsées <http://adelie.application.equipement.gouv.fr/>, un système de stockage, de catalogage et de mise à disposition du patrimoine géographique du Ministère de l'Ecologie, de l'Energie, du Développement durable et de la Mer (ex Equipement). Sont ainsi évoquées la représentation documentaire métier au moyen de schémas de métadonnées et la représentation informatique par l'usage de la technologie XML ; la modélisation (schémas entités-associations, diagramme UML, vocabulaire RDF) ; les axes d'évolution possibles ; la complexité de la mise en œuvre.

[13] DALBIN Sylvie. Représentation et accès à l'information : transformation à l'œuvre. In CALDERAN Lisette, MILLET Jacques, INRIA – Institut National de Recherche en Informatique et en Automatique. Métadonnées : mutations et perspectives. Séminaire INRIA 29 septembre - 3 octobre 2008 - Dijon. Paris, ADBS éditions, 2008. Chapitre 1, p. 9-58. Sciences et techniques de l'information. ISBN 978-2-84365-104
(Une version *pre-print* de ce chapitre, datée du 17 juillet 2008, est disponible à l'adresse <http://www.atd-doc.com/xmedia/publications/INRIA-2008_DALBIN_Chap01.pdf> [consulté le 24 novembre 2009]).

L'auteur propose une introduction au monde des métadonnées et de leur modélisation (processus de création, administration). Elle insiste sur l'importance de la modélisation des métadonnées, et sur la nécessité pour les professionnels du secteur de l'information-documentation de faire évoluer le regard qu'ils portent sur les métadonnées et les systèmes utilisés pour les représenter (modèles, schémas) ainsi que les ressources numériques et leurs usages.

[14] DELAHOUSSE Jean. Métadonnées : décrire pour mieux servir. Documentaliste - Sciences de l'information, août 2008, vol. 45, n°3, p. 73-74. ISSN 0012-4508
L'auteur présente les métadonnées et les référentiels comme deux concepts clés d'évolution pour le Web. Les travaux de normalisation, au niveau des Schémas de métadonnées, des référentiels partagés, et des technologies du Web sémantique, participent à la création de véritables services à valeur ajoutée, dans lesquels les contenus deviennent partageables, échangeables, réutilisables.

[15] DESCHATELETS Jacinthe. MOREL-PAIR Catherine (collaboration). Dossier sur les métadonnées [en ligne]. AIFBD - Association Internationale Francophone des Bibliothécaires Documentalistes, dernière mise à jour le 12 mai 2009. [consulté le 26 novembre 2009]. <http://bibliodoc.francophonie.org/article.php3?id_article=172>
Définition ; Petit lexique des standards et normes d'écriture ; Pourquoi utiliser les métadonnées ? ; Quelles sont les forces et les faiblesses des métadonnées ? ; Le Dublin Core en bref ; Initiation à quelques formats de métadonnées (EAD, TEI, MARC-XML, MODS, BiblioML, PICS, RDF, P3P Project, RSS) ; les URI - identifiants uniformes de ressources.

[16] DUCOURNAU Josiane. Les Métadonnées [en ligne]. Bordeaux, CRDP Aquitaine, mise à jour septembre 2007 [consulté le 4 novembre 2009]. <http://crdp.ac-bordeaux.fr/documentalistes/formation/metadonnees.asp>
Josiane Ducournau, du CRDP d'Aquitaine – Centre Régional de Documentation Pédagogique, propose dans cet article une courte synthèse sur les métadonnées pour les ressources pédagogiques, avec leur organisation selon le modèle LOM – Learning Object Metadata, dont elle propose un graphe <http://crdp.ac-bordeaux.fr/documentalistes/formation/m%C3%A9tadonnees_fichiers/lomfr.pdf> ainsi qu'un schéma heuristique <http://crdp.ac-bordeaux.fr/documentalistes/formation/m%C3%A9tadonnees_fichiers/LOM.pdf>. Elle liste aussi des outils d'édition de métadonnées (comme Educameta), inclut une bibliographie et un glossaire.

[17] FRESSARD Olivier. Métadonnées et valorisation de l'information. BBF – Bulletin des Bibliothèques de France [en ligne], 2006, n°4. [consulté le 09 octobre 2009], p. 94-95. <http://bbf.enssib.fr/consulter/bbf-2006-04-0094-012> ISSN 0006-2006
Compte-rendu d'une journée organisée le 4 avril 2006 par l'ADBS et l'INTD-CNAM, sur un thème qui constitue une question stratégique pour les bibliothèques. Les métadonnées sont nécessaires à la recherche efficace de l'information pertinente. Exposés sur les défaillances du Web, le projet de Web sémantique et l'enjeu des métadonnées « à l'âge de l'accès ».

[18] LEBIGRE Loïc, MASSE Claudine. Où se cachent les métadonnées ? Documentaliste – Sciences de l'information, novembre 2008, vol. 45, n° 4, p. 20. ISSN 0012-4508
Compte-rendu d'une demi journée d'étude ADBS, organisée à Paris, le 7 octobre 2008, dans le cadre du FAN – Forum des acteurs du numérique, et intitulée : « Où se cachent les métadonnées ? Manuelles, automatiques ; pour l'humain, pour la machine… Comment les maîtriser ? ». Cet article également consultable à l'adresse : <http://www.adbs.fr/compte-rendu-ou-se-cachent-les-metadonnees-manuelles-automatiques-pour-l-humain-pour-la-machine-comment-les-maitriser--55426.htm?RH=ACCUEIL>. Programme : <http://www.adbs.fr/ou-se-cachent-les-metadonnees-manuelles-automatiques-pour-l-humain-pour-la-machine-comment-les-maitriser--43458.htm?RH=ACCUEIL>. Supports des cinq interventions : <http://www.adbs.fr/interventions-ou-se-cachent-les-metadonnees-manuelles-automatiques-pour-l-humain-pour-la-machine-comment-les-maitriser--52516.htm?RH=ACCUEIL>.

[19] MOREAU François. Méta-information et économie numérique. In CALDERAN Lisette, MILLET Jacques, INRIA – Institut National de Recherche en Informatique et en Automatique. Métadonnées : mutations et perspectives. Séminaire INRIA 29 septembre - 3 octobre 2008 - Dijon. Paris, ADBS éditions, 2008. Chapitre 7, p. 223-240. Sciences et techniques de l'information. ISBN 978-2-84365-104
La méta-information devient un élément central de la création de valeur dans les industries de contenu. Face à une offre pléthorique, fournir aux consommateurs une méta-information pertinente est essentiel pour leur permettre à la fois de naviguer dans le foisonnement de contenus et d'en tirer le meilleur avantage.

[20] MOREL-PAIR Catherine. Panorama : des métadonnées pour les ressources électroniques [en ligne]. Octobre 2005, 33 p. [consulté le 6 septembre 2009]. <http://hal.archives-ouvertes.fr/docs/00/04/04/73/PDF/Metas_panorama_CMO.pdf>

Les métadonnées, ensembles de données structurées décrivant des ressources physiques ou numériques, sont un maillon essentiel pour l'interopérabilité de l'information et sa gestion. Cet article propose une synthèse sur le contexte d'apparition des métadonnées, situant les initiatives et leurs acteurs. C. Morel-Pair détaille ensuite le jeu d'éléments Dublin Core, central pour l'interopérabilité, l'intérêt du langage XML pour les métadonnées, puis passe en revue d'autres jeux d'éléments complémentaires et leurs formats d'implémentation, avec des exemples d'applications. Elle insiste sur l'intérêt de RDF pour l'interopérabilité, et les applications qui en sont issues.

[21] NISO – National Information Standards Organization. Understanding Metadata [en ligne]. Bethesda, NISO, 2004. 20 p. [consulté le 6 septembre 2009]. <http://www.niso.org/publications/press/UnderstandingMetadata.pdf> ISBN 1-880124-62-9
Publié en 2004, cette brochure se présente comme un guide pour faire le tour des métadonnées : ce que c'est, ce qu'on en fait, comment les structurer. Il décline tour à tour plusieurs schémas de métadonnées et « jeux d'éléments » : Dublin Core, TEI, METS, MODS, EAD, LOM, <indecs> et ONIX, CDWA et VRA, MPEG, FGDC et DDI. Il aborde ensuite la création de métadonnées, l'interopérabilité et l'échange, les perspectives de développement, et termine sur un glossaire des termes.

[22] PUIMATTO Gérard. Les métadonnées : pourquoi, et pour quoi faire ? [en ligne]. Créé le 3 février 2009, enregistré le 10 mars 2009 [consulté le 3 novembre 2009]. 17 p. <http://savoirscdi2.cndp.fr/fileadmin/fichiers_auteurs/PDF_manuels/Les_metadonnees_Puimatto.pdf>
<http://savoirscdi2.cndp.fr/index.php?id=905>
Gérard Puimatto, directeur adjoint du CRDP de l'académie d'Aix-Marseille (Centre Régional de Documentation Pédagogique) explique l'attention qu'il faut porter aux métadonnées, les intérêts qu'elles représentent pour la pratique documentaire. Après une définition, il en rapporte les usages au regard des ressources électroniques, particulièrement des évolutions du Web dont la préoccupation sémantique est toujours plus prégnante ; et décrit les trois grandes catégories de métadonnées (descriptives, structurelles, administratives) et leur localisation. Ce cadrage amène à s'intéresser aux métadonnées dites « modernes », au-delà du Web, qui offrent au moyen de modèles descriptifs, des standards indispensables pour promouvoir l'interopérabilité par la réutilisabilité d'ensembles de données communes. Les formats MARC conçus pour les échanges de données bibliographiques, se sont ainsi vus enrichis de la démarche Dublin Core (DC) pour aboutir à des formats ouverts comme MARC-XML ou LOM et ses variantes, parmi lesquelles LOM-FR. L'auteur met aussi en valeur un modèle technique de plus en plus utilisé pour l'accès à

des serveurs de données documentaires : le moissonnage des métadonnées DC dans des entrepôts d'archives ouvertes au moyen du protocole OAI-PMH.

3. XML, Schémas XML et autres modèles de métadonnées standards ou normés

3.1 Norme XML et recommandations complémentaires – Modélisation XML

[23] <XML>fr : l'espace XML francophone [en ligne]. [consulté le 29 octobre 2009]. <http://xmlfr.org/>

Site de la communauté francophone utilisatrice de la technologie XML proposant un suivi de l'actualité dans le domaine. La rubrique <http://xmlfr.org/documentations/standards/> propose des traductions françaises de recommandations W3C – World Wide Web Consortium relatives à XML et ses compléments, mais ces traductions n'ont toutefois pas valeur de référence, seule la version anglaise publiée sur le site d'origine faisant fois.

[24] XML.org [en ligne]. [consulté le 29 octobre 2009]. <http://www.xml.org/>

Ce site anglophone promeut l'utilisation des standards ouverts en fournissant des informations éducatives et des ressources collaboratives, ainsi qu'un espace d'échanges. Il est hébergé par le consortium OASIS – Advancing open standards for the global information society <http://www.oasis-open.org/>. Une liste de discussion XML-DEV <http://www.xml.org/xml-dev/> est dédiée aux développeurs.

[25] ACCARY-BARBIER Tiphaine, CALABRETTO Sylvie. XML : syntaxe. Techniques de l'Ingénieur, 10 août 2005, H 3 500, 15 p. ISSN 0245-9612

A l'heure où les échanges d'information via le Web, mais aussi sur les Intranets, connaissent un grand essor, la nécessité de disposer de formats fiables et puissants de documents hypermédias se fait plus pressante. Cet article retrace l'apparition de formats conçus pour l'échange d'information documentaire structurée sur Internet ou en Intranet. XML est ainsi apparu comme une solution de compromis combinant la souplesse de SGML apparu dans les années 80 et la simplicité de HTML. Métalangage qui permet à la fois de définir des formats de documents et de créer des documents respectant ces formats, XML constitue « une solution très satisfaisante » pour l'échange de données. Dans un document XML, la mise en forme des données est totalement séparée des données elles-mêmes. Cet article présente les éléments syntaxiques du

langage XML, du modèle de document DTD, et des outils de base complémentaires Xpath et XSLT qui permettent le traitement et la mise en forme.

[26] BLOUIN Arnaud, BEAUDOUX Olivier, LOISEAU Stéphane. Un tour d'horizon des approches pour la manipulation des données du Web. Document numérique, 2008/1-2, vol. 11, p. 63-83. ISSN 1279-5127. ISBN 2-7462-2318-9
Le passage au Web 2.0 a renforcé le principe selon lequel les données doivent être séparées de leurs représentations. Dans le même temps, le nombre de formats de documents XML a grandement évolué, et les bases de données (BDD) sont aussi utilisées pour le fonctionnement des sites Web. Ces constats amènent à mettre en œuvre à la fois des méthodes et des outils dédiés à la Manipulation des Données du Web (MDW). Cet article, s'appuyant sur de très nombreuses sources bibliographiques, particulièrement en anglais, présente et compare les techniques majeures de MDW, afin d'en faire ressortir avantages, inconvénients, mais aussi complémentarités. Les approches décrites sont classées en trois groupes : celles basées sur les instances (les données elles-mêmes), celles basées sur les schémas de données, et enfin celles dirigées par les modèles qui empruntent les principes de l'Ingénierie Dirigée par les Modèles (IDM). Toutes ces approches ont pour buts de permettre l'interopérabilité entre des sources de données (documents XML et BDD), l'intégration des données, ainsi que la création de vues.

