



**HAL**  
open science

# L'exploitation des vocabulaires dans un portail web sémantique. Le cas du portail thématique Plan4Learning de l'IIPÉ-UNESCO

Sid-Ali Chikh

► **To cite this version:**

Sid-Ali Chikh. L'exploitation des vocabulaires dans un portail web sémantique. Le cas du portail thématique Plan4Learning de l'IIPÉ-UNESCO. domain\_shs.info.docu. 2014. mem\_01128428v1

**HAL Id: mem\_01128428**

**[https://memic.ccsd.cnrs.fr/mem\\_01128428v1](https://memic.ccsd.cnrs.fr/mem_01128428v1)**

Submitted on 9 Mar 2015 (v1), last revised 3 Jul 2015 (v2)

**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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0  
International License



CONSERVATOIRE NATIONAL DES ARTS ET METIERS

Ecole Management et Société-Département CITS

INTD

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

Présenté et soutenu par

*Sid-Ali CHIKH*

Le 18 décembre 2014

L'exploitation des vocabulaires dans un portail web  
sémantique

Le cas du portail « Plan4Learning » de l'IIFE-UNESCO

Jury :

Nadia Rais - Professeur associée au Cnam

Jean Delahousse - Consultant

Asunción Valderrama – Directrice de la bibliothèque de l'IIFE-UNESCO

## Promotion 43



# Remerciements

Je tiens à remercier

- Nadia Rais pour sa patience, ses encouragements et ses conseils avisés tout au long de la préparation de ce mémoire,
- Jean Delahousse pour l'intérêt soutenu qu'il a bien voulu porter à ce travail,
- Asunción Valderrama pour sa disponibilité et le soutien qu'elle m'a apporté tant pendant le stage que pour la préparation de ce mémoire,
- Corinne Bitoun, Aurore Brillant, Lynne Sergeant pour la qualité de leur accueil, leur disponibilité durant ma mission,
- Toute l'équipe du centre de documentation et de l'IPE-UNESCO pour l'accueil bienveillant dont j'ai pu bénéficier durant mon intervention

## Notice

CHIKH Sid-Ali. L'exploitation des vocabulaires dans un portail web sémantique. Le cas du portail thématique « Plan4Learning » de l'IPE-UNESCO. 2014. 140 pages. Mémoire, ingénierie documentaire, INTD-CNAM, 2014

**Résumé :** ce mémoire porte sur une stratégie de valorisation de ressources documentaires dans un portail web sémantique. Il s'appuie sur une intervention et sur l'analyse du projet de portail Plan4Learning conduit au sein de l'Institut international pour la planification de l'éducation IPE-UNESCO. L'analyse met en particulier en regard la modélisation et l'exploitation de vocabulaires documentaires et l'élaboration de services innovants en matière d'accès à l'information.

**Descripteurs :** accès à l'information ; base de connaissance, bibliothèque ; donnée liée ; descripteur ; éducation ; facette ; gestion de contenu, glossaire ; graphe ; IPE-UNESCO ; indexation, langage contrôlé, linked open data ; liste d'autorité ; mot-clé ; moteur de recherche ; notice catalographique ; ontologie ; organisation des connaissances ; owl ; plan de classement ; portail documentaire ; publication ; rdf ; recherche d'information ; recherche en texte intégral ; référentiel ; skos ; thésaurus ; UNESCO ; valorisation ; vocabulaires contrôlés ; web de données ; web sémantique.

**Abstract** This paper concerns a strategy of valorisation of documentary resources through a semantic Web portal. It proceed from an intervention and a study of the Plan4Learning portal project led within the international Institute for educational planning IPE-UNESCO. The analysis stresses in particular on the interaction linking the semantic description and the use of documentary vocabularies on the one hand and on the other hand the elaboration of innovative services regarding improving access to information.



# Table des matières

CONSERVATOIRE NATIONAL DES ARTS ET METIERS .....	1
Ecole Management et Société-Département CITS.....	1
INTD .....	1
Remerciements.....	4
Notice.....	5
Table des matières .....	7
Liste des tableaux.....	12
Liste des figures .....	14
Introduction .....	16
Première partie Clés de compréhension .....	19
1 Langages documentaires et web de données .....	20
1.1 Le web de données comme extension du web.....	20
1.1.1 L'empilement des technologies .....	20
1.1.2 Une modélisation de domaines de connaissances.....	23
1.1.2.1 Les ontologies .....	23
1.1.3 Quelques réalisations.....	25
1.1.3.1 DBPédia .....	25
1.1.3.2 Linked Open Data .....	25
1.1.3.3 Data.bnf .....	27
1.2 Les traits d'union entre les vocabulaires contrôlés et le web sémantique .....	29
1.2.1 Des vocabulaires contrôlés... pour la recherche et l'indexation. ....	29
1.2.2 Le thésaurus .....	29
1.2.2.1 Définition et norme.....	29
1.2.2.2 Les types de relations .....	31
1.2.2.3 Les apports principaux en indexation et en recherche.....	31
1.2.3 SKOS : un rôle pivot .....	32
1.2.3.1 Capturer le sens .....	32
1.2.3.2 Les grandes propriétés du modèle .....	33
1.3 Domaines de connaissances et accès aux contenus .....	34
1.3.1 L'adaptation des moteurs de recherche.....	34



1.3.2	Expansion sémantique et recherche à facettes.....	35
	Deuxième partie Le cadre de l'expérience .....	37
2	L'élaboration du portail Plan4Learning de l'IPE .....	38
2.1	Une institution internationale : l'IPE .....	38
2.1.1	Une autonomie relative au sein de l'Unesco.....	38
2.1.2	Une action coordonnée par objectif .....	38
2.1.3	Missions et organisation.....	39
2.1.3.1	Formation assistance et partage de connaissances.....	39
2.1.3.2	Une politique active de partenariat.....	39
2.1.3.3	Une organisation intégrée et multilingue .....	40
2.1.3.3.1	Les unités de recherche.....	40
2.1.3.3.3	Le service d'information .....	40
2.2	Le centre de documentation.....	42
2.2.1	Un fonds spécialisé et des publics experts .....	42
2.2.2	L'activité du centre de documentation.....	42
2.2.2.1	Une équipe restreinte et expérimentée.....	43
2.2.2.2	Une exploitation diversifiée des ressources documentaires.....	43
2.2.2.2.1	La gestion et la maintenance de bases de données spécialisées .....	43
2.2.2.2.2	L'exploitation et la mise à disposition du fonds.....	45
2.2.2.3	Une fonction documentaire étendue.....	45
2.3	La qualité des apprentissages : une thématique prioritaire pour l'IPE .....	47
2.3.1	Un thème de stratégie de moyen terme .....	47
2.3.2	Une feuille de route pour Plan4Learning.....	48
2.3.2.1	Ambition générale et objectif du portail .....	48
2.3.2.2	Périmètre et contenu .....	48
2.4	Le projet Plan4Learning dans son contexte documentaire .....	50
2.4.1	Visibilité sur le web et ouverture à de nouveaux publics .....	50
2.4.2	Une exploration du nouveau format de description bibliographique .....	51
2.5	Le projet de portail en phase opérationnelle.....	52
2.5.1	Pilotage et équipe projet.....	52
2.5.1.1	Les prestataires extérieurs .....	52
2.5.5.2	La répartition des tâches entre prestataires .....	53
2.5.2	Phases d'avancement du projet .....	54
2.5.2.1	Phase 1. Moteur de recherche et accès aux ressources documentaires .....	55
2.5.2.2	Phase 2. Communication et mobilisation des communautés .....	55
2.5.3	La précision des objectifs .....	58
2.5.3.1	L'objet du portail .....	58
2.5.3.2	Les publics cibles.....	58

2.5.3.3	Un portail multilingue.....	58
2.5.3.4	Un périmètre documentaire et éditorial resserré : La qualité des apprentissages .....	59
2.5.3.5	Une valeur ajoutée tirée de l'emploi de vocabulaires structurés.....	59
2.5.3.6	Accès simplifié et navigation enrichie .....	60
2.6	Solution logicielle et modélisation des vocabulaires.....	60
2.6.1	Le dispositif logiciel d'ensemble .....	61
2.6.2	L'édition et l'annotation des notices documentaires.....	63
2.6.2.1	Maintenance des notices documentaires et échanges de données .....	63
2.6.2.2	La structuration des notices.....	63
2.6.2.2.1	Rappel des règles et modalités de description en vigueur.....	64
2.6.2.2.2	Limites du développement.....	65
2.6.3	La gestion des vocabulaires et du référentiel .....	66
2.6.3.1	Caractéristiques générales de la solution .....	66
2.6.3.2	Un domaine modélisé de connaissance.....	66
2.6.3.3	Un réservoir de métadonnées .....	66
2.6.3.4	Une solution logicielle étendue .....	67
2.6.3.5	Principes de gestion des vocabulaires dans la base de connaissance.....	69
2.6.3.6	Les fonctions d'administration.....	70
2.6.4	Le moteur de recherche : alimentation, requête et affichage .....	72
2.6.4.1	Les flux d'alimentation du moteur .....	72
2.6.4.1	La mise en forme des résultats .....	72
2.6.4.1.1	Les options liées à la recherche.....	72
2.6.4.1.2	Les choix concernant l'affichage des pages.....	73
Troisième partie Exploitation des vocabulaires et services rendus.....		76
3	Les vocabulaires du portail et leur modélisation.....	77
3.1	Objectif et limites de la description .....	77
3.1.1	Retour à la modélisation .....	78
3.1.2	Le glossaire sur la planification de l'éducation.....	80
3.1.2.1	Les besoins de l'exploitation .....	80
3.1.2.2	La description des termes du Glossaire dans le référentiel .....	82
3.1.3	Le thésaurus de l'Unesco .....	86
3.1.3.1	Périmètre et structuration .....	86
3.1.4	Le plan de classement de l'IPE .....	90
3.1.5	La liste des pays.....	93
3.1.6	Les listes projets et institutions.....	96
3.1.6.1	Les institutions .....	97
3.1.6.2	La liste des projets .....	98

3.2	Enrichissement du glossaire : retour sur la mission .....	100
3.2.1	Périmètre de l'enrichissement.....	100
3.2.1.1	Un état avancé de l'intégration et de l'exploitation .....	100
3.2.1.2	Les besoins et les services visés.....	101
3.2.2	Ciblage des champs descriptifs et <i>modus operandi</i> .....	104
3.2.2.1	Le choix des champs descriptifs.....	104
3.2.2.2	L'extension de la recherche par les descripteurs du thésaurus.....	105
3.2.2.2.1	Un objectif d'expansion sémantique .....	105
3.2.2.3	L'évitement des liens de type « Association » .....	105
3.2.2.4	Le traitement des équivalences.....	106
3.2.3	Modalités d'exécution .....	106
3.2.3.1	Les ressources mises à disposition .....	106
3.2.3.3	Objectifs .....	107
3.2.4	La réalisation des liens : bilan et retour d'expérience.....	108
3.2.4.1	Les liens vers le plan de classement.....	108
3.2.4.2	Les liens vers les descripteurs du thésaurus.....	108
3.2.4.3	Modalité et durée d'exécution .....	109
3.2.5	Validation .....	111
3.2.5.1	Les champs en cours d'enrichissement.....	112
3.2.5.2	Points de vigilance.....	112
3.2.5.2.1	Contrainte de réalisation et <i>workflow</i> de validation .....	112
3.3	Les services rendus : synthèse .....	114
3.3.1	Des accès simplifiés aux ressources .....	114
3.3.2	La démultiplication des points d'accès aux ressources .....	114
3.3.3	Optimisation des requêtes.....	117
3.3.4	Recherche à facettes .....	117
3.4	Bilan et préconisations.....	119
3.4.1	Remarques .....	119
3.4.1.1	Sur la conformité du dispositif aux objectifs de l'IPE.....	119
3.4.1.2	Sur la conduite du projet : opportunité et contraintes.....	120
3.4.2	Préconisations.....	120
3.4.2.1	Poursuivre l'enrichissement du référentiel avec les données DBpédia.....	120
3.4.2.1.1	Pour l'enrichissement des contenus du portail .....	121
3.4.2.1.2	Pour l'exposition des référentiels.....	121
3.4.2.2	Valorisation pérennisation du référentiel.....	122
3.4.2.3	Fiabiliser et interconnecter les identifiants auteurs et organisations .....	123
	Conclusion .....	126
	Bibliographie .....	130