[27] CALABRETTO Sylvie, ACCARY-BARBIER Tiphaine. XML : gestion de contenus Web. Techniques de l'Ingénieur, 10 août 2006, H 3 502, 14 p. ISSN 0245-9612
Avec le développement d'Internet, le volume d'information textuelle stockée de manière électronique va en s'accroissant. Cet article détaille la gestion de ces contenus Web au moyen de XML. Quatre parties sont distinguées. La 1ère synthétise les méthodes existantes pour stocker des documents XML à l'aide de bases de données (BDD) : deux grands types de documents, basés sur le contenu ou basés sur la structure, et deux grands types de BDD (XML natif, ou classique comme les BDD relationnelles). La 2ème partie détaille les langages de requête XML : Lorel, XML-QL, YATL, XSL, XQL et Quilt. La 3ème aborde XQuery, langage de requête XML proposé par le W3C. La 4ème propose une introduction aux éléments fondateurs du Web sémantique : ontologies, métadonnées, annotations et les langages associés (RDF, OWL,...).

[28] CARFANTAN Jean. PHP & MySQL. Paris, Micro Application, 2004. 436 p. Le poche. ISBN 2-7429-3613-0
Description du langage PHP, qui, couplé au système de gestion de base de données MySQL, permet de créer des pages Web dynamiques. L'auteur, analyste-programmeur et webmestre, en

propose la découverte progressive et livre ses conseils de praticien. Le 9ème chapitre intitulé « Parsez-vous XML ? » (p. 333-357) constitue une introduction à la « galaxie XML » : XML, XSLT, Xpath.

[29] COVER Robin. The CoverPages : online resource for markup language technologies [en ligne]. [consulté le 18 octobre 2009]. <http://xml.coverpages.org/>
Ce site Web en anglais propose une collection de ressources en ligne comme support aux technologies de langages balisés de la famille XML ainsi que d'autres standards relatifs à l'information structurée, permettant d'en suivre l'évolution générale. La rubrique <http://xml.coverpages.org/coreStandards.html> décrit les grandes recommandations et standards utilisés dans les langages balisés : XML, SGML, Schémas XML, XSL/XSLT/Xpath, Xlink, Xpointer, XML Query, CSS, SVG. La rubrique <http://xml.coverpages.org/xmlApplications.html> liste de nombreuses applications XML et initiatives ayant développé des DTD ou des Schémas XML.

[30] OpenWebGroup. Openweb.eu.org [en ligne]. [consulté le 29 octobre 2009]. <http://www.openweb.eu.org/>
Le collectif OpenWebGroup propose sur ce site en français une expertise dédiée aux standards du Web – normes et recommandations du W3C – World Wide Web Consortium, des exemples concrets d'utilisation, de nombreux articles, notamment sur les technologies XHTML, RSS, CSS et DOM.

[31] O'REILLY Tim. XML.com : XML From the Inside Out [en ligne]. [consulté le 29 octobre 2009]. <http://www.xml.com/>
Ce site Web en anglais propose un suivi de l'évolution de la technologie XML : développement, ressources et recommandations.

[32] POUYLLAU Stéphane. La démocratisation du XML documentaire : normes, méthodes, outils pour l'informatisation des e-documents [en ligne]. In La démocratisation du XML, conférence ADBS Aquitaine, Pessac , 27 mai 2008 [consulté le 1er octobre 2009] <http://www.stephanepouyllau.org/data/ppt/pouyllau4adbs-27-05-2008.pdf>
<http://www.cn2sv.fr/IMG/pdf_pouyllau4adbs-27-05-2008.pdf>
L'auteur parle de l'utilisation du XML, utilisé sur toutes les plateformes Web du CNRS – Centre National de la Recherche Scientifique et donne les exemples des sites portails du centre de recherche en histoire des sciences et des techniques et CN2SV, centre national de ressources numériques. Une introduction synthétique au XML et ses applications.

Stéphane Pouyllau, responsable technologique de la plateforme technologique du CRHST – Centre de Recherche en Histoire des Sciences et des Techniques <http://www.crhst.cnrs.fr/> et responsable du CN2SV – Centre National pour la Numérisation de Sources Visuelles <http://www.cn2sv.fr/>, témoigne dans ce support d'intervention de la mise en œuvre de la technologie XML dans ces deux centres du CNRS – Centre National de la Recherche Scientifique. XML est au cœur des plateformes Web développées, pour diffuser l'information, traiter et rendre accessibles les archives (OAI-PMH), assurer la pérennité de données numériques (modèle de gestion OAIS), faire le lien avec les outils de l'édition numérique (revues.org, etc.).

[33] ROLE François. Modélisation et manipulation de documents XML. Paris, Hermès science publications, Lavoisier, 2005. 327 p. ISBN 2-7462-0981-0
Les technologies XML empruntent à différents domaines de recherche : langages formels, bases de données, travaux sur l'édition structurée, etc. François Role propose de cerner les grandes caractéristiques de ces techniques foisonnantes, à l'allure de « maquis normatif » très complexe. Deux grandes parties sont distinguées. La première porte sur les différents modes de représentation possibles pour un document XML, d'un simple texte balisé à une instance d'un modèle de données. La seconde présente les principales techniques de manipulation applicables aux documents, et offre ainsi une vue d'ensemble des traitements possibles (navigation, transformation, requêtage...).

[34] ROULEUX Sophie. La DTD et son langage XML : une application pour la lexicographie contemporaine. Revue de didactologie des langues-cultures et de lexiculturologie, 2006/1, n°141, p. 73-94. ISSN 0071-190X. ISBN 2-2520-3546-3
Illustrant son propos par une application à l'édition d'un dictionnaire, l'auteur explicite dans cet article ce qui définit et caractérise une DTD. Il insiste particulièrement sur les aspects fondamentaux que constituent d'une part la recherche de qualité établie entre le producteur, le concepteur et le lecteur du document, et d'autre part du jeu précis des balises à installer pour coder chaque type d'information de manière homogène.

[35] THOMASSON Jean-Jacques, LONJON Antoine. Modélisation XML. Paris, Eyrolles, octobre 2005. 498 p. Architecte logiciel. Plusieurs extraits significatifs de la version PDF de l'ouvrage, publiée en janvier 2006 (ISBN 2-212-11521-0), sont disponibles sur le site de l'éditeur <http://www.eyrolles.com/Informatique/Livre/modelisation-xml-9782212115215>. ISBN 978-2-212-11521-5
Ce manuel s'adresse aux informaticiens, développeurs, qui, pour concevoir des systèmes d'information, doivent recevoir, gérer et traiter des volumes importants de données XML. Il

décrit les étapes de la démarche de modélisation XML et illustre la démarche par des exemples concrets de cas particuliers. Les auteurs expliquent notamment comment passer du modèle conceptuel UML à un schéma XML, détaillant l'utilisation de la notation graphique du diagramme des classes de UML en vue de la production d'un schéma XML.

[36] THOMASSON Jean-Jacques, SALVATORI Olivier (contribution). Schémas XML. Paris, Eyrolles, 2003. 466 p. ISBN 2-212-11195-9

La recommandation XML Schema est le fruit d'une évolution de plus de vingt ans des concepts d'information structurée. L'auteur témoigne de la solidité de cette technologie, qui, si elle est amenée à évoluer dans sa forme, est durablement installée dans la « galaxie XML ». Les chapitres de ce manuel approfondissent les fonctionnalités de XML Schema (éléments XML, attributs, références à des contenus et identificateurs, composition de modèles via les espaces de noms, typage des données...), puis s'intéressent à la conception d'un schéma XML (traduction d'un modèle de données), sa documentation et ses outils de mise en œuvre.

[37] World Wide Web Consortium (W3C) [en ligne]. [consulté le 3 novembre 2009]. <http://www.w3.org/>

Le World Wide Web Consortium, plus communément appelé par son sigle W3C, est un organisme de standardisation à but non-lucratif, ayant pris la forme d'un consortium chargé de promouvoir la compatibilité des technologies du World Wide Web (telles que HTML, XHTML, XML, RDF, CSS...). Le W3C n'émet pas des normes au sens européen ou international, mais des « recommandations » ayant valeur de standards industriels. Une rubrique est dédiée à XML <http://www.w3.org/XML/>. Une autre présente l'ensemble des recommandations relatives à la technologie XML (XML, XML Namespaces, XML Schema, XSLT, etc.) <http://www.w3.org/standards/xml/>.

Certains documents et recommandations du W3C font l'objet d'une traduction en français, bien que seuls les versions originales en anglais fassent foi ou aient valeur de standards :<http://www.w3.org/Consortium/Translation/French><http://www.w3.org/2003/03/Translations/byLanguage?language=fr>

3.2 Jeux de métadonnées, standards, schémas XML développés

[38] BiblioML [en ligne]. Dernière mise à jour de la version française du site : 11 novembre 2006 [consulté le 25 octobre 2009]. <http://90plan.ovh.net/~adnx/biblioml/doku.php>

Le projet BiblioML, lancé en 1999 par le Ministère de la culture et de la communication – Mission de la recherche et de la technologie et réalisé par AJLSM sous la direction de Martin Sévigny, a abouti à une application XML pour des références bibliographiques et des données d'autorités, basée sur les formats bibliographique et d'autorités UNIMARC. Ce site donne des informations sur BiblioML et AuthoritiesML, qui sont des formats basés sur XML pour l'échange d'enregistrements bibliographiques et d'autorités entre applications.

[39] CDM-fr : présentation de l'offre de formation universitaire [en ligne]. Paris, SDTICE - Sous-Direction des Technologies de l'Information et de la Communication pour l'Education. [consulté le 21 novembre 2009]. <http://cdm-fr.fr/>
Le site collaboratif CDMFR permet depuis 2005 d'accompagner la réflexion et la mise en place de CDMFR, norme expérimentale en cours de rédaction décrivant un Schéma XML basé sur le standard européen CDM d'échange de données sur l'offre de formation entre les établissements d'enseignement supérieur.

[40] LHÉO – Langage Harmonisé d'Échange d'informations sur l'Offre de formation [en ligne]. Paris, DGEFP – Délégation Générale à l'Emploi et à la Formation Professionnelle. [consulté le 18 novembre 2009]. <http://lheo.org/>
Le site officiel du référentiel LHÉO. Ce standard se compose de schémas (grammaires XML), d'un glossaire ainsi que de tables de références représentés sous la forme de fichiers XML. La dernière version 1.3.1 date d'avril 2009 et fait l'objet d'un manuel d'utilisation téléchargeable.

[41] LOM-FR : site du profil français du LOM [en ligne]. [consulté le 3 novembre 2009]. <http://www.lom-fr.fr/>
Site collaboratif de l'implémentation de l'initiative LOM-FR, profil d'application français du standard international LOM – Learning Object Metadata. Ce site propose les résultats du travail du groupe AFNOR-SDTICE mettant en place la norme française LOM-FR. Ce site a pour but initial de protéger l'espace de noms et de proposer en ligne les outils de validation pour permettre la rédaction de fiches LOM-FR cohérentes.

[42] Sup-LOMFR : guide d'usage [en ligne]. Ministère de l'Enseignement supérieur et de la Recherche, SDTICE – Sous-direction des technologies de l'information et de la communication pour l'éducation, [consulté le 3 novembre 2009]. <http://www.sup.lomfr.fr/>
Ce guide d'usage est une aide à la compréhension et à l'utilisation de Sup-LOMFR, une spécialisation de LOM-FR adaptée aux besoins et contraintes spécifiques de l'enseignement supérieur. Il est à destination de tous les acteurs impliqués dans la mise en œuvre d'une

politique de référencement et de partage des ressources pédagogiques de l'enseignement supérieur français. Est également proposée une plaquette de présentation <http://www.sup.lomfr.fr/images/4/4e/Plaquette_presentation_suplomfr.pdf>.

[43] BERARD Raymond, GIBERT Julien. Le sudoc dans Google Scholar. BBF – Bulletin des Bibliothèques de France [en ligne], 2008, n°2. [consulté le 25 octobre 2009], p. 64-66. <http://bbf.enssib.fr/consulter/bbf-2008-02-0064-011> ISSN 0006-2006
L'ABES – Agence Bibliographique de l'Enseignement Supérieur a passé un accord avec Google pour l'indexation du SUDOC – Système Universitaire de Documentation par Google Scholar. Le projet initié au début 2006 a abouti à une mise en production publique en avril 2007. Cet article relative succinctement la mise en œuvre technique : l'échange de données s'effectue au format XML après conversion du format bibliographique MARC. Deux schémas présentent la chaîne de traitement des notices, ainsi qu'une même notice dans les deux formats UNIMARC et MARC XML.

[44] BNF – Bibliothèque Nationale de France. Guide d'utilisation du Dublin Core non qualifié à la BnF [en ligne]. Paris, BNF, version 1.0 du 08 août 2006 [consulté le 25 octobre 2009]. 10 p. <http://bibnum.bnf.fr/oai/20060808_GuideOAI.pdf>
Ce document décrit l'utilisation par la Bibliothèque Nationale de France du format de métadonnées descriptives Dublin Core non qualifié (ou Dublin Core « simple »), dans le cadre du protocole OAI-PMH, dans le but d'accroître la visibilité sur Internet des collections numériques.

[45] BOTTIN Michel. BiblioML : une application XML pour les données bibliographiques et d'autorités, compatible avec (UNI)MARC ; Séminaire des documentalistes des Ecoles d'architecture – Paris-La Villette – 3 Juillet 2003 [en ligne]. Juillet 2003 [consulté le 25 octobre 2009]. <http://www.paris-lavillette.archi.fr/docu/biblioml/fr/index.html>
Michel Bottin, Chef de projet Internet au Ministère de la culture et de la communication présente l'application XML BiblioML pour représenter les données bibliographiques et d'autorités initialement saisies au format UNIMARC. Il met en exergue les inconvénients des formats MARC, les avantages et caractéristiques de BiblioML, les utilisations en cours ou en projet.

[46] CHAN Lois Mai, ZENG Marcia Lei. Metadata interoperability and standardization – A study of methodology Part I. Achieving interoperability at the schema level. D-Lib Magazine [en ligne]. June 2006, vol. 12 n°6. [consulté le 26 novembre 2009]. <http://www.dlib.org/dlib/june06/chan/06chan.html>. ISSN 1082-9873

Cet article publié en deux parties propose une analyse des méthodes qui sont utilisées pour atteindre ou améliorer l'interopérabilité parmi les schémas de métadonnées et les applications, de façon à faciliter la conversion et l'échange de métadonnées, et rendre possible les recherches fédérées et le moissonnage de métadonnées dans des domaines croisés. D'un point de vue méthodologique, la mise en œuvre de l'interopérabilité peut être considérée à différents niveaux d'action : au niveau du schéma, au niveau de l'enregistrement et au niveau de l'entrepôt. Cette première partie s'attache à expliquer les différentes situations dans lesquelles les schémas de métadonnées peuvent être créés ou implémentés, que ce soit pour des projets isolés ou dans le cas d'entrepôts intégrés. Il expose également les approches utilisées au niveau du Schéma.

[47] CHAN Lois Mai, ZENG Marcia Lei. Metadata interoperability and standardization – A study of methodology Part II. Achieving interoperability at the record and repository levels. D-Lib Magazine [en ligne]. June 2006, vol. 12, n°6 [consulté le 26 novembre 2009]. <http://www.dlib.org/dlib/june06/zeng/06zeng.html>. ISSN 1082-9873
Cet article publié en deux parties propose une analyse des méthodes qui sont utilisées pour atteindre ou améliorer l'interopérabilité parmi les schémas de métadonnées et les applications, de façon à faciliter la conversion et l'échange de métadonnées, et rendre possible les recherches fédérées et le moissonnage de métadonnées dans des domaines croisés. D'un point de vue méthodologique, la mise en œuvre de l'interopérabilité peut être considérée à différents niveaux d'action : au niveau du schéma, au niveau de l'enregistrement et au niveau de l'entrepôt. Cette seconde partie expose les approches utilisées au niveau de l'enregistrement, et au niveau de l'entrepôt.

[48] CHERHAL Élizabeth. Le Dublin Core (DC) [en ligne]. 2004. 7 p. [consulté le 21 novembre 2009]. <http://www.enssib.fr/bibliotheque-numerique/document-1809>
Article rédigé par l'auteur à la suite de sa communication orale du 16 novembre 2004 à l'ENSSIB – Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques : « L'indexation des ressources pédagogiques ». E. Cherhal décline l'ensemble des éléments de base, et particulièrement explique la notion de Dublin Core qualifié comprenant des éléments supplémentaires et raffinements d'éléments (*element refinements*). Elle évoque également les schémas d'encodage (vocabulaires contrôlés), les formats physiques disponibles pour son expression, l'utilisation du DC par les serveurs OAI.

[49] CLAVEL Thierry. De MARC à XML : Les nouveaux formats bibliographiques. **In** Il y a une vie après MARC : journée d'étude organisée en hommage à Pierre-Yves Duchemin, l'ENSSIB à

Villeurbanne, 19 novembre 2007 [en ligne]. Villeurbanne, ENSSIB, 19 novembre 2007, [consulté le 24 octobre 2009]. <http://www.enssib.fr/bibliotheque-numerique/notice-1333>.
Support d'une intervention portant sur les formats bibliographiques MARC. Après une présentation de l'origine, des caractéristiques, de la structure et de l'évolution de ces formats depuis leur création en 1965, l'auteur recense les besoins actuels des catalogues des bibliothèques, liste les lacunes de MARC puis les formats XML basés sur MARC qui peuvent leur succéder pour y suppléer : MARC XML, MARXChange, BiblioML, MODS. D'autres jeux et schémas de métadonnées, indépendants de MARC, sont complémentaires : Dublin Core (DC), ONIX, EAD, METS. Des protocoles de requête s'appuyant sur XML ont également été développés : SRU et OAI-PMH.

[50] DALBIN Sylvie. Métadonnées et normalisation. In CALDERAN Lisette, MILLET Jacques, INRIA – Institut National de Recherche en Informatique et en Automatique. Métadonnées : mutations et perspectives. Séminaire INRIA 29 septembre - 3 octobre 2008 - Dijon. Paris, ADBS éditions, 2008. Chapitre 4, p. 113-162. Sciences et techniques de l'information. ISBN 978-2-84365-104
(Une version *pre-print* de ce chapitre, datée du 17 juillet 2008, est disponible à l'adresse <http://www.atd-doc.com/xmedia/publications/INRIA-2008-DALBIN_Chp04_.pdf> [consulté le 24 novembre 2009]).
L'auteur propose une cartographie de jeux de métadonnées ou de certains de ces composants selon trois aspects : le document numérique sous des formats variés, la documentation (documents techniques) vue par d'autres secteurs « que celui de l'infodoc » et des normes proposées par des acteurs de na normalisation autre que l'ISO. Des cadres conceptuels pour représenter les données (FRBR, CRM, OAIS). Modèles pour référencer des documents (RDA, MODS, TEF, DC). Des modèles de documents numériques (DAISY, ePub Books, DocBook, DDI, TEI, EAD, IPTC). Systèmes de représentation de concepts et de dictionnaires (norme TMF, projet de norme sur les thésaurus, SKOS). Fonctions de réservoir, transport et pérennisation (OAI-PMH, XMP, METS). Composants transversaux (identifiants, microformats du Web, composants de gestion de droits). Différents schémas complémentaires du secteur de l'éducation (IMS, SCORM, CDM, LOM) interopérables mais utilisables individuellement.

[51] DCMI – The Dublin Core Metadata Initiative [en ligne]. [consulté le 12 novembre 2009]. <http://www.dublincore.org/>
Site officiel du jeu de métadonnées Dublin Core (DC). Il propose notamment, parmi ses nombreuses ressources, un guide d'utilisation du DC

<http://dublincore.org/documents/usageguide/> ainsi que des « *guidelines* » pour l'implémenter en XML <http://dublincore.org/documents/dc-xml-guidelines/>.

[52] DCMI – The Dublin Core Metadata Initiative. Dublin Core Metadata Element Set, Version 1.1 [en ligne]. 14 janvier 2008 [consulté le 30 octobre 2009]. <http://www.dublincore.org/documents/dces/>
Cette page en anglais reproduit la Recommandation de la DCMI, qui est normalisée ISO 15836 :2009, désignant les quinze éléments composant le Dublin Core simple, aussi appelé Dublin Core non qualifié. [Une traduction en français, réalisée par Anne-Marie Vercoustre, de l'INRIA – Institut National de Recherche en Informatique et en Automatique (dernière mise à jour le 26/03/2002) est consultable à l'adresse <http://www-rocq.inria.fr/~vercoust/METADATA/DC-fr.1.1.html>].

[53] DGEFP - Délégation Générale à l'Emploi et la Formation Professionnelle. Normes descriptives et lisibilité de l'offre de formation : guide de correspondance entre les normes MLO, LHÉO et CDM-fr [en ligne]. [juillet 2009]. [consulté le 21 novembre 2009]. 12 p. <http://storage.canalblog.com/92/36/4007/41977863.pdf>
<http://www.crefor.asso.fr/drupal/system/files/2373_guide_de_correspondance_entre_les_normes_descriptives_de_loffre_de_formation.pdf>
Ce document a été élaboré sous le pilotage de la DGEFP par un groupe de travail interministériel. L'élaboration de ce guide répond à des enjeux de partage d'informations sur l'offre de formation en mettant en évidence les correspondances entre deux langages particulièrement développés en France pour décrire l'offre de formation continue : LHÉO et CDM-fr. Il les situe dans le contexte européen de création à moyen terme d'une norme descriptive, MLO. Un tableau comparatif reprend les deux standards français et MLO-AD (une des deux parties envisagées pour MLO : *MLO-Advertising)*.

[54] FAYON David. Du Web 2.0 (outils sociaux) au Web 3.0 (outils sémantiques) : un élément clé du travail collaboratif. Personnel, mai 2009, n°499, p. 80-81. ISSN 0223-5692
Ce court article rappelle la dimension sociale du Web 2.0, ses applications en entreprise (blog, wiki, réseaux sociaux, syndication de contenu, outils collaboratifs) et ses limites. Le Web sémantique ou Web 3.0 en cours d'élaboration, mais qui existe déjà en partie sur le Web 2.0, notamment avec les informations « taguées » et l'utilisation des microformats, permet de donner une signification aux informations manipulées, d'apporter des précisions pour qualifier l'information.

[55] GAUTHIER Mélanie. Le protocole OAI-PMH et les fonctionnalités de recherche : étude de portails du domaine patrimonial [en ligne]. Mémoire de diplôme supérieur, INTD-CNAM, 2007 [consulté le 24 novembre 2009]. 135 p. <http://memsic.ccsd.cnrs.fr/mem_00000629/fr/>
Le protocole OAI-PMH a été adopté par de nombreuses institutions culturelles désireuses de créer des ponts entre leurs collections, et de se rendre plus visibles sur le Web par l'intermédiaire de portails permettant l'accès à leurs ressources disséminées. Toutefois, cette mise en commun dans des index centralisés de métadonnées de provenance et de type divers pose des questions en termes de recherche. Ce mémoire décrit le fonctionnement de cette technologie, identifie les problèmes les plus fréquemment rencontrés et fournit des préconisations pour y remédier.

[56] ILES Nawel, CHIKH Azeddine, MOTHE Josiane, CHOUITI Sidi Mohammed. Un modèle distribué d'entrepôt pédagogique : utilisation de métadonnées LOM et d'annotations sémantiques [en ligne]. Avril 2008, 14 p. [consulté le 15 novembre 2009]. <ftp://ftp.irit.fr/IRIT/SIG/2008_CEMAFORAD_ICMC.pdf> ou <http://edison.u-strasbg.fr/openconf/papers/65.doc>
Afin faciliter la ré-utilisation des objets pédagogiques produits en local dans des universités, cet article propose un modèle d'entrepôt distribué qui repose sur la description par les métadonnées de LOM et l'indexation sémantique des ressources. Selon cette approche chaque université dispose de son propre entrepôt, contenant des objets pédagogiques, leur description en métadonnées et leur description sémantique. Les métadonnées et les descripteurs de chaque entrepôt sont alors utilisés pour capitaliser les connaissances sur les objets distribués et alimenter un méta entrepôt, comparable à un catalogue accessible par tous. Texte d'une communication lors du Colloque Euro Méditerranéen et Africain d'Approfondissement sur la FORmation A Distance (CEMAFORAD <http://cemaforad4.u-strasbg.fr/>) qui s'est déroulé à Strasbourg, les 9, 10 et 11 avril 2008.

[57] JACQUET Christophe. Métadonnées et Dublin Core [en ligne]. Mise à jour : 21/05/2008 [consulté le 25 octobre 2009]. <http://openweb.eu.org/articles/dublin_core/>
Les métadonnées du Dublin Core (DC) permettent de donner des informations à propos de pages Web. Dans cet article, l'auteur présente la forme de ces descriptions, ainsi que la façon de les intégrer dans des pages Web ((X)HTML).

[58] KATTNIG Cécile. Référentiels, données d'autorité, thésaurus, ontologies, taxonomies… Pour en savoir plus ! BBF – Bulletin des Bibliothèques de France [en ligne], 2008, n°5. [consulté le 22

novembre 2009], p. 102-103. <http://bbf.enssib.fr/consulter/bbf-2008-05-0102-013> ISSN 0006-2006
Cet article relate une journée d'étude organisée le 28 mars 2008 par la commission générale AFNOR CG46 Information et documentation et la BNF, consacrée aux référentiels de représentation des contenus, à leur rôle dans l'interopérabilité des systèmes d'information hétérogènes et à leur exploitation dans le Web sémantique. Ce propos a été abordé en trois points : la structuration des données, les usages actuels et l'interopérabilité sémantique. Les interventions sont consultables à l'adresse <http://www.bnf.fr/PAGES/infopro/journeespro/no-Afnor2008.htm>.

[59] MARESSE Yolène, PUYLAURENT Chamila. Métiers et compétences : les professionnels de l'I-D et le Web de demain. Documentaliste – Sciences de l'information, mai 2009, vol. 46, n°2, p. 22-23. ISSN 0012-4508
Sous le titre « Quel Web de demain ? », une journée d'étude proposée le 7 avril 2009 par l'EBD – Ecole des bibliothécaires documentalistes et l'ADBS – Association des professionnels de l'information et de la documentation invitait à réfléchir sur la place et le rôle que le Web 2.0 et ses successeurs réservent aux professionnels de l'Information et Documentation (I-D). Ceux-ci ont un rôle clé à jouer dans la mise en place du Web sémantique, habitués à gérer des données structurées et des référentiels. Ils peuvent aussi jouer le « rôle d'architectes des données, intermédiaires entre l'expert fonctionnel et l'architecte technique, pour modéliser les données dans les systèmes d'information sur le Web ».

[60] MENET Ludovic. Consolidation d'un modèle conceptuel de données de Master Data Management [en ligne]. Mémoire de Master Recherche, IST - Information Scientifique et Technique, Université de Marne-la-Vallée, 2006 [consulté le 15 novembre 2009]. 82 p. <http://memsic.ccsd.cnrs.fr/mem_00000461/fr/>
Ce mémoire expose une solution informatique développée par la Société Orchestra Network : EBX.Platform, basée sur une architecture XML, visant une démarche de Master Data Management dans un système d'information. L'étude porte sur la vérification et la consolidation d'un modèle conceptuel de données d'EBX.Platform, dont la définition est basée sur la technologie XML Schema. Les données sont unifiées dans un référentiel central (entrepôt de données) dont les performances sont mises en valeur. [Contient une bibliographie et un glossaire.]