## Liste des tableaux

Tableau 1 Les base de données du centre de documentation de l'IPE.....	44
Tableau 2 Les bases de données dédiées à l'activité interne de l'IPE .....	45
Tableau 3 Répartition des interventions des prestataires extérieurs du projet.....	54
Tableau 4 Récapitulatif des standards et fonctionnalités de gestion des vocabulaires et Thésaurus dans ITM (source offre commerciale Mondeca).....	69
Tableau 5 Les quatre profils d'utilisateurs ont des droits par espace de travail dans l'environnement d'ITM.....	71
Tableau 6 Description des attributs d'un terme du glossaire dans la fonction Edition d'ITM et exemple.....	85
Tableau 7 Description des termes descripteurs du thésaurus de l'Unesco dans la notice d'édition d'ITM et exemple. ....	90
Tableau 8 Description des thèmes du plan de classement dans la notice d'édition d'ITM et exemple.....	93
Tableau 9 Description des notices pays dans le mode édition de l'outil d'ITM et exemple ...	96
Tableau 10 Description de la notice de l'IPE « <i>Cambodian women's Development Agency</i> » dans le portail d'édition. ....	98
Tableau 11 notice du programme PASEC de la liste Projet .....	99
Tableau 12 Les champs descriptifs du glossaire, leur vocabulaire lié et leur exploitation visée. .....	102
Tableau 13 Récapitulatif des modalités d'exécution de la mise en lien du glossaire avec le plan de classement et le thésaurus de l'Unesco.....	108
Tableau 14 Extrait du tableau d'enregistrement des mises en lien du glossaire avec le thésaurus et le plan de classement de l'IPE. ....	110
Tableau 15 Récapitulatif de l'avancement et des suites à donner.....	113
Tableau 16 La notice documentaire du rapport de la direction de la Banque mondiale intitulé « <i>Incentives to improve teaching: lessons from Latin America</i> » .....	116
Tableau 17 Propriétés des différentes facettes dans leur libellé en français et leur vocabulaire ou liste sources.....	118



# Liste des figures

Figure 1 L'empilement des couches technologiques du web de données (Layer cake » ..... 21	21
Figure 2 Schéma d'ensemble de la solution logicielle (hors CMS du portail final) (source : présentation Sparna) ..... 62	62
Figure 3 L'interface de recherche et de navigation à facettes dans les notices documentaires du portail d'exploitation. (Source capture d'écran <a href="http://plan4learning.iiep.unesco.org/plan4learning/">http://plan4learning.iiep.unesco.org/plan4learning/</a> ) ..... 74	74
Figure 4 La page dédiée à la recherche et à la navigation alphabétique dans le Glossaire.. 75	75
Figure 5 Extrait du champ de description des mots clés (Keywords) dans l'interface d'édition d'une notice bibliographique ..... 79	79
Figure 6 : Affichage d'une page d'un terme du glossaire dans le portail d'exploitation (Source : capture d'écran)..... 82	82
Figure 7 Figure 1 : Schéma des champs de description du périmètre d'enrichissement indiquant les vocabulaires liés et les services rendus en <i>front office</i> ..... 103	103





# **Introduction**

Le présent travail traite d'un dispositif original de valorisation de ressources d'informations. Il s'appuie sur la mission qui m'a conduit à contribuer à une phase d'élaboration du portail thématique « Plan4Learning » conçu au sein du service Information de l'Institut International pour la Planification de l'Éducation (IIPÉ-Unesco). Ce portail web est dédié à la qualité des apprentissages et à la qualité de l'éducation. Il s'inscrit de plain-pied dans la vocation de cette institution internationale d'apporter aux cadres et aux responsables des programmes et stratégies d'éducation des pays en développement les connaissances et les outils utiles, renouvelés, adaptés à leur sphère d'intervention. Ce projet est innovant à plus d'un titre. Il mobilise les standards du web de données dans un dispositif de publication et de mise à disposition de ressources documentaires et leur articulation à des contenus éditoriaux ambitieux ( étude de controverse, actualités, couverture thématique, forum de collaboration, etc.). Au sein de l'équipe projet, mon intervention sur un glossaire spécialisé m'a conduit tout particulièrement à m'intéresser à la façon dont ce dispositif exploitait dans une architecture technologique innovante une pluralité de langages documentaires destinés à l'indexation et à la recherche d'information. La mission d'enrichissement du référentiel sur lequel porte et s'adosse le projet m'a permis de prendre part très concrètement à l'évaluation d'un vocabulaire, à son alignement avec des référentiels d'indexation de portée générale ( Thésaurus de l'Unesco) ou locale (Plan de classement). Au sein du dispositif cette opération est de bout en bout ajustée à des fonctionnalités et des services destinés aux utilisateurs finaux du futur portail Plan4Learning : navigation, repérage, affichage de définitions multilingues, chemins de découverte et de rebonds dans les vocabulaires et entre les vocabulaires, expansion de la recherche de notices documentaires à partir d'un moteur de recherche sémantique.

Aussi, l'objectif général de mon mémoire tient-il à l'exploration et à la description de la façon dont ces vocabulaires sont exploités et à la compréhension des services utilisateurs auxquels ils sont destinés. Je traite cet aspect en trois parties.

Dans un premier temps je livre sous forme de rappels théoriques et de définitions quelques clés de compréhension des technologies du web de données et de leur articulation récente avec les langages documentaires.

Dans une deuxième partie j'inscris le projet de portail Plan4Learning dans son environnement institutionnel, documentaire et opérationnel.

Dans une troisième partie je livre un exposé des vocabulaires et une synthèse des services générés par leur modélisation. J'enrichis cette description par un retour détaillé sur ma mission et quelques pistes de préconisation.



# **Première partie**

## **Clés de compréhension**

# 1 Langages documentaires et web de données

---

Dans cette première nous présentons comme clés de compréhension, quelques rappels et analyses destinées à éclairer les rapports qu'entretiennent dans un dispositif de valorisation de contenus tel que celui dont traite ce travail, les standards du web de données, les langages documentaires et quelques uns des traitements automatisés de la recherche d'information

## 1.1 Le web de données comme extension du web

Le web de données procède de l'idée formulée par son fondateur concepteur Tim Berners Lee d'étendre le web à un vaste univers de données structurées, reliées entre elles sémantiquement et interprétables par les machines<sup>1</sup>. Il s'agit de dépasser les limites d'un web antérieur, parfois appelé web documentaire dont les langages ne permettaient que de relier des documents au moyen de liens hypertextes – les pages html - assimilables par des utilisateurs mais dont la composition interne, non structurée, demeuraient hors de portée du traitement par les machines afin de constituer une vaste base de données reliées entre elles. Dans ce sens, les technologies du web de données vont progressivement être élaborées et adoptées sous forme de recommandations pour la mise à disposition des données dans des standards garantissant leur interopérabilité, la mise en relations des données entre elles et pour permettre aux données d'être partageables et ré-utilisables.

### 1.1.1 L'empilement des technologies<sup>2</sup>

Plusieurs couches de technologies et de standards contribuent à ce cadre d'échange général. Elles sont représentées dans le désormais fameux « layer cake » du web sémantique. [voir Figure 1]. Cet empilement illustre bien la continuité du web de données avec les piliers technologiques robustes du web « traditionnel » que sont le protocole d'échange http<sup>3</sup> et l'identifiant Uri.

---

<sup>1</sup> Exposé dans la page du W3C consacrée au web sémantique <http://www.w3.org/2001/sw/>.

<sup>2</sup> Les exposés sur les technologies du web quand ils ne renvoient pas à une source s'appuient largement sur éléments de cours suivis pendant la formation à l'INTD et tout particulièrement ceux de Mme Nadia Rais sur le web, standards et enjeux techniques.

<sup>3</sup> Acronyme de HyperText Transfer Protocol est le protocole de communication client serveur qui permet la communication entre le navigateur du poste client et le serveur web

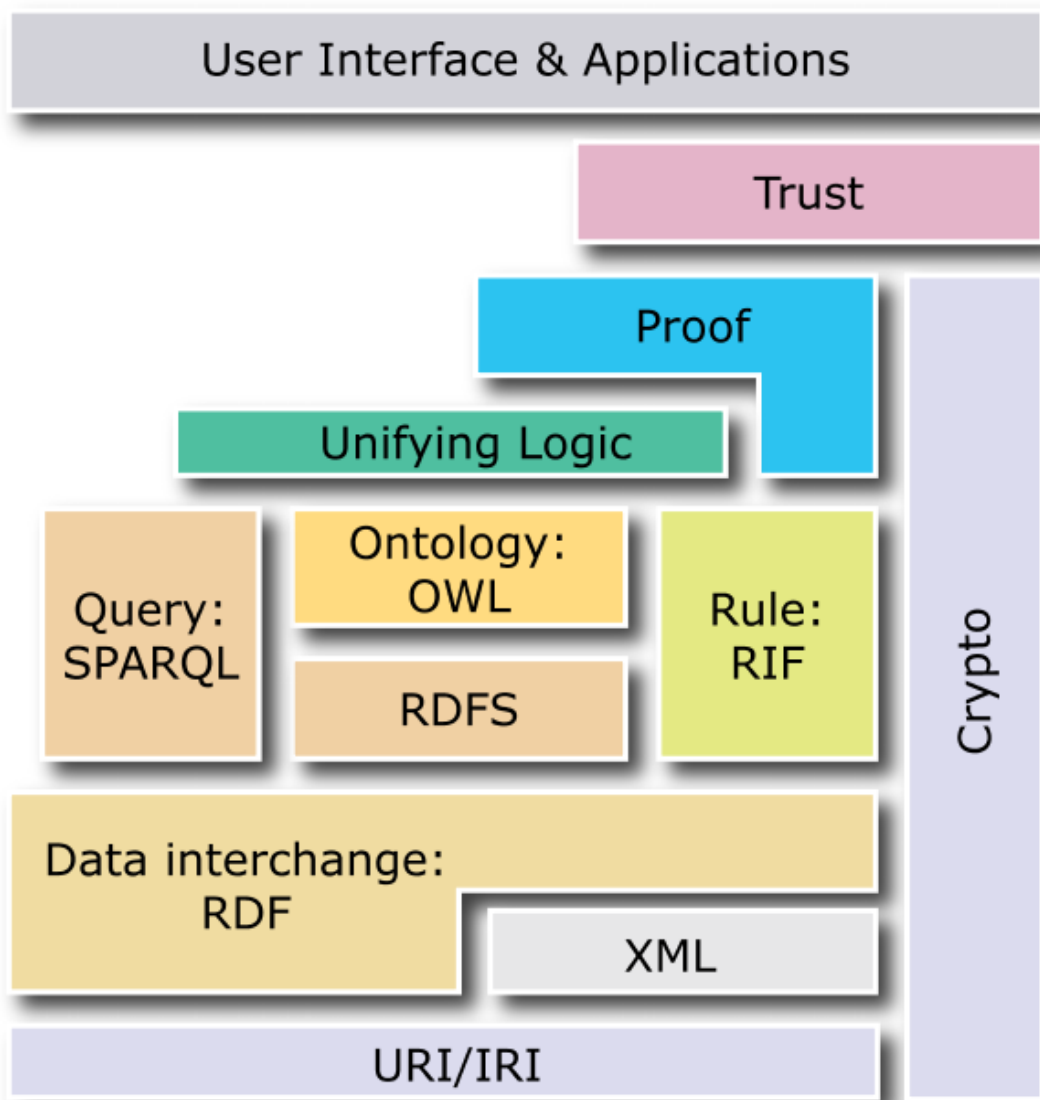


Figure 1 L'empilement des couches technologiques du web de données « Layer cake » (Source : W3C)

**URI** : Uniform Resource Identifier (identifiant de ressource) est un format d'identifiant unique permettant de nommer et de faire référence à n'importe quelle ressource sur le web. Il se décompose en un URN pour le nommage d'une ressource indépendamment de son emplacement et en un URL Uniform Resource Locator qui spécifie la localisation d'une ressource sur internet et le moyen d'y accéder.

**XML** : Extensible Markup Language (Langage de balisage extensible) est un format textuel d'échange de documents et données structurés qui permet d'introduire plusieurs éléments de structuration au moyen de balises sémantiques enserrant des chaînes de caractères (chevrons). Il a été mis au point par le W3C en 1996.

Cette structuration est décisive pour rendre les données calculables par une machine (arithmétique, comparaison). Un document XML se compose ainsi d'un élément racine, de balises sémantiques au nombre non prédéfini et d'un arbre présentant des éléments de structuration de façon hiérarchique.

Son utilisation est déjà très répandue pour l'échange de métadonnées spécifiques à la documentation (Dublin Core, MARC XML, NEWSML pour les informations d'actualité et articles de presse, etc.).

Le standard de structuration XML est en particulier celui qui sert de base au dispositif de fédération de contenus et d'échanges de métadonnées de collections OAI PMH (Archives Initiative Protocol for Metadata Harvesting)<sup>4</sup>. Un document XML peut être transformé au moyen de feuilles de style de type CSS ou XSLT et affichable dans une page web html.