[61] MILLARA Sylvie, CREFOR – Centre de Ressources pour la Formation de Haute-Normandie. Normes de lisibilité de l'offre de formation [en ligne]. Rouen, CREFOR,

juin 2009 [consulté le 21 novembre 2009]. 42 p. Eclairage, n°13. <http://www.crefor.asso.fr/drupal/system/files/Eclairage_2009-13-7.pdf>
Ce dossier fait le point sur les normes de lisibilité de l'offre de formation au travers de ses outils d'information. Après avoir introduit le concept de normalisation, les enjeux, et la variété des vocables employés – notamment métadonnée, norme, standard – il décline les grands acteurs du secteur, internationaux, européens et français. Enfin, il décrit les normes et standards relatifs à la publication de l'offre de formation (CDM, CDM-FR, LHÉO, LHIC, MLO) et à la description de ressources pédagogiques (LOM, LOM-FR, SCORM).

[62] Ministère de l'Éducation nationale, Ministère de l'Enseignement supérieur et de la Recherche, STSI – Service des technologies et des systèmes d'information, SDTICE – Sous-direction des technologies de l'information et de la communication pour l'éducation. Indexation de ressources : métadonnées, normes et standards ; Les enjeux de la normalisation dans le domaine des TICE [en ligne]. Première version : 2002. Dernière mise à jour 5 octobre 2009. [consulté le 8 octobre 2009]. <http://www.educnet.education.fr/dossier/metadata>
Le site Educnet, dédié aux nouvelles technologies pour l'éducation, propose un dossier traitant de l'indexation de ressources, réalisé par le centre de documentation de la SDTICE. Il précise les concepts et montre l'intérêt des métadonnées, des normes et des standards (dont Dublin Core). Il fait le point sur la normalisation de l'indexation des ressources pédagogiques et du domaine éducatif, notamment LOM-FR et CDM-FR, puis offre un aperçu du développement de l'Open Archives Initiative (OAI) dans le domaine de l'enseignement supérieur et de la recherche. En complément : organismes de normalisation et de standardisation, et sites spécialisés pour approfondir le sujet ; bibliographie ; glossaire.

[63] Ministère de l'Enseignement Supérieur et de la Recherche, Direction générale de l'enseignement supérieur, Mission de l'Information Scientifique et Technique et du Réseau Documentaire. Normes et standards : [Bib – Cadres Information professionnelle / Normalisation] [en ligne]. Mis à jour le 28 juillet 2008. [consulté le 2 novembre 2009]. <http://www.sup.adc.education.fr/bib/f_Norm.htm>
Les Bibliothèques Universitaires listent dans ces pages de nombreux standards et normes en vigueur dans le domaine de la gestion documentaire. Particulièrement voir les chapitres Produire un document numérique et ses métadonnées (Préparation, balisage, métadonnées, formats ; Formats d'encodage…) ; Echanger des données bibliographiques ; Décrire un objet documentaire.

[64] MOREL PAIR Catherine, GAUTIER Patricia. DCMI-FR : Liste de discussion pour les initiatives de métadonnées francophones [en ligne]. INIST – Institut de l'Information Scientifique et Technique, [consulté le 25 octobre 2009]. <http://listserv.inist.fr/sympa/info/dcmi-fr>
DCMI-FR est une tribune électronique dédiée aux initiatives francophones de métadonnées utilisant le Dublin Core. La tribune se veut un forum pour les discussions, les échanges et la distribution de nouvelles susceptibles d'intéresser les membres des communautés, entre autres académiques et gouvernementales.

[65] MOREL-PAIR Catherine. Métadonnées et XML : des standards efficients de l'environnement numérique [en ligne]. Ingénierie des systèmes d'information, 2007, vol. 12, n°2, p. 9-39. [consulté le 11 novembre 2009]. Prépublication. <http://www.enssib.fr/bibliotheque-numerique/notice-1842> ISSN 1633-1311. ISBN 978-2-7462-1913-7
Les métadonnées sont des ensembles de données structurées décrivant des ressources physiques ou bien numériques. Le processus de « curation digitale » repose sur les métadonnées : c'est un ensemble de pratiques visant à maintenir et ajouter de la valeur à un corpus fiable d'information électronique pour les utilisations actuelles et futures (Cf. Chris Rusbridge du DCC – Digital Curation Center <http://www.dcc.ac.uk/> lors d'une présentation du 9 juin 2005 : "The challenge of managing and preserving e-Research" <http://www.apsr.edu.au/documents/rusbridge_NLA_talk.pdf>). L'utilisation de standards, pour les éléments de métadonnées comme pour leur format d'implémentation, est indispensable. Le langage XML est le format d'implémentation privilégié de nombreux standards de métadonnées. C. Morel-Pair propose ainsi un panorama complet : intérêt et fonctions des métadonnées, interopérabilité des ressources, technologie XML, schémas XML développés, applications et initiatives, intérêt de RDF comme outil XML d'une interopérabilité accrue entre applications.

[66] NUTTIN Guillaume. Nouvelles de la guéguerre des normes en infodoc. Archimag, septembre 2009, n°227, p. 26-28. ISSN 0769-0975
Actuellement, la démarche normative dans le domaine de l'information et de la documentation est l'objet de multiples enjeux. Les normes sont décriées au regard de leur coût et des influences dont elles peuvent faire l'objet, elles sont concurrencées par le foisonnement des formats ou standards, notamment ouverts.

[67] PECCATTE Patrick. Métadonnées : une initiation : Dublin Core, IPTC, Exif, RDF, XMP, etc. [en ligne]. Première version : août 2002. Dernière mise à jour : 13 décembre 2007. [consulté le 8 octobre 2009]. <http://peccatte.karefil.com/Software/Metadata.htm>

L'auteur décrit tout du long d'une conséquente page Web le domaine des métadonnées. Son objectif est « d'orienter le lecteur dans le dédale » que constituent concepts, recommandations et initiatives en la matière. Sont abordées plusieurs techniques fondamentales relatives aux métadonnées : Dublin Core (DC), RDF, XMP (eXtensible Metadata Platform). Mais il développe plus particulièrement celles qui sont appliquées aux images (IPTC et IPTC Core, Exif, Geocodage, DIG35, JPEG2000 et JPX) et à la presse (PRISM, NewsML, NITF).

[68] PETERLONGO Marie. Vers une correspondance entre langages maison et Classification décimale de Dewey dans le cadre d'une utilisation de la norme LOM-FR [en ligne]. Master 2 Stratégies documentaires et valorisation de l'information, Université Lyon 3 Jean Moulin - Faculté de Lettres et Civilisations, 2007 [consulté le 21 novembre 2009]. 88 p. <http://memsic.ccsd.cnrs.fr/mem_00000566/fr/>
Mémoire relatant une mission d'amélioration de taxonomies disciplinaires utilisées par des enseignants, en correspondance avec la Classification décimale de Dewey pour l'indexation de ressources pédagogiques, dans le cadre du profil d'application français LOM-FR. Description du LOM, puis exposé de l'utilisation de LOM-FR à l'INRP - Institut National de Recherche Pédagogique et à l'ENS - Ecole Normale Supérieure de Lyon, et réflexion sur l'interopérabilité entre systèmes de descriptions de ressources, notamment entre langages documentaires.

[69] VATANT Bernard. Web sémantique et Web social : un mariage de raison. Documentaliste – Sciences de l'Information, février 2009, vol. 46, n°1, p. 59-60. ISSN 0012-4508
Le courant social et le courant sémantique du Web sont tous deux nés au milieu des années 90. Les technologies du Web social sont aujourd'hui utilisées par presque tous les internautes. Quant aux technologies sémantiques, moins visibles, elles sont déjà implicitement présentes au cœur des applications Web 2.0. Le courant social et le courant sémantique sont en train de converger pour structurer le Web de demain.

[70] VATANT Bernard. Des métadonnées à la description des ressources : les langages du web sémantique. In CALDERAN Lisette, MILLET Jacques, INRIA – Institut National de Recherche en Informatique et en Automatique. Métadonnées : mutations et perspectives. Séminaire INRIA 29 septembre - 3 octobre 2008 - Dijon. Paris, ADBS éditions, 2008. Chapitre 5, p. 163-194. Sciences et techniques de l'information. ISBN 978-2-84365-104
Après avoir défini le concept de ressource, l'auteur présente plusieurs langages spécialisés du Web sémantique de la famille RDF : RDFS, OWL et SKOS. Il esquisse quelques principes de mise en œuvre de ces langages dans un environnement intégré de gestion de métadonnées.

[71] WENZ Romain. L'avenir des catalogues : formats, données, outils, usages [en ligne]. Mémoire d'étude, Diplôme de conservateur de bibliothèque, ENSSIB, 2009 [consulté le 24 octobre 2009]. 107 p. <http://www.enssib.fr/bibliotheque-numerique/document-21205>
L'avenir des catalogues fait actuellement l'objet de nombreuses réflexions. En France, la BNF – Bibliothèque Nationale de France et l'ABES – Agence bibliographique de l'enseignement supérieur notamment mènent des travaux d'analyse et de remise en question dont les conclusions auront une importance pour les bibliothèques françaises. Ce mémoire évoque aussi le développement de nouveaux modèles de structuration de l'information bibliographique (FRBR – *Functional Requirements for Bibliographical Records*), une réflexion sur les règles de catalogage (ISBD consolidé, RDA – *Ressource Description and Access*), une évolution des formats (avec le langage XML), et des évolutions plus fondamentales comme le Web sémantique.

4. Outils méthodologiques ; ouvrages usuels sur les enquêtes et entretiens ; terminologie

→ Sur les méthodes d'enquêtes et entretiens

→ Sur la terminologie et les définitions de termes techniques

[72] ACCART Jean-Philippe, RETHY Marie-Pierre. Le métier de documentaliste. Paris, Editions du Cercle de la librairie, 2008. 463 p. ISBN 978-2-7654-0961-8
Portant sur les multiples aspects du métier de documentaliste dans la société de l'information, cet ouvrage général permet d'appréhender en première approche les notions d'informatique documentaire, de banque de données, de normes, ou encore de formats de données.

[73] ARBORIO Anne-Marie, FOURNIER Pierre, DE SINGLY François. L'observation directe. 2e éd. Refondue. Paris, Armand Colin, 2008. 128 p. Collection L'enquête et ses méthodes n°128. ISBN 978-2-200-35437-4
Guide relatif à l'observation directe de pratiques sociales. Cette méthode permet de compléter les informations recueillies par le seul discours des acteurs (notamment recueillies par entretien), en reconstituant les pratiques par l'observation des situations où elles se développent.

[74] BALTA François, MULLER Jean-Louis. La systémique avec les mots de tous les jours : changer en douceur. 3ème édition. Issy-les-Moulineaux, ESF éditeur, 2009. 153 p. Les guides pratiques de la Cegos. ISSN 1146-5778. ISBN 978-2-7101-1997-5

La systémique est une méthode qui, en contextualisant les problèmes rencontrés, permet « d'adapter nos modes de pensée aux besoins actuels de l'entreprise et d'anticiper ceux de demain pour mieux y répondre ». F. Balta et J.-L. Muller proposent l'exposé de méthodes et pratiques au travers de sept chapitres, afin d'aider à adopter la systémique en entreprise. Le 3ème chapitre s'intitule « utiliser l'objectif pour avancer » et propose de repenser la notion d'objectif, qui est un but visé, sans la confondre avec deux autres : le problème, description de la difficulté rencontrée, et la solution, moyen d'atteindre ce but.

[75] BELLENGER Lionel, COUCHAERE Marie-Josée. Les techniques de questionnement : savoir poser les bonnes questions. 2ème éd. Issy-les-Moulineaux, ESF Editeur, 2002. 151 p. Collection formation permanente. ISBN 2-7101-1607-3
Cet ouvrage propose une étude approfondie du questionnement et de ses techniques. Découpé en quatre exposés, ce guide explore l'acte essentiel de communication que constitue le questionnement, l'art d'interroger, les effets d'influence du questionnement, et enfin l'état d'esprit à adopter lorsque l'on se place en situation de questionner autrui.

[76] BERDUGO Alain, MAHL Robert, JEAN Gérard (direction). Guide du management des systèmes d'information : thèmes et termes essentiels. Paris, Hermès Science Publications, 2002. 622 p. ISBN 2-7462-0524-6
Guide traitant des thèmes et termes couverts par le management des systèmes d'information dans toutes leurs dimensions : métier, technologie et environnement. Reprise de définitions relatives à trois concepts clés dans la 1ère partie du mémoire : système d'information, interopérabilité de systèmes d'information, entrepôt de métadonnées.

[77] BLANCHET Alain, GOTMAN Anne. L'enquête et ses méthodes : l'entretien. Paris, Nathan, 1992. 125 p. Collection 128. ISBN 2-09-190652-2
Cet ouvrage pose les conditions de validité de la méthode, précise le cadre technique de la situation d'interrogation et définit les principes de l'analyse des discours produits. Il est utile pour la préparation, la réalisation (chapitre 3) et l'analyse des entretiens effectués dans le cadre d'une enquête.

[78] BOULOGNE Arlette [et al.], INTD-er. Vocabulaire de la documentation. Paris, ADBS éditions, 2004. 334 p. Sciences et techniques de l'information. ISSN 1762-8288. ISBN 2-84365-071-2
Cet ouvrage propose des définitions du vocabulaire technique spécialisé dans le domaine de l'information-documentation. Le site de l'ADBS – Association des professionnels de l'information

et de la documentation, propose aussi cet ensemble de termes sur son site Web <http://www.adbs.fr/vocabulaire-de-la-documentation-41820.htm> [consulté le 2 novembre 2009].

[79] CACALY Serge, LE COADIC Yves F., POMART Paul-Dominique, SUTTER Eric. Dictionnaire de l'information. 3e éd. Paris, Armand Colin, 2008. 295 p. ISBN 978-2-200-35132-8
Dictionnaire utilisé pour compléter la définition des concepts clés suivants : métadonnées, bases de données, XML.

[80] CACALY Serge, MELOT Michel, POMART Paul-Dominique, SUTTER Eric, LE COADIC Yves F. Dictionnaire encyclopédique de l'information et de la documentation. Paris, Nathan, 1997. 634 p. Collection réf. ISBN 2-09-190528-3
Cet ouvrage rédigé par près de 100 spécialistes, rassemble les principaux éléments de connaissance qui fondent les sciences de l'information et de la documentation. Ouvrage ancien, il permet néanmoins rétrospectivement, de resituer certains vocabulaires et concepts, parmi lesquels : Banque de données (par Bernard Marx, p. 59-62). Interopérabilité (par Yves F. Le Coadic, p. 364-365). Format (par Philippe Martin, p. 227-231).

[81] FROMAN Bernard, GOURDON Christophe. Dictionnaire de la qualité. Saint-Denis-La Plaine, AFNOR, 2003. 224 p. ISBN 2-12-467821-3
Cet ouvrage propose plus de 800 définitions de termes français, des équivalents en anglais, adjoints d'informations normatives commentées. Parmi ces termes : norme, modèle de données.

[82] HARVATOPOULOS Yannis, LIVIAN Yves-Frédéric, SARNIN Philippe. L'art de l'enquête : guide pratique. Paris, Eyrolles, 1989. 137 p. ISBN 978-2-212-04052-4
Dans sa deuxième partie « Construire des outils performants », ce guide méthodologique donne notamment de nombreux conseils relatifs à la conception et la réussite d'un entretien. Il aborde ainsi l'élaboration du guide d'entretien semi-directif (ESD), les conditions d'efficacité de l'ESD puis sa conduite.

5. Pour en savoir plus

[83] BACHIMONT Bruno. Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents. Paris, Hermès science publications, Lavoisier, 2007. 279 p. Science informatique et SHS. ISBN 978-2-7462-1369-2
Une note de lecture de cet ouvrage est proposée par Sylvie Dalbin dans la revue Documentaliste – Sciences de l'information, mai 2008, vol. 45, n°2, p. 69-70.

[84] BERMES Emmanuelle, MAIGNIEN Yannick, NICOLAS Yann, FLEURY Christine. Moteurs et catalogues : recherches en cours ; dossier. Arabesques [en ligne], juillet-août-septembre 2009, n°55, [consulté le 25 novembre 2009], p. 4-11 <http://www.abes.fr/abes/DocumentsWebAbes/abes/arabesques/Arabesques55.pdf> ISSN 1269-0589

[85] BERMES Emmanuelle. Les moteurs de recherche : petit précis de mécanique à l'usage des bibliothèques numériques. BBF - Bulletin des Bibliothèques de France [en ligne], 2007, n°6. [consulté le 25 novembre 2009], p. 5-10. <http://bbf.enssib.fr/consulter/bbf-2007-06-0005-001> ISSN 0006-2006

[86] ERTZSCHEID Olivier. Moteurs de recherche : des enjeux d'aujourd'hui aux moteurs de demain. In CALDERAN Lisette, MILLET Jacques, INRIA – Institut National de Recherche en Informatique et en Automatique. Métadonnées : mutations et perspectives. Séminaire INRIA 29 septembre - 3 octobre 2008 - Dijon. Paris, ADBS éditions, 2008. Chapitre 2, p. 59-88. Sciences et techniques de l'information. ISBN 978-2-84365-104

[87] GANDON Fabien. Le futur du Web à la lecture des recommandations du W3C. In CALDERAN Lisette, MILLET Jacques, INRIA – Institut National de Recherche en Informatique et en Automatique. Métadonnées : mutations et perspectives. Séminaire INRIA 29 septembre - 3 octobre 2008 - Dijon. Paris, ADBS éditions, 2008. Chapitre 8, p. 241-271. Sciences et techniques de l'information. ISBN 978-2-84365-104

[88] GODART Claude, PERRIN Olivier. Les processus métiers : concepts, modèles et systèmes. Paris, Hermès science publications, 2009. 366 p. Traité IC2 – Informatique et systèmes d'information. ISBN 978-2-7462-2300-4

[89] MBARKI Mohamed. Gestion de l'hétérogénéité documentaire : le cas d'un entrepôt de documents multimédia [en ligne]. Thèse de doctorat, Spécialité Informatique, Université Paul Sabatier – Toulouse III, 2008 [consulté le 22 novembre 2009]. 268 p. <http://thesesups.ups-tlse.fr/185/>
Une liste d'écrits de cet auteur est proposée à la page <http://www.irit.fr/-Publications-?code=2564&nom=Mbarki%20Mohamed>.

[90] MESGUICH Véronique, PARENT Guylène. Recherche fédérée à l'Infothèque universitaire Léonard-de-Vinci. In Guide pratique : publications et ressources numériques. Paris, IDP -

Information Documentation Presse, janvier 2009. p. 34-37. Guide pratique Archimag, n°35. ISSN 0769-0975

[91] OCLC - Online Computer Library Center. What librarians and users want : OCLC report [en ligne]. Dublin, OCLC, mars 2009, [consulté le 26 novembre 2009]. 68 p. <http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf>. ISBN 1-55653-411-6
Un synopsis de ce document est également disponible en français : « Catalogues en ligne : les souhaits des utilisateurs et des bibliothécaires : rapport d'OCLC, document synoptique », daté septembre 2009 [consulté le 26 novembre 2009]. 24 p. <http://www.oclc.org/reports/onlinecatalogs/213724cfb_%20Online_Catalogs_Synopsis.pdf> ISBN-10 1-55653-418-3 / ISBN 978-1-55653-418-8.

[92] RICHY Hélène, DESPRÉS Sylvie. Métadonnées, ontologies et documents numériques. Techniques de l'Ingénieur, novembre 2007, H 7 155 v2 - H 7 155, 19 p. + 2 p. ISSN 0245-9612

[93] SEDES Florence. Métadonnées et nouveaux systèmes d'information. Paris, Hermès, Lavoisier, 2007, 162 p. ISBN 2-7462-1913-1

[94] VOLANT Christiane (direction). L'information dans les organisations : dynamique et complexité. Tours, Presses Universitaires François-Rabelais, 2008. 363 p. Collection Perspectives "Villes et Territoires", n°19. ISBN 978-2-86906-239-9

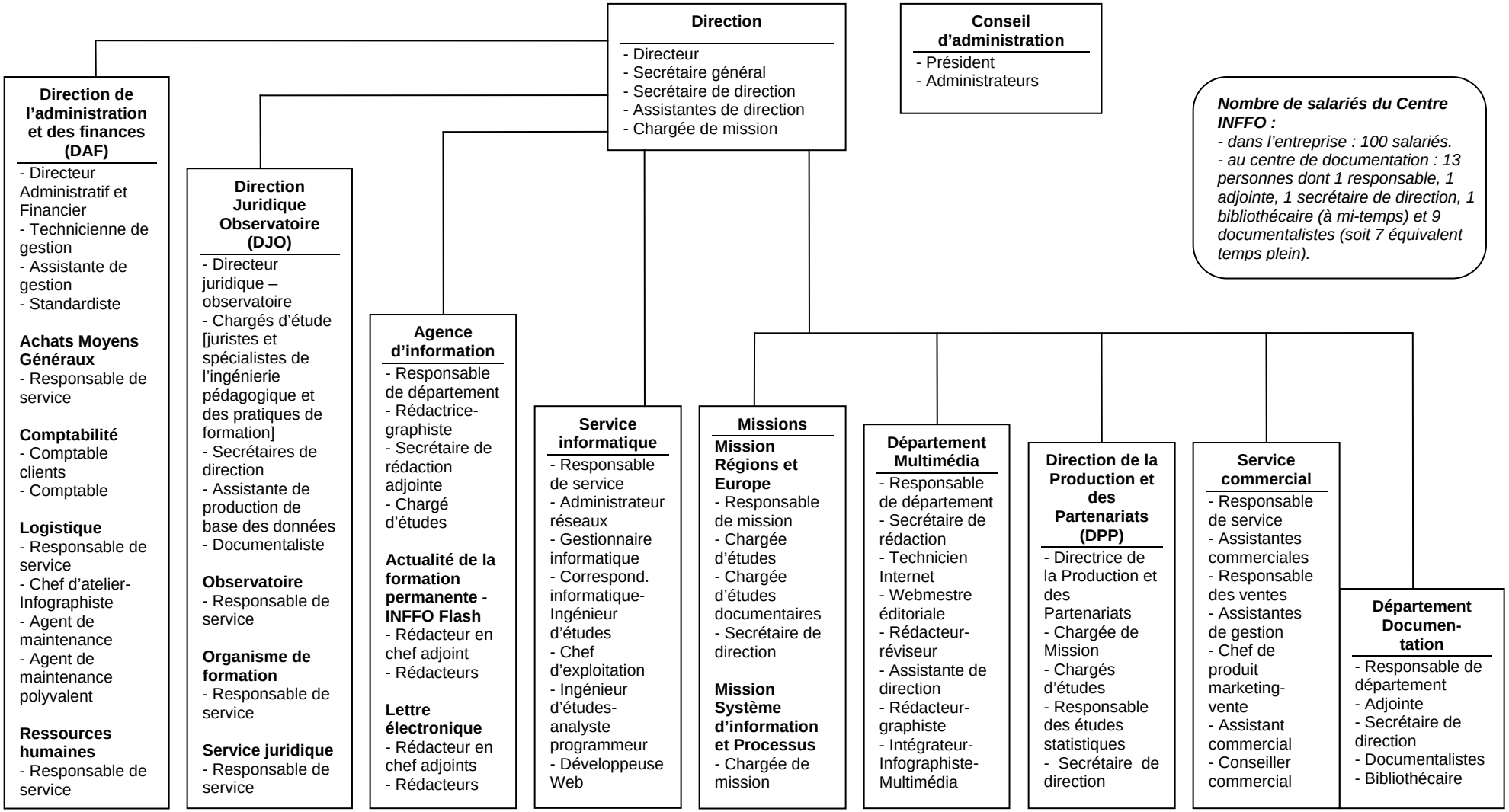
Annexes

Annexe 1 Sigles et acronymes relatifs aux technologies

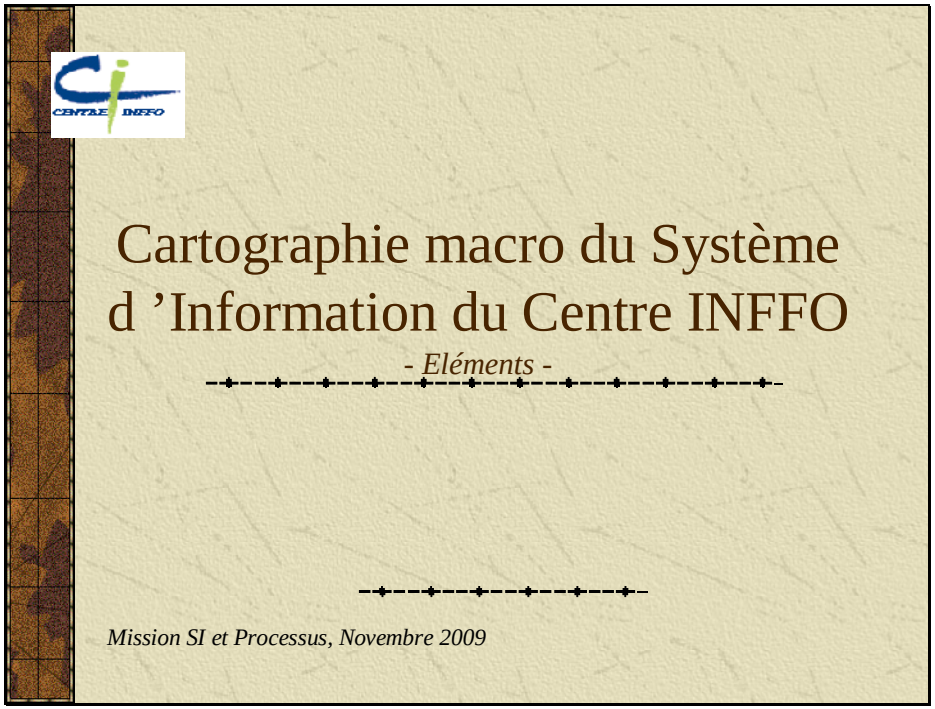
Abréviation	Développé
API	Application Programming Interface
BiblioML	Bibliographic Markup Language
CDM	Course Description Metadata
CSS	Cascading Style Sheets
DC	Dublin Core
DOI	Digital Object Identifier
DOM	Document Object Model
DTD	Document Type Definition
EAD	Encoded Archival Description
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
ISBD	International Standard Bibliographic Description
LHÉO	Langage Harmonisé d'Échange d'informations sur l'Offre de formation
LOM	Learning Object Metadata
MARC	MAchine Readable Cataloging
MCD	Modèle Conceptuel de Données
MDW	Manipulation des Données du Web
MLO	Metadata for Learning Opportunities
MPEG	Moving Pictures Experts Group
METS	Metadata Encoding and Transmission Standard
MODS	Metadata Object Description Schema
OAI-PMH	Open Archives Initiative – Protocol for Metadata Harvesting
ONIX	Online Information eXchange
OWL	Web Ontology Language
PDA	Portable Digital Assistant
RDF	Resources Description Framework
RDFS	RDF Schema