**RDF** : Resource Description Framework est le premier standard du web de données. Il est une recommandation du W3C depuis 2004. Son objectif est de partager les métadonnées pour toutes ressources au moyen d'un modèle et de plusieurs syntaxes dont XML [ GANDON, p.18]

Son organisation en triplets est la base du web de données liées sur le web.

L'acronyme RDF articule trois éléments :

- une ressource matérielle ou immatérielle identifiée par un URI : cela peut être une page web, un objet, un événement, etc.
- les caractéristiques des ressources et de leurs relations
- le modèle et la syntaxe des descriptions.

Ce modèle permet de décrire toutes les ressources au moyen de la syntaxe minimale Sujet-verbe-complément du type « L'Assemblée nationale est située à Paris » exprimé en langage naturel.

Une déclaration RDF se décrit quant à elle par Sujet-prédicat-objet.

- Le sujet est la ressource : l'Assemblée nationale ;
- Le prédicat, le nom de la propriété de la ressource : « est située à » ;
- L'objet, la valeur ou la caractéristique exprimée par une chaîne de caractère « Paris ». Il peut aussi être tout autre littéral comme un nombre ou une date.

---

<sup>4</sup> Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) est un protocole informatique développé par l'Open Archives Initiative afin d'échanger des métadonnées. Il permet de constituer et de mettre à jour automatiquement des entrepôts centralisés où les métadonnées de sources diverses peuvent être interrogées simultanément.  
[http://fr.wikipedia.org/wiki/Open\\_Archives\\_Initiative\\_Protocol\\_for\\_Metadata\\_Harvesting](http://fr.wikipedia.org/wiki/Open_Archives_Initiative_Protocol_for_Metadata_Harvesting)

La description d'un triplet est représentée sous forme de graphe de relations. Celui-ci illustre qu'une ressource peut être le sujet ou l'objet d'autres déclarations.

L'ensemble des nœuds et des liens forment un graphe potentiellement infini de données liées décrivant des ressources formellement identifiées (par URI) sur le web.

**SPARQL** (Simple Protocol and RDF Query Language) : est le langage qui permet de faire des requêtes sur les graphes RDF. L'interrogation propre à SPARQL repose sur une forme symétrique des triplets RDF : Ressource – propriété – valeur. Elle se distingue en cela des langages d'interrogation des bases de données relationnelles de type SQL dont elle s'inspire. Elle permet de rechercher, modifier, d'ajouter, ou supprimer des données RDF sur la base d'une mise en vis-à-vis « matching ». « Poser une question en SPARKL « consiste, résume Fabien Gandon à écrire un graphe requête pour lequel on recherche des occurrences dans le graphe cible » [GANDON, p. 50]. Le format de résultat est normalisé pour être présenté soit en RDF soit en XML. Des recommandations additionnelles ont étendu les fonctions de requêtes à l'extraction de données RDF inscrite dans une page web.

## **1.1.2 Une modélisation de domaines de connaissances**

### **1.1.2.1 Les ontologies**

RDFS et OWL (Ontologie Web Langage) introduisent les schémas et les ontologies empilés qui forment les couches supérieures du web sémantique.

Ils s'agit de modèles plus ou moins avancés (dits légers et lourds) d'organisation des données et de caractérisation de leur relation identifiés par les URI.

La conceptualisation modélise un espace de connaissances nourri des spécifications concernant les composantes du domaine (classes, catégories, instance etc.), la structuration du domaine (hiérarchie, ex : subsumption de la catégorie voiture dans la catégorie véhicule) le type de relation ( inférences logiques,) qu'elles entretiennent ainsi que les contraintes propres enserrant ces relations.

Dans son acception la plus large, une ontologie consiste dans « la formulation exhaustive et rigoureuse de la conceptualisation d'un domaine » [GANDON, p. 84.]

Sur le plan opératoire, précise le chercheur, les ontologies «(...) fournissent un référentiel pour construire un système à base de connaissances et permettre des inférences pour la recherche d'information, la gestion des connaissances, l'enrichissement automatique des données etc. et pour assurer l'interopérabilité entre les systèmes qui la partagent ».



Les domaines représentés dans le formalisme des ontologies peuvent être de toute nature : santé, tourisme, commerce de biens etc.

Le schéma RDF Schéma est rangé dans les ontologies dites légères. Il permet de nommer des classes et des propriétés, de donner la signature de ces propriétés et de définir une organisation hiérarchique de ces classes et propriétés. [GANDON, p.88 ]

OWL est un schéma plus sophistiqué. Au delà de la représentation hiérarchisée de classes et de propriétés il introduit un éventail plus large permettant d'exprimer pour les classes et leur propriété des relations d'équivalence, d'égalité, de contraire, de symétrie, de cardinalité.

Nous verrons un peu plus loin comment le modèle SKOS (simple knowledge organisation system) système simple d'organisation des connaissances peut être présenté comme un trait d'union parmi les plus consensuels entre les technologies du web sémantique et les théoriciens et les professionnels du traitement des langages documentaires.

Comme y invite Bernard Vatant, [Vatant, 2008] il est permis de s'en tenir à une conception plus pragmatique des ontologies, délestée en somme d'une prétention élevée à la scientificité. L'auteur opte en particulier s'agissant des ontologies destinées à faciliter l'accès à l'information et aux connaissances pour une acception de type contractuel. L'ontologie est le produit d'un compromis : « les éléments définis dans une ontologie résultent d'un consensus entre utilisateurs sur les choses à décrire et la façon de les décrire. (...) Le fait d'inclure ou non telle règle dans l'ontologie est donc plus une question d'accord social que d'objectivité scientifique. L'expression de ce consensus dans un langage formel permet de mettre à plat la terminologie, d'explicitier les désaccords éventuels et les approximations nécessaires, le niveau de granularité désiré, les compromis liés aux contraintes de performance du système ou à l'intelligibilité des interfaces utilisateurs, etc. »

La gradation pourtant bien réelle qui sépare le web « tout court » du web de données, reste souvent méconnue des millions d'utilisateurs qui utilisent le web quotidiennement. Reste que le mouvement enclenché depuis la fin des années 1990 n'a pas manqué d'attester de sa pertinence et de ses potentialités auprès de nombreux spécialistes, organisations publiques ou privées et évidemment de professionnels de la préparation, de la recherche et de la diffusion d'informations.

Sans les hiérarchiser, on peut rapporter les avantages trouvés et exploités par le modèle universel de représentation de connaissance RDF.

- La possibilité de relier et d'échanger des données au moyen d'URI
- L'exploitation de données décloisonnées des silos de bases de données (comme par exemple rendre les données de catalogues bibliographiques accessibles aux moteurs de recherche voire sortir de la notion même de catalogue)

- Le caractère indéfiniment extensible d'un graphe reliant et enrichissant du même coup la description d'un domaine de connaissance
- La mise à disposition par la publication de source de données différentes réutilisables.

### 1.1.3 Quelques réalisations

#### 1.1.3.1 DBPédia

L'initiative DBPédia figure parmi les premières expérimentations de RDF. L'initiative a été lancée en 2007 par une collaboration entre l'université de Mannheim, l'université libre de Berlin et l'Open Link Software. Elle consiste dans l'extraction et la publication sous licence libre de l'ensemble des métadonnées issues de ressources de l'encyclopédie collaborative et multilingue Wikipédia.

Les descriptions mises à disposition sur le site web de DBPédia s'appuient autant sur rdf que sur des schémas de descriptions plus spécialisés comme FOAF<sup>5</sup> pour représenter les personnes ou Dublin Core<sup>6</sup> pour les métadonnées des ressources numériques et physiques.

La version anglaise de DBPédia décrit ainsi plus de 4,5 millions de ressources (personnes, lieux, œuvres artistiques, organisations<sup>7</sup>).

La maintenance de DBPédia en français est assurée par la plateforme Semanticpedia<sup>8</sup>. Elle est animée par un collectif composé de l'équipe Wimmics de l'Inria, du Ministère de la culture et de la communication et de l'association Wikimedia France.

#### 1.1.3.2 Linked Open Data

L'initiative Linked Open data portée au sein du W3C par le groupement Semantic web education and outreach ( SWEO) donne un coup d'accélérateur à l'expansion du partage de

---

<sup>5</sup> Friend Of A Friend est un vocabulaire publié sous licence creatives commons permettant de décrire des personnes les relations entre elles, leurs activités photos, calendriers, web blogs, etc. (...) il est développé depuis 2000 par une communauté de contributeurs à laquelle participe le W3C. )

<sup>6</sup> Le Dublin Core (DC) est un schéma de données génériques qui sert à décrire les ressources numériques et physiques. Il contient 15 champs de descriptions (Titre, auteur, éditeur, etc.) et une version enrichie. [DCMI](#) (Dublin Core Metadata Initiative) est une organisation créée dans la perspective du mouvement linked open data qui discute des composantes sémantiques du schéma Dublin core.

<sup>7</sup> Dont 49 000 institutions éducatives renseignées dans 125 langues.

<sup>8</sup> [www.semanticpedia.org/](http://www.semanticpedia.org/)

données liées en impulsant ce mouvement dans la sphère des données ouvertes. Dans cette voie, Tim Berners Lee a formulé en 2010 une gradation de la qualité des données ouvertes qui va de une étoile pour des données non filtrées, éventuellement dégradées, et mises en ligne avec n'importe quel format, à cinq étoiles pour les données liées à d'autres données susceptibles d'être contextualisées et enrichies<sup>9</sup>

En France le ministère de la Culture est à bien des égards à l'avant scène de cette innovation.

---

<sup>9</sup> Le classement est accessible [en ligne] [http://fr.wikipedia.org/wiki/Donn%C3%A9es\\_ouvertes](http://fr.wikipedia.org/wiki/Donn%C3%A9es_ouvertes)

### 1.1.3.3 Data.bnf

La bibliothèque nationale de France a expérimenté une exposition et la mise en relation des données de son catalogue à travers la plate forme data.bnf.

Celle-ci articule le modèle conceptuel FRBR<sup>10</sup> élaboré au sein de l'IFLA<sup>11</sup> pour la description des notices bibliographiques et les standards du web sémantique pour relier ses données aux pages équivalentes dans d'autres sources de données. Le référentiel d'indexation des sujets RAMEAU<sup>12</sup> est aligné aux sujets de la Bibliothèque du Congrès, de la Bibliothèque nationale allemande, de l'Insee, des archives de France, thésaurus Agrovoc ; les auteurs à d'autres sources telles que le répertoire internationale des identifiants auteurs (ISNI), DBPédia, etc. Les données sont exposées sur le web au travers des différentes déclinaisons de RDF. Elles sont récupérables par un terminal d'interrogation SPARQL *Endpoint*.

Pour la Bnf, les standards du web sémantique sont explicitement présentés comme une solution à des problèmes anciens « (...) des bibliothèques, en particulier autour de la gestion de formats divers et de l'échange de données. Ils offrent une présence nouvelle de ces ressources sur le web en les rendant plus facilement accessibles et réutilisables et en les liant à des ressources complémentaires. »<sup>13</sup>

---

<sup>10</sup> Le modèle FRBR (Functional Requirements for Bibliographic Records / Fonctionnalités requises des notices bibliographiques) est un modèle conceptuel de données bibliographiques élaboré par un groupe d'experts de l'IFLA de 1992 à 1997. Il a été officiellement approuvé par le Comité permanent de la Section de catalogage de l'IFLA le 5 septembre 1997, et le rapport final du groupe d'experts a été publié dès l'année suivante aux éditions Saur.  
[http://www.bnf.fr/fr/professionnels/modelisation\\_ontologies/a.modele\\_FRBR.html](http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.modele_FRBR.html)).

<sup>11</sup> <http://www.ifla.org/>

<sup>12</sup> « Le répertoire d'autorité-matière encyclopédique et alphabétique unifié est un langage d'indexation matière. Ce langage documentaire est utilisé, en France, par la Bibliothèque nationale de France, les bibliothèques universitaires, de nombreuses bibliothèques de lecture publique ou de recherche ainsi que plusieurs organismes » privés. <http://rameau.bnf.fr/>

<sup>13</sup> L'exposé détaillé du projet est accessible sur le site data.bnf.fr ; voir aussi le projet Isidore en matière de diffusion des publications en sciences humaines et sociales ou le plus récent projet Joconde Lab et son exploitation de DBpédia pour optimiser l'accès multilingue aux ressources des musées de France. La présentation du projet est disponible sur le site de l'Institut de recherche et d'innovation, IRI.  
<http://www.iri.centrepompidou.fr/experimentations/joconde-lab/>).

L'expérience très vite survolée de Data.bnf, illustre que le cadre et les langages du web de données représentent au moins une double opportunité pour les praticiens de l'information qu'il s'agissent des professionnels de la documentation, des bibliothèques et plus généralement des acteurs qui « nativement » ont chacun dans sa sphère une expérience de l'information structurée. On pense bien sûr à l'immense savoir historiquement accumulé pour identifier, classer, accéder et mettre à disposition des documents. On pense du même coup à l'élaboration et au maniement des vocabulaires contrôlés utilisés pour l'indexation, le catalogage, et la recherche, aux répertoires de métadonnées discutés et standardisés (Dublin Core) et stabilisés pour la description des ressources documentaires sur papier ou sur format numérique.

La première opportunité est de nature socio-technique. Elle repose sur l'affinité très forte de la sphère professionnelle de l'information et de la documentation avec les ambitions du web de données et tout particulièrement son effort visant à asseoir son idéal de partage et d'expansion infinie de la connaissance sur la formalisation avancée de données structurées. Sur ce terrain, le monde de la documentation a de bonnes raisons de se considérer en pays de connaissance [COTTE, 2011]. A l'échelle du tournant numérique et de la transformation des rapports à l'information induite par le web et les moteurs de recherche de type Google, le web de données autorise une jonction originale entre des savoirs et des interrogations anciennes et des technologies innovantes qui redessinent les frontières de l'information.

La seconde opportunité est de type stratégique en ce sens qu'elle est susceptible d'orienter l'action.

Le web de données ouvre des possibles. Ses standards d'interopérabilité et d'échanges sont autant d'instruments disponibles pour une reconfiguration des services d'accès à l'information, des rapports avec des utilisateurs et des publics dont les pratiques de recherche d'information sont fortement refaçonnées par la généralisation du web, l'interrogation en langage naturel, la navigation et la découverte optimisée de contenus, etc. Le décloisonnement des bases de données, l'ouverture des catalogues, la publication de leur données de description sur web, la création de portail destinés à des publics ciblés dessinent quelques uns des ingrédients de ce qu'on appellera tout au long de ce travail une stratégie de valorisation des ressources.

## **1.2 Les traits d'union entre les vocabulaires contrôlés et le web sémantique**

La deuxième série de notions que l'on voudrait introduire ici est celle de langage documentaire. Nous nous arrêterons principalement sur la définition et la structuration de la forme Thésaurus.

Nous aborderons ensuite comment les spécifications Skos importent dans le web de données la structuration propre au thésaurus, mais dispose d'une souplesse de description qui permet à ce modèle de représenter d'autres formes de vocabulaires métiers, listes d'autorités, glossaires, qui participent de la modélisation d'un domaine de connaissance.

### **1.2.1 Des vocabulaires contrôlés... pour la recherche et l'indexation.**

L'ADBS donne la définition suivante du langage documentaire : « langage contrôlé et normalisé utilisé dans un système documentaire pour l'indexation et la recherche. Un langage documentaire permet de représenter de manière univoque les notions identifiées dans les documents et dans les demandes des utilisateurs, en prescrivant une liste de termes ou d'indices, et leurs règles d'utilisation. »<sup>14</sup>

La définition distingue les langages combinatoires ou post-coordonnés dont les éléments peuvent être combinés entre eux a posteriori lors de l'indexation ou de la recherche (thésaurus, classifications à facettes) les langages pré-coordonnés, contenant des combinaisons de notions établies a priori (classifications hiérarchiques, listes de vedettes-matières) ; les langages poly-hiérarchiques (chaque terme du langage peut avoir plusieurs termes génériques de niveau immédiatement supérieur) ou mono-hiérarchiques. -

Elle englobe les grandes familles de langages classificatoires ( la classification de Dewey 1876, la CDU) ; les répertoires de vedettes matière ( RVM de la bibliothèque de l'université de Laval-Québec, Rameau) utilisée pour l'indexation libre dans les bibliothèques, la forme thésaurus, d'apparition plus récente.

### **1.2.2 Le thésaurus**

#### **1.2.2.1 Définition et norme**

« Dans sa plus simple expression rappelle Michèle Hudon [HUDON, p. 21], le thésaurus est une liste de mots simples ou d'expressions ayant valeur de terme dans un domaine plus ou

---

<sup>14</sup> Association des professionnels de l'information et de la documentation (ADBS)  
[http://www.adbs.fr/langage-documentaire-17593.htm?RH=OUTILS\\_VOC](http://www.adbs.fr/langage-documentaire-17593.htm?RH=OUTILS_VOC)

moins limité de la connaissance. Le terme est défini ici comme la représentation linguistique (symbolique) d'un concept unique : ainsi « Alphabétisation », « Autobus scolaire », « marketing » et « orientation professionnelle » sont des termes.

Une définition normalisée de 1981 introduit quant à elle les éléments de composition, de structuration et les principales finalités du thésaurus. « Un thésaurus est une liste d'autorité organisée de descripteurs et de non-descripteurs obéissant à des règles terminologiques propres et reliées entre eux par des relations sémantiques (hiérarchiques, associatives, ou d'équivalence). Cette liste sert à traduire en un langage artificiel dépourvu d'ambiguïté des notions exprimées en langage naturel.

La nouvelle norme thésaurus 25964 -1 ( en anglais, non traduite )<sup>15</sup> est plus explicite encore, elle définit le thésaurus comme un : « Vocabulaire contrôlé et structuré dans lequel les concepts sont représentés par des termes et organisé de façon à rendre explicites les relations entre les concepts et dont les termes préférentiels sont accompagnés d'entrées sur les synonymes et les quasi-synonymes. L'objectif d'un thésaurus est de guider l'indexeur et le chercheur d'information à sélectionner le même terme préférentiel ou bien une combinaison de termes préférentiels pour représenter un sujet donné. Pour cette raison un thésaurus est optimisé pour l'exploration humaine et la couverture terminologique d'un domaine. »<sup>16</sup>

Cette norme décrit des spécifications et des recommandations très étendues sur les règles d'organisation des thésaurus. Nous nous en tenons aux principes très généraux de structuration qui pour l'essentiel demeurent inchangés par delà des désignations nouvelles.

Un thésaurus est composé :

- de termes préférentiels TP (appelés antérieurement Termes Descripteurs) ce sont les termes retenus pour l'indexation (noms communs, noms propres, unitermes ou composés etc.) ;
- de termes non préférentiels TNP ( ex-non descripteurs) donc non retenus pour l'indexation mais liés aux termes préférentiels par l'invitation à « Employer » ( symbole EMP) un terme

---

<sup>15</sup> <https://www.iso.org/obp/ui/fr/#iso:std:iso:25964:-1:ed-1:v1:en> [accès partiel] « 2.62 thesaurus controlled (2.12) and structured vocabulary (2.56) in which concepts (2.11) are represented by terms (2.61), organized so that relationships between concepts (2.11) are made explicit, and preferred terms (2.45) are accompanied by lead-in entries for synonyms (2.58) or quasi-synonyms (2.47) Note 1 to entry: The purpose of a thesaurus is to guide both the indexer and the searcher to select the same preferred term or combination of preferred terms to represent a given subject. For this reason a thesaurus is optimized for human navigability and terminological coverage of a domain. »

<sup>16</sup> C'est moi qui traduit.

préférentiel correspondant. Symétriquement le terme préférentiel indique son équivalent par « Employé pour » (symbole EP).

- de notes explicatives optionnelles (NE) qui précisent le périmètre d'emploi d'un terme préférentiel.
- de listes annexes qui regroupent des termes ( nom géographiques, organisme, noms de pays, etc) non reliés au termes préférentiels.
- De mots outils ( des termes préférentiels qui aident à qualifier d'autres termes préférentiels).

Le thésaurus peut enfin organiser le regroupement thématique de concepts dans des Microthésaurus appelés désormais Groupes de concepts dans la nouvelle norme.

### **1.2.2.2 Les types de relations**

Les relations entre les termes sont de trois types :

- Equivalence : relation de synonymie entre un terme préférentiel et un terme non préférentiel symbolisée par EM et EP
- Hiérarchie : relation d'inclusion qui exprime un rapport de supériorité ou de subordination entre concepts et entre termes. Elle se décline en relation générique / spécifique (de type genre espèce symbolisé par TG – terme générique TS terme spécifique), partitive ( de type tout partie) et d'instance ( est un spécimen de)
- Association : relation réciproque par paires de termes préférentiels relevant de hiérarchies différentes dont on veut souligner les affinités pour optimiser la recherche (symbolisé par TA).

### **1.2.2.3 Les apports principaux en indexation et en recherche**

L'élaboration, l'entretien ( mise à jour, gestion des candidats Préférentiel), et l'emploi de ces langages contrôlés visent principalement à pallier la faiblesse du langage naturel ( entendu comme non contrôlé) dont les ambiguïtés, les imprécisions et les variations le rendent impropre à décrire de manière précise et univoque le contenu de documents.

En proposant une liste de termes normalisés pour représenter des concepts, un thésaurus par exemple affinera l'indexation en restreignant la prolifération de formes susceptibles d'être employées pour décrire le même concept et autorisera du même coup une indexation sur le terme le plus précis.

Dans une procédure d'interrogation d'un système documentaire, le recours à un terme contrôlé ( descripteur de thésaurus) corrigera le bruit ( la proportion de résultat non



pertinents) engendré par le traitement des moteurs de recherche qui pour traiter leur requête indexent l'ensemble des composantes d'un texte ou d'un document sous la forme d'une succession de chaînes de caractères séparées par des blancs.

La structure hiérarchique des thésaurus optimisera encore la recherche en permettant au moteur de recherche d'interroger un concept décrit par des termes reliés par lien d'équivalence et encore d'élargir une requête aux termes équivalents, et aux termes compris dans sa hiérarchie ( auto-postage ascendant ou descendant).

### **1.2.3 SKOS : un rôle pivot**

#### **1.2.3.1 Capturer le sens**

Selon Antoine Isaac [ISAAC 2011] Skos<sup>17</sup> peut être décrit comme une tentative d'importer dans le web de données, la part sémantique qui échappe presque par nature aux modélisations formelles de type OWL. « La question de l'interopérabilité se pose pour les ontologies au niveau sémantique, » est-il expliqué. Ce que l'auteur nomme des systèmes d'organisation de connaissances (KOS) renvoie très précisément aux vocabulaires contrôlés de type thésaurus, taxonomie, classifications. Ces référentiels décrivent localement des connaissances au moyen de concepts et d'inférences dotés de signification pour les humains. Ils sont en cela très éloignés du raisonnement formel des ontologies. Skos s'emploie à capturer ce sens par la représentation de concept (conceptSchéma), de données terminologiques ( Alt Label Pref Label) attachées à ces concepts, de relation sémantiques entre concepts (générique/ spécifique) et de notes d'applications.

Autrement dit, Skos est un modèle de description de concept dont les déclinaisons en font une forme particulièrement adaptée aux structures des langages documentaires de type thésaurus dont il restitue assez bien l'essentiel des relations et du sens.

Tandis que sur le plan formel, Skos reste imbriqué dans les modèles d'inférences d'ontologie OWL. De sorte que comme le souligne [GANDON, p.127] « les concepts décrits dans SKOS sont traités comme des instances de classe ou de propriétés d'OWL et que les raisonnements qui peuvent être faits sur des données Skos sont des raisonnements sur des instances de classe et de propriété. »

La recommandation Skos explicite cette nuance importante : « Pour comprendre cette distinction, il faut savoir que la "connaissance" explicitée dans une ontologie formelle est exprimée à l'aide d'un ensemble d'axiomes et de faits. Un thésaurus ou une classification sont d'une nature complètement différente, et n'expriment pas d'axiomes ou de faits. Un

---

<sup>17</sup> Skos est une recommandation du W3C depuis 2009

thesaurus ou une classification cherchent plutôt à identifier et à décrire, à travers la langue naturelle et d'autres moyens informels, un ensemble d'idées distinctes les unes des autres, qui sont parfois commodément appelées des "concepts". Ces "concepts" peuvent être arrangés et organisés en différentes structures, souvent des hiérarchies ou des liens associatifs. Ces structures, cependant, n'ont aucune sémantique formelle, et ne peuvent pas être interprétées comme des axiomes formels ou des faits sur le monde avec un degré de confiance suffisant. En réalité ces structures n'ont jamais eu cette prétention, leur seul objectif étant de fournir une carte pratique et intuitive d'un certain domaine, qui peut être utilisé comme une aide à l'organisation et à la recherche d'objets, tels que des documents, qui ont trait à ce domaine.

Rendre explicite la "connaissance" contenue dans un thesaurus ou une classification d'une façon formelle, demande que ce thesaurus ou cette classification soit re-travaillé sous forme d'une ontologie formelle. »<sup>18</sup>

### **1.2.3.2 Les grandes propriétés du modèle**

Le modèle SKOS organise un système de connaissance dans un schéma de concept Concept Scheme identifié par un URI. Ce système agrège des concepts eux mêmes identifiés par des URI. Parmi les nombreuses propriétés du modèle nous ne listons ici que ceux qui sont les plus en résonance avec la structuration d'un thesaurus et quelques unes des propriétés qui lui sont propres.