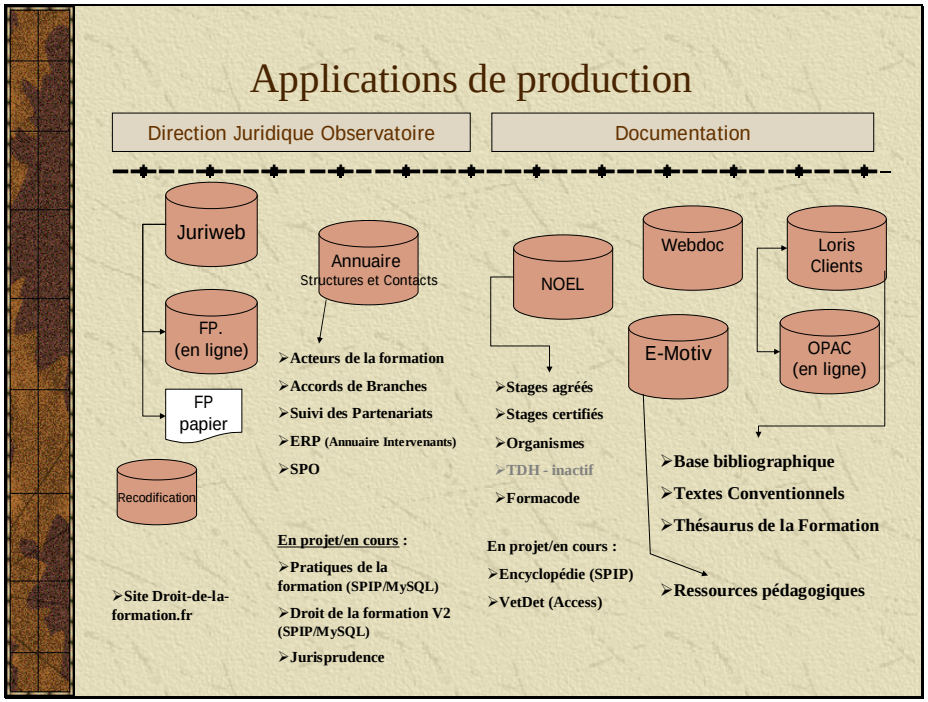
RSS	RDF Site Summary / Rich Site Summary / Really Simple Syndication
SGBD	Système de Gestion de Base de Données
SGML	Standard Generalized Markup Language
SIGB	Système Intégré de Gestion de Bibliothèque
SKOS	Simple Knowledge Organization System
SQL	Structured Query Language
SRU	Search Retrieval via URL
TEF	Thèses Electroniques Françaises
TEI	Text Encoding Initiative
UML	Unified Modeling Language
UNIMARC	UNIversal MARC
URI	Universal Resource Identifier
URL	Uniform Resource Locator
XML	eXtensible Markup Language
XML-QL	Query Language for XML
XPath	XML Path Language
XQL	XML Query Language
XSL	eXtensible Stylesheet Language
XSLT	eXtensible Stylesheet Language Transformation
XTML	eXtensible HyperText Markup Language

Annexe 2 Organisation fonctionnelle du Centre INFFO



Annexe 3 Cartographie macro du Système d'Information du Centre INFFO





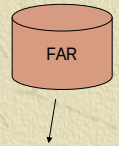
Fonctions de production… et support

Europe/Régions

Pas d'application informatique (à vérifier)

Outil bureautique : Pack Office 2003

- Extranets E-For
- Études
- Affiches…



- Aides des conseils régionaux
- Outils de programmation et de partenariats des Conseils régionaux

Agence

Outil bureautique : Pack Office 2003

PAO :
Photoshop (?)
Illustrator (?)
InDesign
Xpress

- L'Inffo Formation (ex-Inffo-Flash)
- AFP
- Archives Inffo-Flash/L'Inffo

Bases de données à l'aide de SPIP

+ mailing-List avec WAnewsletter

- Le Quotidien de la Formation
- Site Actualité-de-la-formation.fr

Fonctions de production… et support

DPP

Applications informatiques (sur l'intranet ICI)

- Gestion des partenariats
- Suivi des activités (avec DAF) (en cours)
- Gestion des appels d'offre (en projet)

Outil bureautique : Pack Office 2003

- Catalogue

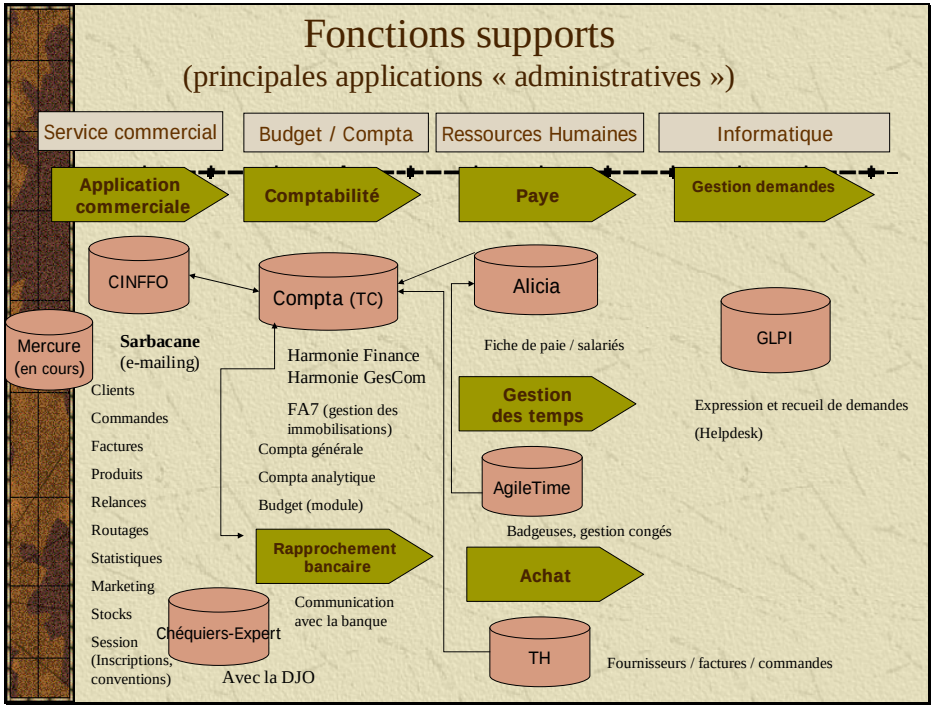
Editions/Multimédia

Fireworks

SPIP

PAO :
Photoshop
Illustrator
InDesign
Xpress

- Sites Internet et extranets (Centre INFFO, Portail Orientation-Formation…)
- POF



Applications collaboratives

Intranet « ICI »

Gestion d'annonces d'information

Rubriques :

. La vie au Centre INFFO / Ressources humaines
. Suivi des activités et produits
. Sources et ressources documentaires
. Espace collaboratif
. Institutions représentatives du personnel (IRP)
. Gestion administrative
. Assistance informatique et bureautique
. Partenaires et conventions
. Boîte à outils

Applications informatiques (certaines figurent déjà sur les pages précédentes)

- **SPO - Suivi des Prestations Orales**

  DJO/DAF/Service Commerciale

- **ERP - Espace de Ressources Pédagogiques)**

  DJO/Service commercial (ouvert en lecture à toutes les équipes)

- **Suivi des partenariats**

  DPP

- **BUCI - Base des Utilisateurs du Centre INFFO, gérant notamment le Trombinoscope**

  Qualité SI et Processus/DAF

- **Suivi des activités**

  DPP/DAF

- **Chéquiers-Experts – Gestion des Questions Juridiques**

  DJO/Service commercial

Outils bureautiques

- **Open-Office**

  (en cours)

- **Pack Office 2003**

Annexe 4 Formats UNIMARC et ISBD : exemples de références extraites de la base LORIS

Trois exemples de références bibliographiques - ouvrage, article de revue, texte conventionnel - extraites de la base de données bibliographiques du Centre INFFO, telles qu'elles apparaissent dans l'interface de saisie LORIS. Elles traduisent un format documentaire adapté, utilisant les champs UNIMARC nécessaires à ses besoins. Elles sont proposées dans les deux formats UNIMARC et ISBD à titre de comparaison.
Les textes apparaissant en rose [nature du texte] ont été ajoutés pour expliciter le contenu de certaines zones UNIMARC.

OUVRAGE

Etiquette UNIMARC

LABEL nam0 22 9 450

005 30/01/2006

010 $a 92-896-0349-6

100 $a d u y0frey0103 ba

101 0 $a Français

102 $a France

200 1 $a Formation et enseignement professionnels $e la clé pour un meilleur avenir - priorités à moyen terme du CEDEFOP pour la période 2006-2008 $f CEDEFOP. Centre européen pour le développement de la formation professionnelle

210 $a Luxembourg $c Office des publications officielles des communautés européennes $d 2005

215 $a 51 p.

330 $a Le présent document définit les objectifs stratégiques et les priorités qui guideront le Conseil de Direction du CEDEFOP dans l'adoption des programmes de travail du Centre pour 2006, 2007 et 2008. Il met en lumière les défis qu'il conviendra de relever dans le domaine de la formation et de l'enseignement professionnels (FEP) pour réaliser les objectifs fixés par le Conseil européen de Lisbonne et exposés dans le programme de travail "Education et formation 2010". Il décrit la contribution que le CEDEFOP est invité à apporter à la réalisation de ces objectifs. La mission du CEDEFOP est de soutenir les politiques de FEP aux niveaux européen, national et sectoriel et d'aider ses partenaires à réaliser leurs

objectifs. A cette fin, le Centre fournit des résultats de recherche, des informations, des connaissances et un appui au développement documenté des politiques.

610 $5 CEDEFOP $5 PROSPECTIVE $5 FORMATION PROFESSIONNELLE CONTINUE $5 ENSEIGNEMENT TECHNIQUE PROFESSIONNEL $5 EDUCATION PERMANENTE $5 EUROPE

710 $a CENTRE EUROPEEN POUR LE DEVELOPPEMENT DE LA FORMATION PROFESSIONNELLE

856 $e 30/01/2006 $u http://www2.trainingvillage.gr/etv/publication/download/panorama/4050_fr.pdf $a http://www.trainingvillage.gr/

[zone 856$e date de la consultation et de l'accès - $u adresse électronique normalisée - $a nom du serveur]

Format ISBD

Formation et enseignement professionnels : la clé pour un meilleur avenir - priorités à moyen terme du CEDEFOP pour la période 2006-2008 / CEDEFOP. Centre européen pour le développement de la formation professionnelle

Luxembourg : Office des publications officielles des communautés européennes, 2005. - 51 p.

Le présent document définit les objectifs stratégiques et les priorités qui guideront le Conseil de Direction du CEDEFOP dans l'adoption des programmes de travail du Centre pour 2006, 2007 et 2008. Il met en lumière les défis qu'il conviendra de relever dans le domaine de la formation et de l'enseignement professionnels (FEP) pour réaliser les objectifs fixés par le Conseil européen de Lisbonne et exposés dans le programme de travail "Education et formation 2010". Il décrit la contribution que le CEDEFOP est invité à apporter à la réalisation de ces objectifs. La mission du CEDEFOP est de soutenir les politiques de FEP aux niveaux européen, national et sectoriel et d'aider ses partenaires à réaliser leurs objectifs. A cette fin, le Centre fournit des résultats de recherche, des informations, des connaissances et un appui au développement documenté des politiques.. - ISBN 92-896-0349-6

CEDEFOP / PROSPECTIVE / FORMATION PROFESSIONNELLE CONTINUE / ENSEIGNEMENT TECHNIQUE PROFESSIONNEL / EDUCATION PERMANENTE / EUROPE

URL: http://www2.trainingvillage.gr/etv/publication/download/panorama/4050_fr.pdf

ARTICLE DE REVUE

Etiquette UNIMARC

LABEL naa0 22 9 450

005 20/05/2005

100 $a d u y0frey0103 ba

101 0 $a Français

102 $a France

200 1 $a Les jeunes qui sortent sans diplôme de l'enseignement supérieur $e parcours de formation et insertion professionnelle $f Gwenaëlle Thomas

210 $a Marseille $c CEREQ. Centre d'études et de recherches sur les qualifications $d 2003 $9 09

330 $a La moitié des 750 000 jeunes qui chaque année débutent leur vie active ont poursuivi des études après le baccalauréat. Mais plus de 90 000 d'entre eux ont quitté l'enseignement supérieur sans y obtenir de diplôme. Sur un marché du travail où le diplôme constitue un signal fort pour les employeurs, ces jeunes s'insèrent dans des conditions moins favorables que leurs homologues diplômés. Certains tirent néanmoins bénéfice de leur poursuite d'études. Tel est le cas de ceux qui après un baccalauréat général ou technologique ont tenté d'obtenir un BTS - Brevet de technicien supérieur (ou un DUT - Diplôme universitaire de technologie). En revanche, poursuivre en filière professionnelle n'apporte pas d'avantages significatifs aux bacheliers professionnels s'ils n'obtiennent pas de diplôme. Mais ce sont les jeunes ayant échoué au DEUG - Diplôme d'études universitaires générales qui semblent les plus pénalisés lors de leurs débuts dans la vie active, tout particulièrement lorsqu'ils sont titulaires d'un baccalauréat technologique ou professionnel.