Ainsi SKOS décrit :

- La désignation de concept en langage naturel au moyen de littéraux : terme préférentiel - Pref Label ; Terme alternatif synonyme (qui n'existe pas dans le thesaurus) Alt label ; libellé caché - hidden label
- Les relations sémantiques entre concepts de type générique/spécifique : Broader Term / Narrower term
- Des notes d'explications : Scope note
- Des relations d'association : Related term
- Des définitions : definition
- Des synonymes : alt label
- Des propriétés de mise en correspondance de concepts issus de schémas différents : Mapping relation

---

<sup>18</sup> Traduction réalisée et mise à disposition par Thomas Francart accessible sur <http://www.sparna.fr/skos/SKOS-traduction-francais.html>

La conversion du Thésaurus Agrovoc<sup>19</sup> en Skos et sa publication par la FAO illustrent le partage, l'enrichissement par interconnexion de plusieurs vocabulaires permis par la modélisation Skos. A partir d'Agrovoc la FAO a organisé sa mise en relation avec les concepts de pas moins de six référentiels tels que Rameau, Eurovoc, Gemet. Il comprend 32 000 concepts disponibles en 20 langues et est publié par la FAO.

Enfin, comme le souligne Jean Delahousse [DELAHOUSSE, 2011], les domaines de connaissances modélisés dans Skos ne s'appuient pas exclusivement sur les vocabulaires contrôlés de type thésaurus et schéma de classifications. Ils peuvent être étendus à des terminologies locales exploitées dans SKOS par une extension Skos-XL gérant les terminologies.

## **1.3 Domaines de connaissances et accès aux contenus**

### **1.3.1 L'adaptation des moteurs de recherche**

S'il est permis de présenter les standards du web de données comme autant de nouveaux instruments disponibles pour les stratégies de valorisation de contenus, quels sont les solutions et avantages ou services qu'ils permettent en terme d'accès au contenu et d'optimisation de la recherche ? Ces services procèdent comme le rappelle Stéphane Pouyllau [POUYLLAU, 2011] des efforts accomplis par les concepteurs de moteur de recherche pour adapter leur traitement automatisé (indexation du texte intégral) aux modes sémantiques de structuration et de mise relation des données introduits par les standards cumulés URI et RDF.

La plate forme avancée de collecte, recherche et de diffusion de l'information scientifique en sciences humaines et sociale Isidore s'appuie sur les standards sémantiques. Selon Stéphane Pouyllau, ce cas illustre trois familles de services rendus possibles par l'enrichissement du moteur de recherche au moyen de données structurées :

---

<sup>19</sup> Le thésaurus Agrovoc <http://aims.fao.org/fr/agrovoc> et Aujourd'hui, AGROVOC est un schéma de concepts SKOS-XL et est disponible sous la forme d'un ensemble de données liées sur le Web de Données, aligné sur 13 autres systèmes d'organisation des connaissances multilingues en lien avec l'agriculture. Il est possible de télécharger AGROVOC ainsi que d'y accéder par l'intermédiaire de web services ou d'un SPARQL *endpoint*. qui couvre tous les domaines d'intérêt de l'Organisation des nations unies pour l'alimentation et l'agriculture FAO (la nutrition, l'agriculture, la pêche, la foresterie, l'environnement, etc.).

Rapprochement des informations en mettant au jour les relations entre publications scientifiques et diverses sources de données (archives, fonds documentaires, etc.) et diversification des modes de représentation de l'information ; la validation de l'information au moyen de leur mise en relation ; augmentation de la pertinence des résultats de recherche permise par l'indexation des ressources par des vocabulaires contrôlés.

### **1.3.2 Expansion sémantique et recherche à facettes**

Les traitements avancés des moteurs de recherche mettent à disposition des stratégies de valorisation et d'accès au contenu des solutions plus ou moins lourdes ou coûteuses. On n'en mentionnera que quelques unes permises par la structuration des vocabulaires et l'organisation des référentiels de connaissance en ontologie de domaine.

- L'expansion sémantique

« L'expansion sémantique peut être décrite comme la capacité des moteurs de recherche à étendre automatiquement les termes de la requête de l'utilisateur aux synonymes, acronymes, aux traductions ou aux concepts plus spécifiques ( auto-postage descendant / ascendant) [DELAHOUSSE, 2011]. Comme on l'a vu il est facilité par l'indexation initiale des ressources sur des vocabulaires métiers contenant ce type de structuration.

- La recherche à facette

Elle consiste dans la possibilité donnée à l'utilisateur de filtrer les résultats de sa recherche de documents et de naviguer au moyen d'axes ou de critères affichés comme autant de points de vues complémentaires sur fonds de ressources documentaires. « Une classification à facettes associe à chaque donnée de l'espace de recherche un certain nombre d'axes explicites de filtrage, par exemple des mots clés issus d'une analyse de texte, des métadonnées stockées dans une base de données, etc. »<sup>20</sup>

Une variante de recherche à facettes est décrite par Jean Delahousse comme « taxonomies dynamiques. » [DELAHOUSSE, 2011] Elles reposent sur une taxonomie de publication définie comme « l'ensemble des rubriques disponibles pour le filtrage des contenus d'un site.

---

<sup>20</sup> Les facettes se déploient le plus souvent sous forme de liste ( de métadonnées, de mots clés) dont le contenu peut être hiérarchisé. Titre, Date de publication, langues, mots clés etc. enrichis de données statistiques de contenus associés automatiquement mis à jour. [http://fr.wikipedia.org/wiki/Recherche\\_%C3%A0\\_facettes](http://fr.wikipedia.org/wiki/Recherche_%C3%A0_facettes)

» Les listes utilisées peuvent provenir de taxonomie de publications ou bien de listes créées automatiquement à partir d'extraction d'entités nommées<sup>21</sup> identifiées dans les contenus. Elles constituent une modalité avancée de l'anticipation éditoriale des besoins utilisateurs en permettant leur adaptation (introduction de nouvelles rubriques, modifications des contenus regroupés dans rubriques stables).

Les ontologies de domaine formulées en Skos par exemple autorisent la pré-configuration de taxonomies de publications. Les deux notions ontologie et taxonomies de publications renvoient à des objets de nature différente mais leur modélisation permet leur réutilisation. [DELAHOUSSE, p.182] « L'ensemble de ces objets constituent les instances d'une ontologie. Si le contenu des ressources terminologiques et des taxonomies est bien différent pour chaque domaine métier, la modélisation qui permet de les gérer est indépendante du domaine métier. ».

---

<sup>21</sup> « La reconnaissance d'entités nommées est une sous-tâche de l'activité d'extraction d'information dans des corpus documentaires. Elle consiste à rechercher des objets textuels (c'est-à-dire un mot, ou un groupe de mots) catégorisables dans des classes telles que noms de personnes, noms d'organisations ou d'entreprises, noms de lieux, quantités, distances, valeurs, dates, etc. »  
source [http://fr.wikipedia.org/wiki/Reconnaissance\\_d%27entit%C3%A9s\\_nomm%C3%A9es](http://fr.wikipedia.org/wiki/Reconnaissance_d%27entit%C3%A9s_nomm%C3%A9es)

# **Deuxième partie**

## **Le cadre de l'expérience**

## 2 L'élaboration du portail Plan4Learning de l'IIEPE

---

L'élaboration du portail thématique Plan4Learning au sein du centre de documentation de l'IIEPE délimite le cadre resserré de l'expérience qui a nourri le présent travail. Nous l'introduisons par un rappel plus large de son environnement institutionnel et une présentation de l'organisation et de l'activité du centre de documentation.

### 2.1 Une institution internationale : l'IIEPE

#### 2.1.1 Une autonomie relative au sein de l'Unesco<sup>22</sup>

L'IIEPE est fondé en 1963 comme institution internationale faisant partie intégrante de l'Unesco. Son mandat est de renforcer la capacité des pays à planifier et gérer leur système d'éducation. Son action est principalement tournée vers les pays en développement. L'institut a son siège basé à Paris et compte également un bureau à Buenos Aires (Argentine) et, depuis octobre 2013, à Dakar (Sénégal) suite à l'intégration du pôle de Dakar à l'IIEPE<sup>23</sup>.

Le directeur est nommé par le directeur général de l'Unesco, mais l'IIEPE a son propre conseil d'administration, ses statuts, ses locaux, son personnel et son budget.

Le conseil d'administration est l'organe exécutif. Il est composé de douze membres, choisis pour leur compétence et siégeant à titre personnel.

En tant qu'organe exécutif, il fixe la politique globale de l'IIEPE et décide de la nature de ses activités dans le cadre de la politique générale de l'Unesco. Il détermine également son programme de travail et ses objectifs.

#### 2.1.2 Une action coordonnée par objectif

Les objectifs stratégiques de l'IIEPE sont définis pour une période de quatre ans, dans une stratégie de moyen terme. Toute activité de l'Institut est réalisée et évaluée par rapport à

---

<sup>22</sup> Les statuts sont consultables sur le portail de l'IIEPE

<http://unesdoc.unesco.org/images/0015/001579/157990fo.pdf>

<sup>23</sup> Source : rapport d'activité de l'IIEPE 2012-2013 p.12 consultable

[http://www.iiep.unesco.org/fileadmin/user\\_upload/About\\_IIEP/pdf/IIEP\\_in\\_Action\\_2012-13\\_-\\_FR\\_-\\_web.pdf](http://www.iiep.unesco.org/fileadmin/user_upload/About_IIEP/pdf/IIEP_in_Action_2012-13_-_FR_-_web.pdf)

ses priorités stratégiques, les thématiques prioritaires, les actions à entreprendre ainsi que les ressources à leur affecter. Il arrête le programme de travail et le budget de l'Institut et s'assure de la mise en œuvre des activités prévues.

## **2.1.3 Missions et organisation**

### **2.1.3.1 Formation assistance et partage de connaissances**

L'activité de l'IIEP relève de missions de formation, d'assistance technique et de recherche et partage de connaissances expertes

Au titre de la formation destinée aux professionnels de l'éducation, le principal volet est le programme de formation approfondi ( PFA) Dispensés au siège, les cours du PFA préparent chaque année pendant six mois, une trentaine de stagiaires de différents pays aux responsabilités de décideur en matière de gestion et de planification de systèmes d'éducation. Les programmes spécialisés sur la planification du secteur de l'éducation (PSE), les universités d'été annuelles, des modules de formation intensive, des débats stratégiques complètent l'éventail des leviers de formation.

Les programmes d'assistance technique ciblent en particulier les gouvernants confrontés à des situations de crise.

Enfin, au titre de la Recherche et du partage de connaissances les interventions portent sur les diverses stratégies de développement des capacités à l'œuvre à différents échelons : ministères de l'Éducation, systèmes éducatifs, organismes internationaux. Ces travaux anticipent les évolutions et permettent à l'institut de formuler des propositions pour adapter en conséquence ses programmes et ses activités.

L'action du centre de documentation s'inscrit pleinement dans la mission de partage des connaissances.

### **2.1.3.2 Une politique active de partenariat**

Une politique active de partenariat et d'assistance conduit les experts à soutenir plusieurs réseaux d'établissements de formation et de recherche à la planification à l'échelle des pays membres. L'IIEP est à l'origine de réseaux comme le SACMEQ (Consortium de l'Afrique australe et orientale pour le pilotage de la qualité de l'éducation) et l'ANTRIEP (Réseau asiatique d'institutions de formation et de recherche en planification de l'éducation). Il assure également le secrétariat du GITE (Groupe international de travail sur l'éducation). »<sup>24</sup>

---

<sup>24</sup> source page site IIEP



### 2.1.3.3 Une organisation intégrée et multilingue

Pour conduire ses missions, l'IIPE s'appuie sur un effectif de 70 personnes<sup>25</sup> réparties entre unités de recherche, services administratif et informatique et, depuis janvier 2014, un service d'information regroupant la bibliothèque, les sites web et l'unité publication.

#### 2.1.3.3.1 Les unités de recherche

Les unités de recherche et d'assistance sont composées d'économistes, de sociologues, de statisticiens, de démographes, d'experts en technologie éducative, d'analystes des politiques et planificateurs. Ces spécialistes sont affectés à des thématiques. Ils oeuvrent à la production de connaissances et au transfert des savoirs au travers de formations et de programmes d'assistance. Le bilinguisme (anglais et français) est de règle dans le fonctionnement quotidien de toute l'organisation. Les formations sont dispensées en français et en anglais.

#### 2.1.3.3.3 Le service d'information

- L'unité publication

Elle participe activement à la valorisation et à la visibilité de l'action de l'IIPE au travers de deux missions principales :

La gestion du catalogue des publications de l'IIPE d'abord. Il s'agit d'un fonds de publications multilingue (fr, en, es) d'environ 1500 références portant sur les concepts, les méthodes, les instruments et techniques relatives à la planification et à la gestion de l'éducation. Elle assure la consultation de ses ressources proposées sous format papier et électronique au moyen d'un site internet multilingue (fr, en, es) dédié<sup>26</sup>.

En 2014, les références disponibles en ligne sont regroupées dans six thèmes : compétences, emploi et développement social (74)<sup>27</sup> ; égalité de l'éducation pour les groupes cibles (135) ; gestion et administrations des systèmes éducatifs (159) ; niveaux, contenus et modalités éducatives (342) ; planification de l'éducation (207) ; économie de

---

<sup>25</sup> Chiffres de 2013

<sup>26</sup> <http://publications.iiep.unesco.org/>

<sup>27</sup> Les nombres entre parenthèses correspondent aux références disponibles sur le site des publications à la date de consultation soit le 31 octobre 2014.

l'éducation (119). En amont de la diffusion, l'unité de publication gère également la chaîne de fabrication des publications de l'IPE.

Une mission de communication large ensuite qui englobe la gestion du site web de l'IPE. Elle est également responsable de la publication et de la diffusion sur support électronique et papier de la lettre d'information trimestrielle de l'organisation.

## **2.2 Le centre de documentation**

Créée en 1963, le centre de documentation est pleinement investi dans la mission de partage de connaissance de l'IPE. Sa vocation première est de collecter, gérer et diffuser l'ensemble de la littérature spécialisée sur la planification de l'éducation.

### **2.2.1 Un fonds spécialisé et des publics experts**

Le centre collecte et diffuse l'essentiel de la littérature spécialisée traitant de la planification de l'éducation et son contexte économique et social. Il est le lieu de référence pour la documentation des systèmes éducatifs des pays en développement.

Le catalogue de la bibliothèque contient à ce jour

- 33 000 ouvrages documents et rapports
- 400 titres de périodiques

La documentation comprend ainsi une sélection de publications de l'UNESCO, les travaux et rapports émanant d'organismes internationaux tels que la Banque mondiale, l'OCDE, l'UNICEF, Union Européenne, etc.

La forte diversification du fonds tient à la présence à côté des ouvrages, de nombreux rapports et travaux de recherche, documents officiels publics ( plans d'éducation de pays membres), de comptes rendus d'activité de l'IPE ( conseil d'administration). La conservation et la gestion des documents internes à l'IPE relève de la compétence du centre de documentation. Le fonds est multilingue. Les publications sont disponibles en anglais (à 65%), en français, en espagnol et pour une part en arabe, en portugais et autres langues.

Pour l'essentiel, les utilisateurs du centre sont :

- des spécialistes de la planification et de la gestion de l'éducation (formateurs, chercheurs) ;
- des stagiaires en formation à l'IPE ou à l'UNESCO ;
- des étudiants de 3ème cycle en sciences de l'éducation ou en développement.

La plupart des lecteurs sont bilingues voire trilingues.

Le centre de ressources a toutefois pris la mesure de l'élargissement de son rayonnement auprès de publics élargis constaté par l'évolution des consultations de son catalogue en ligne et l'essor des visites de ces portails thématiques.

### **2.2.2 L'activité du centre de documentation**

### **2.2.2.1 Une équipe restreinte et expérimentée**

L'animation et le fonctionnement sont assurés par la directrice du centre de documentation.

Celle-ci est entourée de trois documentalistes spécialisées :

- Une documentaliste chargée de l'administration et de la gestion des bases de données
- Une documentaliste chargée des acquisitions et de la veille
- Une documentaliste responsable de l'animation du service d'échange d'information dédié aux problématiques du VIH de la santé en lien avec l'éducation<sup>28</sup>

### **2.2.2.2 Une exploitation diversifiée des ressources documentaires**

#### 2.2.2.2.1 La gestion et la maintenance de bases de données spécialisées

L'administration et la maintenance du fonds sont assurées au moyen de bases de données distinctes. Deux types de bases peuvent être distinguées :

Les bases de données en accès libre ( catalogue de la bibliothèque, bases thématiques) et les bases de données relatives à l'activité interne de l'IPE dont l'accès est réservé

---

<sup>28</sup> Ce portail remodelé sur CMS *Drupal* pendant l'été 2014 est accessible à l'adresse suivante : <http://hivhealthclearinghouse.unesco.org/>

Nom de la base	Nature	Volume	Exploitation en <i>front office</i>
EPIDOC	Notices bibliographiques en trois langues (en, es, fr) du catalogue général de la bibliothèque	33 000 notices + collection titres de périodiques	Accessible par le catalogue en ligne EPIDOC. sur le site de l'IIE
ETICO	Entrées sur le thème de l'éthique et la corruption	Plus de 450 ouvrages, projets et normes.	Portail thématique public
PLANIPOLIS	Plans et politiques d'éducation des pays rendus publics.	Plus de 2100 plans issus de 194 pays	Portail thématique Planipolis public
HIV Health Clearinghouse	Fonds multilingue dédié au sida, à la santé et à l'éducation	7000 références	Portail thématique public

Tableau 1 Les base de données du centre de documentation de l'IIE

GB (Governing Board)	Fiches projets et activité (GB) établies dans le cadre des plans stratégiques à moyen terme	En ligne (accès réservé)
INTEPI	Documents administratifs Produits de l'IPE rapports de mission, rapports d'évaluation, supports de présentations des chercheurs, etc.) contributions à des séminaires	En ligne (accès réservé)

Tableau 2 Les bases de données dédiées à l'activité interne de l'IPE

L'administration et l'exploitation de ces bases de données se heurtent depuis plusieurs années aux limites du logiciel gratuit de GBD mis à disposition par l'Unesco CDS ISIS. (arrêt du développement, ancienneté de la mise à jour) Plusieurs travaux ont étudié les modalités de son remplacement.<sup>29</sup>

#### 2.2.2.2.2 L'exploitation et la mise à disposition du fonds

Ces activités s'appuient sur un large éventail d'instruments :

- Consultation du catalogue de la bibliothèque en ligne sur le site web de l'Institut
- Interface d'accès public ou restreint aux bases de données documentaires
  - Conception et alimentation de portail thématiques : Planipolis, HIV-Health
  - Clearinghouse, ETICO sur l'éthique et la corruption, Capacity Development, Plan with Youth
- Production de bibliographies, contributions régulière à la newsletter de l'IPE.

#### 2.2.2.3 Une fonction documentaire étendue

L'activité des documentalistes couvre au total un périmètre d'intervention étendu qui comprend :

- Les grandes fonctions de la chaîne documentaire

<sup>29</sup> Un point très complet mais déjà daté a été réalisé dans [Sobral-Boyer, 2010].

La veille automatisée<sup>30</sup> relayée par des outils collaboratifs (Netvibes, outils de partage de signets, etc.) la collecte, le catalogage, la structuration de l'information au moyen de langages documentaires (thésaurus de l'Unesco, plan de classement, etc.), la recherche, l'accueil du public et la gestion des prêts de la bibliothèque, la conservation des documents<sup>31</sup>.

➤ Un support systématique aux unités fonctionnelles

Celui-ci comprend :

L'appui aux chercheurs au travers de recherches bibliographiques de veille thématique, et de personnalisée.

L'assistance des formateurs avec lesquels ils collaborent pour la mise à jour des bibliographies attachées aux supports de cours ; la mise à disposition des stagiaires des documents de référence sous format électronique.

L'aide et l'accompagnement des stagiaires que les documentalistes orientent dans les outils et les ressources du centre documentaire.

➤ Une assistance technique « hors les murs ».

Dans le cadre de partenariats établis au titre de l'appui au renforcement des capacités des Etats avec certains pays comme le Maroc et le Cambodge les documentalistes assistent techniquement certains pays dans la mise en place et la gestion de leur centre de ressources. Par ailleurs, le centre de documentation est aussi en charge de l'animation et du soutien (par l'envoi de livres et documents spécialisés notamment) d'un réseau international de 130 bibliothèques dépositaires qui, pour des raisons de moyens ne peuvent acquérir la littérature spécialisée.

➤ Une forte implication dans les projets transversaux

Recherches bibliographiques sur le rapport GMR Global monitoring Report, rapport mondial du suivi de l'éducation pour tous<sup>32</sup> ;

Mise à jour des bibliographies multilingues pour les supports des universités d'été

---

<sup>30</sup> Sur outil KBcrawl

<sup>31</sup> Pour les documents rares, documents internes et le fonds exclusif

<sup>32</sup> Global monitoring Report, rapport mondial du suivi de l'Education pour tous établi annuellement par l'Unesco dans la lignée du programme Education pour tous lancé à Dakar en 2000, voir <http://www.unesco.org/new/fr/education/themes/leading-the-international-agenda/efareport/the-report-and-efa/>).

## **2.3 La qualité des apprentissages : une thématique prioritaire pour l'IPE**

### **2.3.1 Un thème de stratégie de moyen terme**

La mise en place d'un portail thématique dédié à la qualité des apprentissages figure parmi les actions envisagées dans la stratégie à moyen terme 2014-2017<sup>33</sup>

Dans ce document, l'IPE programme ses thématiques prioritaires et consigne les axes d'actions destinés à coordonner et mobiliser les différentes unités opérationnelles de l'organisation. La qualité des apprentissages y est posée comme thématique prioritaire n°2. Celle-ci prend acte des défis que posent l'aggravation des inégalités et la persistance de la pauvreté, l'adaptation aux changements démographiques et surtout la modification de la demande de compétences, et ce, en dépit des progrès accomplis dans le cadre des programmes Education pour tous et de l'objectif n° 2 du Millénaire. Dans ce cadre, les actions envisagées en matière de soutien, de recherche et de partage de connaissance sont décrites comme suit : *« Dans la poursuite de sa tâche, l'IPE fournira aux responsables politiques et aux planificateurs de nouvelles analyses et synthèses des facteurs et des stratégies qui permettent d'améliorer les acquis de l'apprentissage. L'Institut aidera également les ministères de l'éducation à développer leur capacité d'interpréter les données nationales sur l'apprentissage et d'élaborer sur cette base des politiques et des plans pour améliorer l'apprentissage. Un nouveau portail en ligne, « Planifier l'apprentissage », ainsi que les activités de formation, d'assistance technique et de conseils stratégiques de l'IPE mettront l'accent sur cette priorité. »*<sup>34</sup>

---

<sup>33</sup> Stratégie à moyen terme SMT 2014-2017 consultable [en ligne]

[http://www.iiep.unesco.org/fileadmin/user\\_upload/About\\_IIEP/pdf/MTS\\_FR\\_2014\\_2017.pdf](http://www.iiep.unesco.org/fileadmin/user_upload/About_IIEP/pdf/MTS_FR_2014_2017.pdf)

<sup>34</sup> SMT, op cit p. 10



## 2.3.2 Une feuille de route pour Plan4Learning

Au sein de l'institution, et plus généralement dans la sphère des travaux menés dans le domaine à l'échelle internationale, l'accent nouvellement placé sur la qualité des apprentissages marque le renouvellement des programmes et des actions longterm focalisés sur le thème de l'accès à l'éducation.

Une note de recherche<sup>35</sup> restitue clairement les contours et les enjeux de cette priorité. Ce faisant, elle esquisse la feuille de route pour la réalisation du futur portail par lequel l'IIEP entend apporter une contribution concrète, en esquisant son ambition générale, ses objectifs, son premier périmètre thématique.

### 2.3.2.1 Ambition générale et objectif du portail

- Pallier aux carences constatées en matière de soutien et d'analyses centrées sur la qualité des apprentissages pour les pays du sud en dépit d'une demande montante des intéressés.
- Comblent l'absence d'action coordonnée et de plateforme publique centralisant l'information et l'évaluation des actions d'apprentissage, les débats, les initiatives, les retours d'expériences, les choix d'évaluation et toutes recherches idoines menées à l'échelle du monde.<sup>36</sup>

Un objectif ensuite :

Le portail doit fonctionner comme une base documentée pour la compréhension des besoins, des politiques et des stratégies visant à l'amélioration des apprentissages.

### 2.3.2.2 Périmètre et contenu

L'apprentissage y est décrit comme l'ensemble des compétences cognitives en matière de calcul, de lecture et d'écriture ainsi que l'élargissement de ce savoir faire en vue des apprentissages approfondis. Le périmètre scolaire envisagé s'étend de la prime éducation au primaire et au secondaire. Les expériences et résultats d'apprentissage du post-secondaire et de l'éducation non formelle ne sont pas couvertes. Un moindre accent enfin doit être porté aux questions d'accès à l'éducation et de fidélisation scolaire (retention en anglais).