463 $6 BREF - CENTRE D'ETUDES ET DE RECHERCHES - 1987 - P - 2003 - 01/09/2003 - 200 - Septembre 2003 $3 00250 $t BREF $9 1987 $f CEREQ. Centre d'Etudes et de Recherches sur les Qualifications $v 200, septembre 2003, pp. 1-4 $x 0758-1858

[zone 463 « Niveau de l'unité matérielle » $6 N° périodique - $3 N° notice du périodique- $t Titre de la revue - $9 Date de 1ère publication de la revue - $f 1ère mention de responsabilité - $v N° volume, mois/année pagination - $x ISSN]

610 $5 ENSEIGNEMENT SUPERIEUR $5 SORTIE FORMATION $5 JEUNE $5 INSERTION PROFESSIONNELLE $5 PARCOURS FORMATION $5 ENSEIGNEMENT SUPERIEUR TECHNOLOGIQUE $5 DEUG DIPLOME ETUDES UNIVERSITAIRES GENERALES $5 DUT DIPLOME UNIVERSITAIRE DE TECHNOLOGIE $5 BTS BREVET TECHNICIEN SUPERIEUR $5 EDUCATION RECURRENTE

700 1 $a Thomas Gwenaëlle

856 $e 18 mars 2003, 12h00 $u http://www.cereq.fr/cereq/b200.pdf $a CEREQ http://www.cereq.fr/

Format ISBD

Les jeunes qui sortent sans diplôme de l'enseignement supérieur : parcours de formation et insertion professionnelle / Gwenaëlle Thomas

BREF, n°200, septembre 2003, pp. 1-4

Marseille : CEREQ. Centre d'études et de recherches sur les qualifications, 2003

La moitié des 750 000 jeunes qui chaque année débutent leur vie active ont poursuivi des études après le baccalauréat. Mais plus de 90 000 d'entre eux ont quitté l'enseignement supérieur sans y obtenir de diplôme. Sur un marché du travail où le diplôme constitue un signal fort pour les employeurs, ces jeunes s'insèrent dans des conditions moins favorables que leurs homologues diplômés. Certains tirent néanmoins bénéfice de leur poursuite d'études. Tel est le cas de ceux qui après un baccalauréat général ou technologique ont tenté d'obtenir un BTS - Brevet de technicien supérieur (ou un DUT - Diplôme universitaire de technologie). En revanche, poursuivre en filière professionnelle n'apporte pas d'avantages significatifs aux bacheliers professionnels s'ils n'obtiennent pas de diplôme. Mais ce sont les jeunes ayant échoué au DEUG - Diplôme d'études universitaires générales qui semblent les plus pénalisés lors de leurs débuts dans la vie active, tout particulièrement lorsqu'ils sont titulaires d'un baccalauréat technologique ou professionnel.
ENSEIGNEMENT SUPERIEUR / SORTIE FORMATION / JEUNE / INSERTION PROFESSIONNELLE / PARCOURS FORMATION / ENSEIGNEMENT SUPERIEUR TECHNOLOGIQUE / DEUG DIPLOME ETUDES UNIVERSITAIRES GENERALES / DUT DIPLOME UNIVERSITAIRE DE TECHNOLOGIE / BTS BREVET TECHNICIEN SUPERIEUR / EDUCATION RECURRENTE
URL: http://www.cereq.fr/cereq/b200.pdf.

TEXTE CONVENTIONNEL

Etiquette UNIMARC

LABEL naa0 22 9 450
005 06/01/2006
100 $a d u y0frey0103 ba
101 0 $a Français
200 1 $a Accord du 18 janvier 2005 portant sur le titre XVI sur la formation professionnelle
209 $a Accord [nature du texte]
219 $a 18/01/2005
229 $a 18 janvier 2005
249 $a ASET0550134M
301 $a 656
309 $a Oui [texte en vigueur]
330 $a Les dispositions du titre XVI de la convention collective des commissaires priseurs relative à la formation professionnelle sont redéfinies : revalorisation des taux des contributions, conditions de mise en oeuvre du contrat et de la période de professionnalisation, du droit individuel à la formation, des formations organisées dans le

cadre du plan de formation. Un observatoire prospectif des métiers et des qualifications est créé.

463 $9 1999 $t BULLETIN OFFICIEL DES CONVENTIONS COLLECTIVES $v 05/5, 5 mars 2005, pp. 107-112

610 $5 COMMISSAIRE PRISEUR $5 PARTICIPATION ENTREPRISE $5 OPCA $5 PLAN FORMATION

619 $a CONTRAT PROFESSIONNALISATION $a PERIODE PROFESSIONNALISATION $a DROIT INDIVIDUEL FORMATION $a OBSERVATOIRE METIER QUALIFICATION

629 $a PARTICIPATION ENTREPRISE $a OPCA $a CONTRAT PROFESSIONNALISATION $a PERIODE PROFESSIONNALISATION $a PLAN FORMATION $a DROIT INDIVIDUEL FORMATION $a OBSERVATOIRE METIER QUALIFICATION

719 $a Chambre nationale des commissaires-priseurs - CNCPJ $a Syndicat national des maisons de ventes volontaires - SYMEV $a Fédération des services CFDT - FS CFDT $a Syndicat des personnels des commissaires-priseurs judiciaires et des sociétés de ventes volontaires CFE CGC $a Confédération générale du travail Force Ouvrière - CGT-FO

729 $a Etudes et organismes professionnels des commissaires-priseurs

Format ISBD

Accord du 18 janvier 2005 portant sur le titre XVI sur la formation professionnelle

BULLETIN OFFICIEL DES CONVENTIONS COLLECTIVES, n°05/5, 5 mars 2005, pp. 107-112

Les dispositions du titre XVI de la convention collective des commissaires priseurs relative à la formation professionnelle sont redéfinies : revalorisation des taux des contributions, conditions de mise en oeuvre du contrat et de la période de professionnalisation, du droit individuel à la formation, des formations organisées dans le cadre du plan de formation. Un observatoire prospectif des métiers et des qualifications est créé.

COMMISSAIRE PRISEUR / PARTICIPATION ENTREPRISE / OPCA / PLAN FORMATION

Candidats descripteurs: CONTRAT PROFESSIONNALISATION / PERIODE PROFESSIONNALISATION / DROIT INDIVIDUEL FORMATION / OBSERVATOIRE METIER QUALIFICATION

Thèmes: PARTICIPATION ENTREPRISE / OPCA / CONTRAT PROFESSIONNALISATION / PERIODE PROFESSIONNALISATION / PLAN FORMATION / DROIT INDIVIDUEL FORMATION / OBSERVATOIRE METIER QUALIFICATION

Signataires: Chambre nationale des commissaires-priseurs - CNCPJ / Syndicat national des maisons de ventes volontaires - SYMEV / Fédération des services CFDT - FS CFDT / Syndicat des personnels des commissaires-priseurs judiciaires et des sociétés de ventes volontaires CFE CGC / Confédération générale du travail Force Ouvrière - CGT-FO

Convention collective: Etudes et organismes professionnels des commissaires-priseurs.

Annexe 5 Exemples d'utilisation des données pour la production de dossiers documentaires

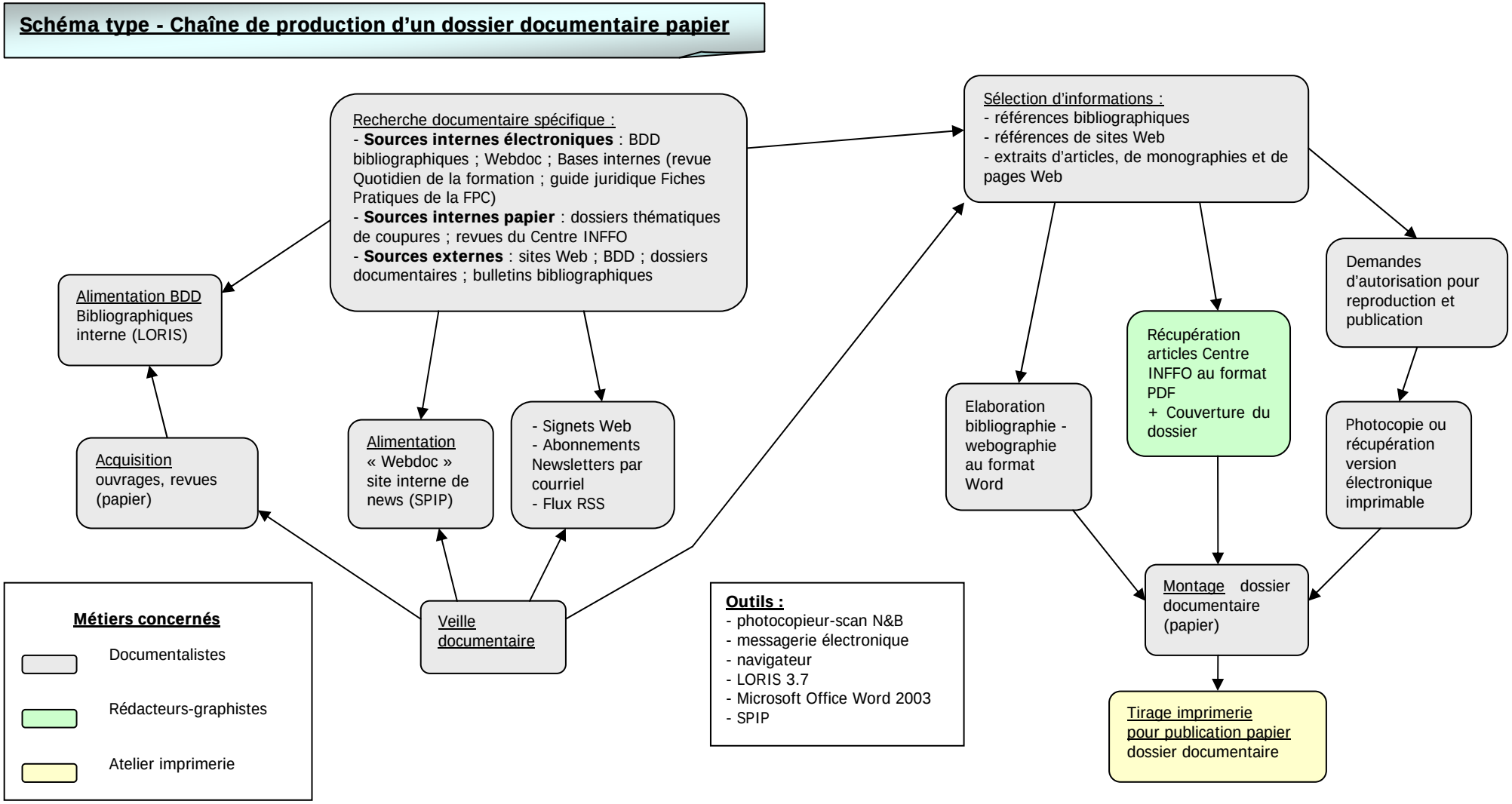
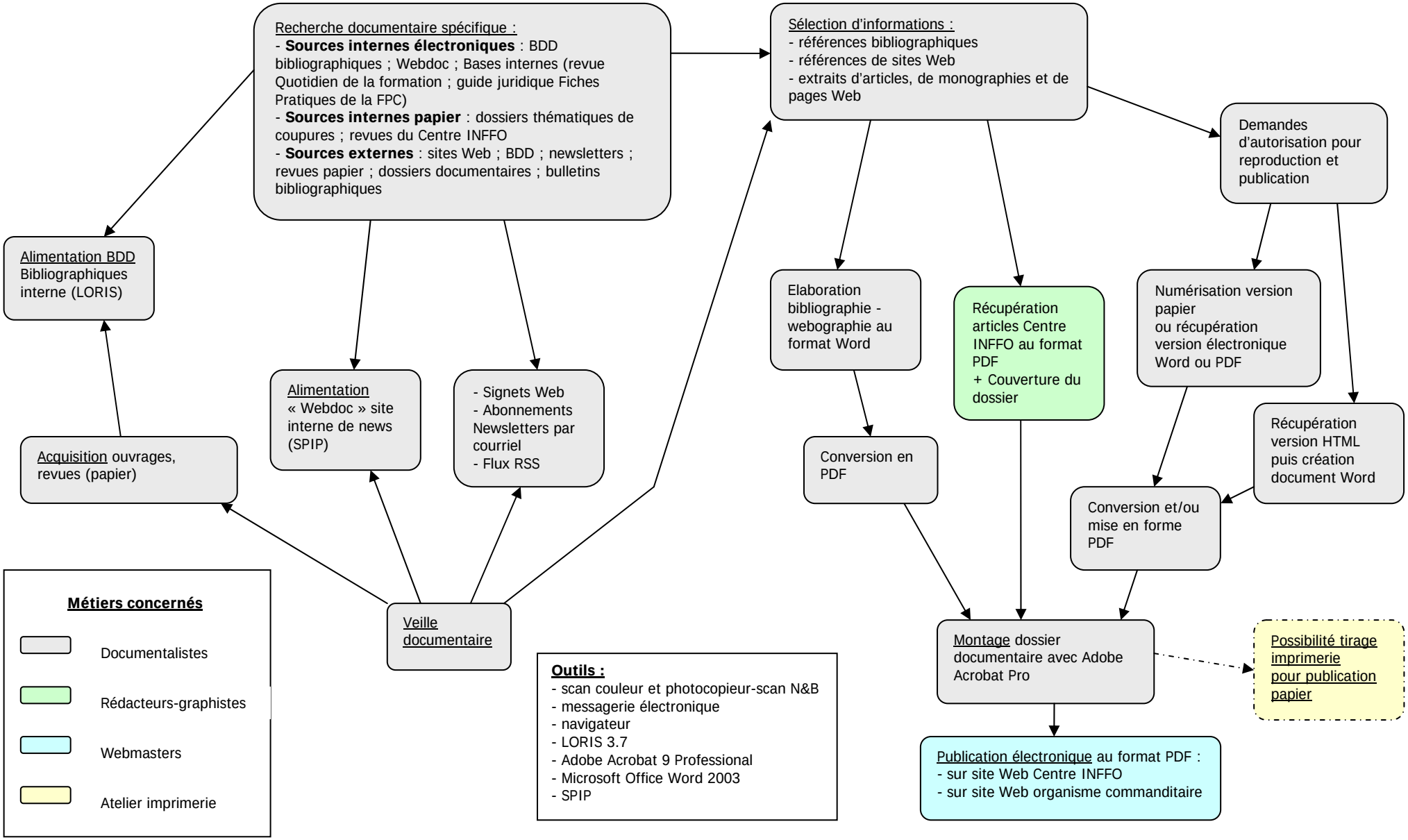


Schéma type - Chaîne de production d'un dossier documentaire électronique



Annexe 6 Site interne Webdoc sous logiciel SPIP

Nom de la base et Type d'objet	Type de base	Service gestionnaire	Utilisation des données Publication sur le Web
Webdoc : rédaction d'actualités sur la formation repérées sur le Web	Base MySQL produite via SPIP	Documentation	• sous logiciel SPIP, en diffusion interne, à l'exception des **Textes officiels** diffusés au jour le jour sur le site **Droit de la formation** <http://www.droit-de-la-formation.fr/spip.php ?page=textesofficiels-rubrique>.