---

<sup>35</sup> « Why quality and learning are important for post-2015 »

<http://www.iiep.unesco.org/fr/recherche/defis-et-nouvelles-tendances/cours-organises-a-liipe/quality-and-learning.html> [en ligne]

<sup>36</sup> « At present there is no public space that brings together the information and evaluation of learning interventions, debates, implementation experiences, assessment options, and relevant research from around the world. »

Dans ce cadre une première ébauche de contenu est formulée :

- Il s'agit d'un portail public qui offre un accès facilité à une information dispersée et peu accessible au moyen de contenus synthétiques, équilibrés, fiables et abordables.
  - Rapporter des initiatives conduites en veillant aux attentes spécifiques émanant des planificateurs, des décideurs politiques des acteurs de la société civile et des financeurs.
  - Donner des outils pour comprendre des données complexes et techniques.
  - Proposer des analyses de contenu et de débats spécifiques traitant des résultats de l'apprentissage : - Quelles politiques pour améliorer l'apprentissage ? Comment intégrer un volet apprentissage dans un processus de planification de l'éducation ? Comment un ministre de l'Education choisit-il entre les différents outils et approches disponibles ? etc.
- La spécification plus avancée des besoins formera la première étape de préparation opérationnelle du projet de portail Plan4Learning.

## **2.4 Le projet Plan4Learning dans son contexte documentaire**

L'élaboration du portail s'inscrit dans un moment particulier de la politique documentaire du centre de documentation. Nous entendons par-là non plus les grands principes de son fonctionnement tels qu'on les a rappelé plus haut, mais plus spécifiquement les problématiques et les réflexions menées sur l'état de son dispositif documentaire, de ses outils, de l'évolution de ses publics, et les pistes très concrètes envisagées pour le faire évoluer. Il s'agit autant d'enseignements issus de recherches menées au cours des quatre dernières années que de prestations d'audit et d'interventions commandés. Par leur existence même, ces travaux, attestent de la grande ouverture du centre de documentation aux innovations technologiques et à une veille réflexive sur ses propres pratiques. Deux « chantiers » ou objets de réflexions méritent d'être mentionnés. Un premier porte sur la prise en compte avancée de l'évolution des publics et de leur mode d'accès aux catalogues. Un second a trait à l'exploration des conditions et opportunités d'un nouveau traitement des notices documentaires propice au décloisonnement des bases de données et à l'ouverture du centre à l'interopérabilité et au partage permis par les applications du web sémantique.<sup>37</sup>

### **2.4.1 Visibilité sur le web et ouverture à de nouveaux publics**

Une première recherche significative et contemporaine du projet a porté, sur la stratégie de numérisation à privilégier s'agissant des ressources du centre et de leur mise à disposition et du suivi de leur consultation en ligne [Polewka,]. Il s'agissait de prendre la mesure des opportunités offertes par la numérisation pour : s'adapter à la généralisation des consultations des fonds en ligne, optimiser la visibilité de ses ressources disponibles ; articuler la satisfaction des besoins d'un public proche local et le ciblage du public potentiellement plus important des utilisateurs distants et des nouveaux profils d'utilisateurs moins bien connus (chercheurs, décideurs politiques, journalistes, etc.). Le succès du portail Planipolis qui, depuis 2007 permet la consultation des plans officiels et publics des gouvernements des pays membres attire près de 100 000 visiteurs unique par mois a

---

<sup>37</sup> Au titre du contexte documentaire on a déjà mentionné les réflexions engagées pour la migration visant à remplacer l'outil de GBD gratuit CDS Isis analysées dans [Sobral Boyer, 2010]

grandement encouragé cette réflexion et l'analyse plus générale de l'évolution des publics et de leur mode d'accès aux ressources du centre. Le centre de documentation fonctionne déjà sur un modèle hybride articulant bibliothèque virtuelle et catalogue physique. Ces travaux laissent entrevoir qu'une plus grande visibilité du catalogue sur le web servirait autant l'optimisation des consultations par les publics proches et locaux qu'elle profiterait à des publics élargis. L'engagement dans cette voie est il aussi rappelé requiert un travail préalable sur le décloisonnement des ressources, l'adoption de solutions logicielles pour la gestion des bases de données pleinement compatible avec le web, l'introduction de standards de description des ressources permettant l'interopérabilité et leur exposition sur le web. [Polewka, p.83].

### **2.4.2 Une exploration du nouveau format de description bibliographique**

De façon tout aussi prospective, le centre s'est très largement prêté à une analyse de la pertinence et de la faisabilité de l'inscription de ces notices catalographiques dans le modèle FRBR. La particularité de ce modèle est d'introduire dans la structuration des notices et de leur description le point de vue de l'utilisateur. Il consiste très schématiquement à réordonner les composantes d'une notice et à les relier à d'autres selon un ordonnancement logique. Pour l'IPE il s'agissait d'envisager la possibilité de donner à l'utilisateur le moyen « d'accéder de manière groupée à des informations interconnectées jusqu'alors éparpillées dans différentes bases de données. »[Bergonnier Signoles p. 89]

Deux types d'avantages corrélés sont soulignés :

En terme de valorisation des ressources d'une part, l'exploration a mis l'accent sur la démultiplication des points d'accès aux ressources permise par ce modèle et la facilité de leur identification.

En terme d'alignement et d'ouverture et de partage des données, l'harmonisation des données de nature hétérogène localisées dans différentes bases autorise non seulement l'alignement de celles-ci au sein de l'IPE, mais aussi leur rapprochement avec les données provenant d'autres institutions extérieures. L'information des notices bibliographiques devenant exploitable par les applications du web de données.

## 2.5 Le projet de portail en phase opérationnelle

### 2.5.1 Pilotage et équipe projet.

Sous l'autorité de la directrice de l'IPE, la conduite du projet est assurée par la directrice du centre de documentation de l'IPE, Mme Asunción Valderrama en lien avec l'unité Publication qui gère le site institutionnel de l'IPE.

Au sein du service de documentation l'équipe projet s'appuie sur la collaboration d'une documentaliste spécialisée dans l'interrogation et la maintenance des bases de données.

A l'instar des différents projets transversaux auxquels le centre a coutume de collaborer, le portail Plan4learning ne peut mobiliser une équipe même réduite à plein temps.

Pour la finalisation des spécifications fonctionnelles et techniques, le projet fait intervenir un chef de projet. En qualité de prestataire maîtrise d'ouvrage/maitrise d'œuvre celui-ci assure la liaison entre les demandes et besoins formulés par les équipes métiers et les différents prestataires extérieurs. Il conduit également les formations à la prise en main des outils logiciels livrés (hors CMS).

Dans les limites du budget alloué, l'équipe projet est depuis la phase préparatoire ponctuellement renforcée par des personnels contractuels ou stagiaires pour l'alimentation en contenu éditorial et en notices, pour la conduite de tests et l'enrichissement des ressources documentaires (vocabulaires, notices, etc.).

Le projet est financé pour l'essentiel par la contribution volontaire de la fondation William and Flora Hewlett.<sup>38</sup>

#### 2.5.1.1 Les prestataires extérieurs

Deux catégories de prestataires ont été retenues pour des interventions séquencées en deux phases successives.

En phase 1 l'appel d'offre a abouti à la sélection d'un consortium composé de la société Mondeca<sup>39</sup>, éditrice du logiciel de gestion de vocabulaires et thésaurus ITM (Intelligent Topic manager), de l'Institut de recherche et d'innovation (IRI) pour la mise en place d'un outil d'édition et d'annotation des notices bibliographiques *Open Source* et la société de consulting Sparna pour la conduite du projet.

---

<sup>38</sup> Le détail du soutien étalé sur quatre ans est public et accessible [en ligne] sur la cartographie des contributions de la fondation. <https://maps.foundationcenter.org/>

<sup>39</sup> <http://www.mondeca.com/fr/>

Pour l'intervention en phase 2 l'agence de web design Le Hub a été retenue pour l'élaboration du portail final dans un CMS *open source* (Symfony).

#### **2.5.5.2 La répartition des tâches entre prestataires**

Le tableau ci-dessous récapitule les missions de chacun des prestataires.

Phases du projet	Composantes	Prestataires			
		Société Mondeca	IRI	Sparna	Agence le HUB
<u>Phase 1</u>  Recherche et accès aux ressources	<ul style="list-style-type: none"> <li>• Intégration du moteur de recherche SolR</li> <li>• Livraison d'un portail d'exploitation</li> <li>• Une ontologie OWL</li> <li>• Outil de gestion des vocabulaires (ITM)</li> <li>• Outil de gestion des notices et annotations</li> </ul>	X		X	
<u>Phase 2</u>  Développement du portail web public « Plan4Learning»	<ul style="list-style-type: none"> <li>• Intégration du portail d'exploitation</li> <li>• Développement du portail web final sur outil de gestion de contenu CMS (Symfony) « Plan4learning »</li> </ul>	(X)		X	X

Tableau 3 Répartition des interventions des prestataires extérieurs du projet

## 2.5.2 Phases d'avancement du projet

Préparé très en amont au cours de l'année 2012, le projet entre dans sa phase opérationnelle avec l'appel d'offre pour la réalisation en phase 1 en novembre 2012.

La sélection du premier groupe de prestataire intervient avril 2013. La mise en service du portail est prévue pour la fin de l'année 2014 ou le début 2015.

Nous ne mentionnons ici que les phases relatives à la sélection et aux livrables attendus des prestataires extérieurs n°1 n°2 et n° 3.

Ce calendrier n'intègre donc pas les phases d'avant projet ni les séquences de réception de tests et de validation. Il n'intègre pas non plus le temps consacré par l'équipe interne à l'alimentation des contenus éditoriaux et à l'enrichissement des vocabulaires.

### **2.5.2.1 Phase 1. Moteur de recherche et accès aux ressources documentaires**

- **Objectif** : Intégration du moteur de recherche sémantique donnant accès aux notices documentaires / modélisation des vocabulaires dans référentiel sémantique.
- Rappel du travail préparatoire : étude d'opportunité, analyse des besoins. L'exploration préalable du fonds documentaire et la définition des besoins en matière de recherche, de navigation, identification et préparation des vocabulaires documentaires pertinents de leur nature et de leur format d'exploitation.
- **Année 2012**  
Novembre 2012 : Appel d'offre pour le choix du prestataire numéro 1
- **Année 2013**  
Mars 2013 : Sélection de l'offre  
Septembre 2013 : Début du déploiement
- **Année 2014**  
Janvier-Mars : Fin du déploiement et tests

### **2.5.2.2 Phase 2. Communication et mobilisation des communautés**

- Objectif : Développement et la livraison du portail final Plan4Learning
- Phase préparatoire : définition du traitement de l'information, structuration du site, spécifications techniques versées dans le cahier des charges phase 2<sup>40</sup>.

#### **Année 2014**

- Janvier : appel à proposition
- Février : réception des offres
- Mars : choix du prestataire n°2
- Mai : développement du prototype
- Juin : premiers tests et ajustements
- Juillet : fin du développement

---

<sup>40</sup> source : document interne



→ Fin juillet : recette et déploiement

**Mise en service prévue : Début 2015**

**Figure :** Capture d'écran de la page d'accueil du portail final Plan4Learning en phase de construction en juillet 2014 ( mise en service prévue début 2015).



## **2.5.3 La précision des objectifs**

### **2.5.3.1 L'objet du portail**

Plan4Learning est un portail Web de documentation et d'information sur la qualité de l'éducation. Son ambition est de permettre aux acteurs de l'éducation de mettre en oeuvre des politiques et des stratégies éducatives centrées sur l'amélioration de la qualité des acquis éducatifs et visant à ce que les systèmes éducatifs répondent de manière adéquate aux besoins de la société.

L'objectif principal vise à la mise à disposition, à l'accès et à l'utilisation de l'information et des objets documentaires associés à cette thématique<sup>41</sup>.

### **2.5.3.2 Les publics cibles**

La liste des publics cibles intègre des profils d'experts tout en élargissement le spectre à des utilisateurs qui ne sont pas forcément des professionnels investis de façon opérationnelle dans des missions de gestion et de planification de l'éducation. Sont identifiés :

- Les planificateurs
- Les cadres des ministères de l'éducation,
- Les décideurs politiques
- Les chercheurs et universitaires
- Les organisations issues de la société civile (ONG)
- Les journalistes

### **2.5.3.3 Un portail multilingue**

Le choix du multilinguisme fait d'emblée droit à la vocation internationale et multiculturelle de l'Unesco. Le portail est proposé en trois langues : le français, l'anglais et l'espagnol. Les versions et traductions dans ces trois langues sont prévues dès la phase de lancement pour la navigation, la consultation et le renseignement des notices documentaires mises à disposition. Seules les contributions postées sur le forum de discussion du portail ne sont

---

<sup>41</sup> Présentation du projet par Madame Asunción Valderrama, directrice du centre de documentation à la réunion des bibliothèques des nations unies le 18 septembre 2013. <http://www.educationinnovations.org/content/international-institute-educational-planning-iiiep> [en ligne]

pas traduites. L'introduction d'autres langues pouvant faire l'objet de développements ultérieurs.