Présentation de Webdoc : constats, réflexions

Un site d'information interne – dénommé « **Webdoc** » - a été mis en place à partir de janvier 2006 par quelques documentalistes, visant à partager, à mutualiser, et à capitaliser toute information reçue de manière électronique, de façon à en faciliter et améliorer l'exploitation.
L'utilisation du système de publication **SPIP** <http://www.spip.net/> prit le relais de répertoires, utilisés comme « boîtes d'information » pour stocker des documents électroniques d'actualité depuis 2004, et partagés sur le réseau informatique interne.

Les informations synthétisées et sourcées, portent sur le domaine de la formation et les thèmes afférents, les publics, les acteurs, les pratiques, les politiques. Les sources sont multiples : courriels, extraits de lettres électroniques en ligne, de sites Web, ou même encore de flux RSS.

Imaginé au départ comme un moyen d'échange d'informations, de pistes, de « signaux faibles », issus notamment d'une veille sur le Web, l'outil **Webdoc** s'est rapidement retrouvé à la charge d'une documentaliste, qui eut pour mission d'effectuer une veille quotidienne spécifique pour alimenter ce produit, et rédiger une note de veille stratégique hebdomadaire, synthétisant chaque semaine les évènements clés, à destination de l'équipe de Direction du Centre INFFO.

Webdoc est produit sous **SPIP** pour permettre une publication immédiate d'informations. Mais leur organisation, le moteur de recherche peu fiable, l'impossibilité d'une indexation riche au moyen d'un langage documentaire (comme les termes issus du « Thésaurus de la formation ») font que cet outil ne constitue pas une véritable Base de données. Son exploitation en est limitée à un rôle d'alerte.
De plus, il n'est pas compatible avec les autres bases documentaires sous LORIS ou développées en interne. Ainsi, particulièrement, **Webdoc** et la **base de données bibliographiques sous LORIS**, contiennent des informations en doublons ou bien des informations qui ne se recoupent pas (rapports, études, articles publiés), nécessitant des requêtes séparées.

Enfin, Webdoc n'est pas articulé à un autre produit d'information interne : « **Les Brèves** ». Ces dernières, quotidiennes, à usage exclusivement interne, sont produites depuis une date beaucoup plus ancienne par le Département documentation. Il s'agit d'un produit constitué d'une revue de la presse quotidienne nationale « papier » et de sommaires de revues et magazines spécialisés reçus au centre de documentation dans un format « papier ». « Les Brèves » font l'objet d'une numérisation au format PDF (image statique) mais en aucune manière elles ne sont « reliées » à « Webdoc ». L'Intranet propose depuis sa page d'accueil un flux RSS menant à chacun des deux produits, mais seul le flux provenant de Webdoc autorise un affichage dynamique.

Constat : il apparaît aujourd'hui que pour mieux exploiter les actualités, plus de collaborations pour l'enrichissement du Webdoc, un meilleur outillage collectif/individuel en amont pour la veille documentaire, et un dépouillement de périodiques ou lettres davantage au format électronique, seraient

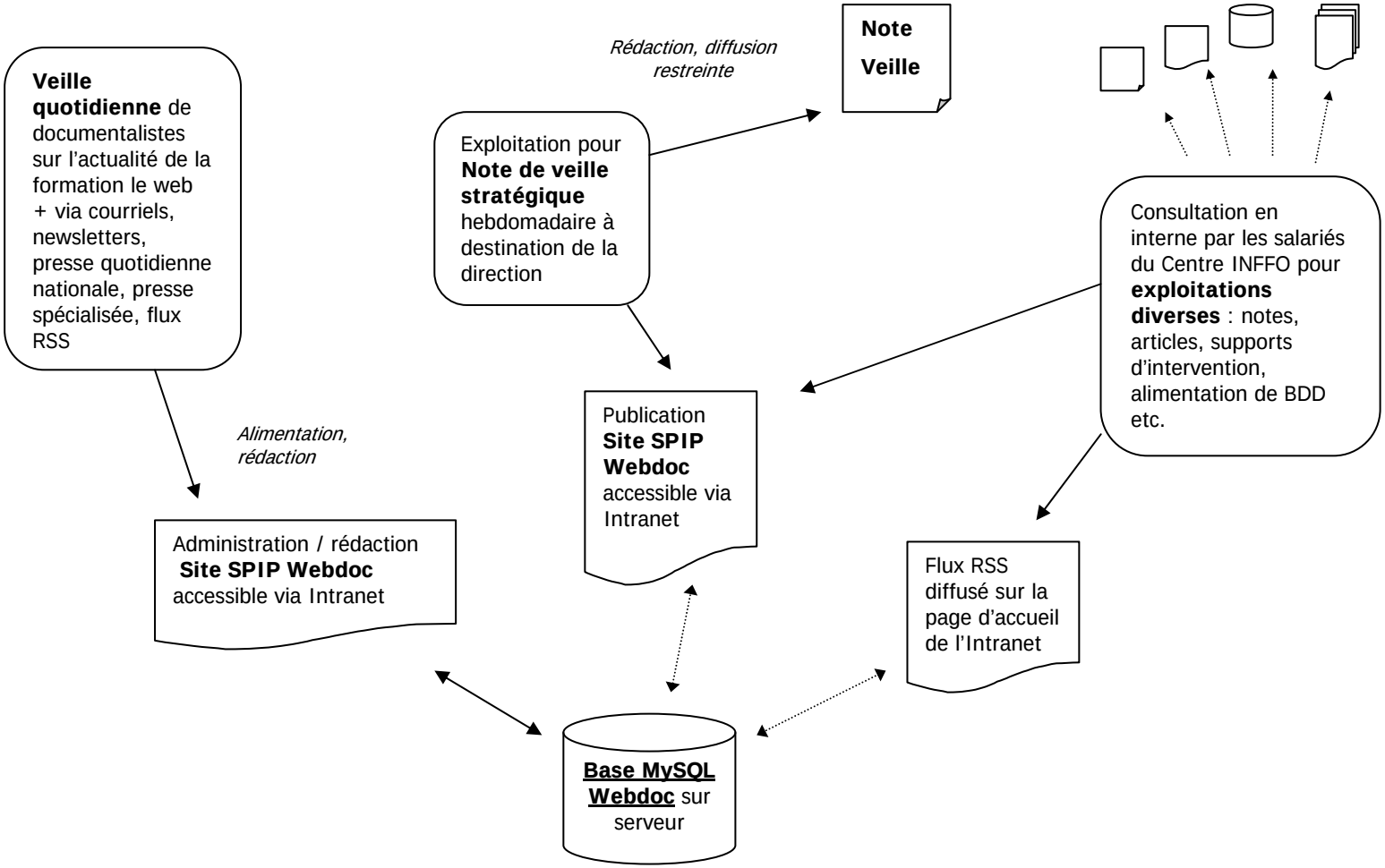
nécessaires. Les résultats de la veille documentaire, plus intégrés, assureraient l'appui des rédacteurs en charge d'alimenter les actuels et futurs sites Internet, en accès libre ou en accès privilégié pour les clients et partenaires du Centre INFFO.

Structure de la base MySQL Webdoc

La base MySQL qui contient les données saisies sous SPIP pour produire le site Webdoc est constituée de **tables SPIP** (version SPIP 2.0.9). Certaines ne sont pas utilisées. Néanmoins, le tableau ci-dessous les liste toutes, telles qu'elles figurent dans la base MySQL Webdoc au 20/11/2009.

Table	Enregistrements	Table	Enregistrements
spip_articles	12 555	spip_mots_syndic	0
spip_auteurs	23	spip_ortho_cache	0
spip_auteurs_articles	12 555	spip_ortho_dico	0
spip_auteurs_messages	8	spip_petitions	1
spip_auteurs_rubriques	14	spip_referers	2 376
spip_breves	1	spip_referers_articles	825
spip_documents	257	spip_resultats	154
spip_documents_liens	249	spip_rubriques	32
spip_forum	2	spip_signatures	0
spip_groupes_mots	2	spip_syndic	6
spip_messages	3	spip_syndic_articles	0
spip_meta	125	spip_test	1
spip_mots	33	spip_types_documents	98
spip_mots_articles	17 631	spip_urls	6 095
spip_mots_breves	0	spip_versions	3 267
spip_mots_documents	0	spip_versions_fragments	40 005
spip_mots_forum	0	spip_visites	304
spip_mots_rubriques	0	spip_visites_articles	4 890

Workflow depuis l'entrée des données jusqu'à leur exploitation et publication



Annexe 7 Base Juriweb : production des « Fiches pratiques de la formation continue »

Nom de la base et/ou Type d'objet	Type de base	Service gestionnaire	Utilisation des données Publication sur le Web
Juriweb : base de production du guide juridique « Les Fiches pratiques de la formation continue »	Via le serveur d'application Zope (SGBD PostgreSQL)	Direction Juridique Observatoire	• **Base Fiches Pratiques** sur le site **Droit de la formation** <http://www.droit-de-la-formation.fr/> • Publication papier annuelle **« Les Fiches pratiques de la formation continue »** (sous la forme de deux volumes) <http://www.centre-inffo.fr/-Les-Fiches-pratiques-de-la-.html>.

Présentation de Juriweb et de ses données

Juriweb est produite au Centre INFFO par l'ensemble des juristes-chargés d'étude. L'interface logicielle qu'ils utilisent est un serveur d'application appelé Zope. Les données sont stockées dans une base Zope, appelée ZODB. Il ne s'agit pas d'une base relationnelle, mais d'une base de données objet.

Les objets contenus se présentent sous forme non pas de tables liées entre elles, mais d'une arborescence d'objets techniques et métier.

- les objets techniques sont les pages dynamiques, les dossiers, les gestionnaires d'authentification, de cookies, de sessions, de templates, l'indexeur des données (Zcatalog), les schémas utilisateurs, etc.

- les objets métier sont par exemple les recueils, les unités (chapitres, fiches, paragraphes...), les encadrés, les descripteurs utilisés pour l'indexation, les sources juridiques, etc.

Une modélisation des données de la base n'a pas été réalisée dans le cadre de ce mémoire. Les informations qui précèdent ont été fournies par l'École des Mines de Paris, qui gère cette base pour le compte du Centre INFFO. L'accès complet aux données de Juriweb (via une interface Web d'administration), ainsi qu'au code source Python de l'application, donnerait en particulier accès aux définitions de toutes les classes d'objets métier[109].

Juriweb permet de produire le guide juridique « Les Fiches pratiques de la formation continue », publié chaque année au format papier, mais également consultable sur le site Droit de la formation <http://www.droit-de-la-formation.fr/>. Ce document est constitué d'un corpus composé de sections, chapitres, fiches, sous-fiches. Cette structure, visible tant dans le document papier que dans les textes présents sur le Web, semble représenter le modèle des données gérées. Cette structure éditoriale semble constituer une sorte de « structure sémantique » qui pourrait être exploitée au travers d'un Schéma (cf. chapitre 1 de la Quatrième partie du mémoire).

D'autres métadonnées sont adjointes, reconnaissables dans le formulaire de requête dans la base : profils d'utilisateurs finaux, natures de sources juridiques, « catégories » de textes, et thèmes, qui sont en fait des descripteurs. En effet, l'indexation des fiches composant « Les Fiches pratiques », est opérée au moyen d'un langage documentaire appelé **ThesJuris**, qui se présente sous la forme d'une liste de descripteurs et synonymes, sans hiérarchie. Ces termes se sont inspirés du vocabulaire du **Thésaurus de la formation**, lors de la création de cette liste, pour les besoins spécifiques de l'indexation des fiches juridiques consultables sur le site Web Droit de la formation.

La version papier des Fiches pratiques, propose également un index, avec des mots-clés proposant une entrée à double niveau. Mais les mots-clés représentés et les renvois vers les fiches ne corroborent pas systématiquement les descripteurs composant ThesJuris et l'indexation des fiches réalisée pour la publication sur le site Web. Une harmonisation serait à opérer.

---

[109] Le site officiel du logiciel Zope présente un mode d'emploi <http://www.zope.org/Documentation/Books/ZopeBook/2_6Edition> (cf. la section 3 relative à l'architecture de Zope, et la section 6 relative à l'interface d'administration – ZMI).

Annexe 8 Guide d'entretien pour l'enquête menée auprès des utilisateurs internes des BDD

→ Axe principal de l'entretien : la production de la base de données, de l'alimentation en informations à l'exploitation des données gérées ; les métadonnées renseignées.

Activité professionnelle	Quel emploi occupez-vous au Centre INFFO ? Quelle est la fonction exercée au regard de la base ?
Présentation de la base de données	Bref historique de la base D'où viennent les données initiales ? Nature du fonds référencé dans la base : types de documents, types de supports gérés, thèmes traités, seulement références ou gestion de documents primaires (GED) ?
Documents relatifs à la base éventuellement disponibles	Manuels relatifs au logiciel (utilisation, administration) Supports de formation Cahiers des charges relatifs à la constitution de la base, reprise de données Formats documentaires, descriptifs de tables Normes ou standards respectés pour le format des données cataloguées ou référencées Langage documentaire utilisé pour l'indexation ?
Production de la Base	Enchaînement des tâches relatives à la production de la base et l'exploitation des données : sélection, indexation, analyse/résumé, saisie, relectures/vérifications, validation (circuit), recherche, export-extraction-publication-diffusion.
Coordination - collaboration	Avec qui êtes-vous en contact pour votre activité sur cette base ? Avec quelles personnes ou quels services êtes-vous amené(e) à vous coordonner et travailler ? Combien de personnes contribuent à l'alimentation de cette base ? Modalités de répartition du travail tout du long du circuit de production ?
Points d'attention relatifs aux données, difficultés, besoins	Difficultés rencontrées Besoins pressentis relatifs aux informations traitées : données à ajouter, normes ou standards à suivre Questionnement relatif à la chaîne de production