#### **2.5.3.4 Un périmètre documentaire et éditorial resserré : La qualité des apprentissages<sup>42</sup>**

Le projet s'inscrit pleinement dans la thématique définie dans la stratégie à moyen terme et cible des ressources et des contenus éditoriaux centrés sur les apprentissages et la qualité de l'éducation.

Ce périmètre a donné lieu à un travail d'identification des ressources dans les bases de données et dans les vocabulaires documentaires servant à l'indexation et à la recherche des notices bibliographiques. Il a permis dans un premier temps de mettre à disposition des prestataires près de 3000 notices bibliographiques issues de la base de données EPIDOC. Un travail ultérieur d'homogénéisation a conduit à réduire le volume de lancement à 1000 notices en privilégiant notamment les notices permettant la mise à disposition des documents en texte intégral ou au minimum à leur résumé.

La thématique de la qualité des apprentissages circonscrit pareillement les sujets traités dans la partie éditoriale du site (*topics*, analyse de controverses, forums, synthèse du forum, fil d'actualités, agenda). Elle intègre enfin la publication de données, sur la recherche, les programmes et les débats relatifs à l'évaluation des systèmes éducatifs (méthodologie, pertinence des indicateurs d'évaluation, etc.) à l'échelle nationale, régionale et internationale.

#### **2.5.3.5 Une valeur ajoutée tirée de l'emploi de vocabulaires structurés**

Le portail doit tirer une grande partie de sa valeur ajoutée de l'exploitation pertinente de plusieurs langages documentaires.<sup>43</sup> Il s'agit des vocabulaires structurés multilingues ( fr, en, es,) destinés à l'indexation et à la recherche documentaire ( thésaurus de l'Unesco, Plan de classement du catalogue de l'IIEP), d'un glossaire spécialisé dans la gestion et la planification de l'éducation et de listes d'autorités (de pays, de projets, d'institutions.)

Le projet s'appuie sur la gestion de ces terminologies (description, mise à jour, enrichissement, alimentation) sur leur modélisation dans un référentiel sémantique ainsi que sur leur indexation dans un moteur de recherche sémantique permis par des exports au format SKOS. Les utilisateurs du site doivent retirer de ce dispositif au moins trois familles

---

<sup>42</sup> Le choix du nom à donner au portail s'est dans un premier temps porté sur « *Improving learning Outcomes* »

<sup>43</sup> L'objectif rappelé dans cahier des charge phase 2 en introduction des spécifications fonctionnelles et techniques en matière de navigation enrichie entre les contenus et les ressources. (Source : cahier des charge phase 2. document interne)

d'avantages ou de services sur lesquels nous reviendrons plus en détail (cf. Infra exploitation des vocabulaire et services). Ils délimitent un objectif étendu en matière de valorisation des ressources en terme de visibilité, de facilité d'accès, de sélection et d'options de découverte.

#### **2.5.3.6 Accès simplifié et navigation enrichie**

Il s'agit de répondre au besoin exprimé de faciliter l'interrogation de la base de données en langage naturel et dans plusieurs langues / Interrogation multilingue ; auto-complétion ; interrogation par mots clés, filtres par facettes, filtre par cartographie.).

Une navigation optimisée est par ailleurs envisagée pour faciliter les passages :

Entre les différentes ressources mises à disposition (liens à partir de mots clés, d'éléments descriptifs tels que nom de pays, de projet, d'auteurs, etc.) ;

Entre les éléments des vocabulaires (exemple : navigation à l'intérieur du glossaire par termes associés) permettra d'enclencher recherche toutes les notices bibliographiques contenant le même terme) ;

Entre les contenus éditoriaux et les ressources documentaires (annotation des contenus éditoriaux par les langages d'indexation des notices documentaires).

## **2.6 Solution logicielle et modélisation des vocabulaires**

Les objectifs de valorisation de ressources formulés dans le projet recouvrent des fonctions avancées en matière d'interrogation multilingue, de recherche en langage naturel, de navigation, de rebonds dans les ressources documentaires et de découvertes. Elles sont permises par le déploiement d'une solution logicielle ad hoc.

L'exposé que j'en donne aide prioritairement à la compréhension des grands principes d'articulation de ces composantes. Il se nourrit d'exemples issus des vocabulaires et de notices exploités dans le cadre du projet. Il pointe enfin en tant que nécessaire les principales limitations données contractuellement aux développements des outils pour coller à une description réaliste du dispositif et non à une énumération abstraite de potentialités.

## 2.6.1 Le dispositif logiciel d'ensemble<sup>44</sup>

Comme l'illustre la figure [figure 2] ci-après, la solution logicielle retenue met en relation trois blocs fonctionnels et un portail d'exploitation des flux du moteur de recherche soient :

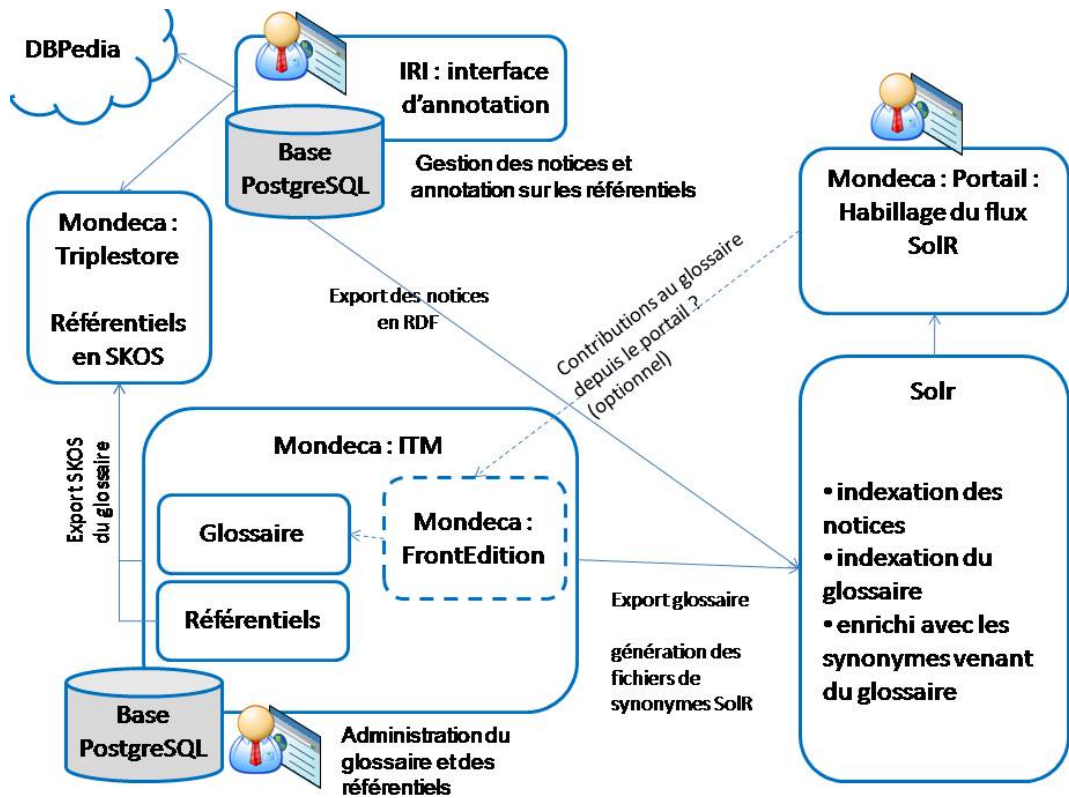
- Un outil de gestion et d'indexation des notices documentaire développé par l'IRI ;
- Un outil de gestion et de maintien du référentiel et des vocabulaires - Intelligent Topics Manager (ITM) (Mondeca) ;
- Ces deux outils de *back office* alimentent le moteur de recherche *open source* SolR ;
- Un portail d'exploitation des flux du moteur de recherche SolR.

Formellement, ce portail d'exploitation est le résultat d'un habillage au moyen d'une feuille de style (XSLT) du flux de requête (en langage XML) issu du moteur de recherche SolR. Cet habillage est destiné à permettre l'affichage dans une page web (html) de ce flux mis en forme. Dans sa version finale le portail d'exploitation offre les fonctions d'accès, de recherche et de filtres (facettes) de navigation et de rebonds dans les notices d'une part et dans le glossaire d'autre part.

Le schéma ne fait pas apparaître l'outil CMS *Open Source* (Symfony) dans lequel sont structurés les contenus et gérés l'affichage des contenus du portail final. Comme on l'a indiqué le développement du CMS est réalisé par second prestataire. Il intervient en phase avancée du projet après l'installation de la solution présentée ici. Il doit intégrer le portail d'exploitation issu de SolR et le soumettre à un habillage adapté à la charte graphique du portail final.

---

<sup>44</sup> La description des composantes logicielles de la solution retenue s'appuie pour partie sur la documentation interne au projet : offre commerciale des prestataires, cahier des charges phase 2. Je remercie l'équipe projet de m'avoir donné accès à ces documents de travail. Il va de soi que je suis seul comptable des éventuels oublis ou maladroresses dont serait porteuse la lecture très synthétique que j'en restitue ici.



**Schema v4**

Figure 2 Schéma d'ensemble de la solution logicielle (hors CMS du portail final), source : présentation Sparna.

## 2.6.2 L'édition et l'annotation des notices documentaires

### 2.6.2.1 Maintenance des notices documentaires et échanges de données

Ces fonctions sont assurées par l'outil d'annotation et de gestion des notices développé par l'Institut de recherche et d'innovation IRI. Il s'agit d'un ensemble logiciel *open source* dont la mise au point participe des recherches de l'IRI sur l'annotation sémantique.

Au sein de la solution mise à disposition ce bloc fonctionnel exerce deux rôles :

- Pour l'alimentation et la gestion des notices tout d'abord, l'outil déploie un environnement pour :

importer des notices existantes et maintenues dans les bases de données de l'IPE ;  
créer de nouvelles notices documentaires ; gérer les notices par des opérations d'enrichissement, de modifications, de suppression.

- Pour les échanges de données au sein du dispositif logiciel

- Il importe les métadonnées descriptives issues du référentiel (interrogation quotidienne de l'entrepôt RDF) pour proposer des annotations dans les champs contrôlés des notices documentaires. Par exemple les champs Theme ; Mots-clé; pays, institutions pourront être renseignés dans l'outil d'édition sur des champs contrôlés alimentés par les vocabulaires correspondants, en l'espèce : le plan de classement de l'IPE, le thésaurus de l'Unesco, la liste normalisée des institutions, etc.

- Il exporte les notices au format RDF pour permettre leur indexation dans le moteur de recherche SolR et leur diffusion dans le portail d'exploitation (l'exportation s'effectue toutes les nuits).

### 2.6.2.2 La structuration des notices

Elle est composée à la fois de champs en texte libre et de champs contrôlés par les valeurs du référentiel. Pour le projet, trente cinq champs ont été paramétrés.

Ils permettent la reprise des descriptions en vigueur et l'introduction des champs spécialement conçus pour suggérer des valeurs à partir des métadonnées extraites du gestionnaire de vocabulaires.



Les avantages attendus de la solution tiennent à un environnement unique et homogène pour les fonctions de maintien des notices et à un élargissement du périmètre de l'indexation et de communication.

#### 2.6.2.2.1 Rappel des règles et modalités de description en vigueur

Les notices concernées à ce stade du projet sont extraites de la base de données catalographique EPIDOC. Après une première importation de 3000 unités, la sélection a réduit le nombre de notices exploitées dans l'outil d'annotation à 1000 unités.

Au sein de l'IIEP les codes et standards de descriptions sont définis par le Format commun de communication<sup>45</sup> et de la norme MARC ISO 2709 pour l'échange informatique des notices bibliographiques.

Les notices bibliographiques sont indexées au moyen du thésaurus de l'UNESCO et du plan de classement du centre de documentation. Des ajouts possibles de mots clefs spécifiques à l'IIEP sont prévus pour la recherche. Les nouveaux mots clés (candidats descripteurs du Thésaurus) sont validés par le gestionnaire du thésaurus au siège de l'Unesco.

Les outils logiciels employés habituellement sont deux déclinaisons du logiciel gratuit CDS ISIS pour la structuration et la gestion multilingue.

La structuration se fait dans le logiciel CDS ISIS for windows (Winisis). Les notices sont saisies dans une grille de catalogage en anglais. La grille prévoit la possibilité de saisir les titres en d'autres langues lorsque le document est publié en plusieurs langues. Toolisis est l'utilitaire qui fait des équivalences entre vocabulaires à partir de table de correspondance entre le français, l'espagnol l'anglais pour l'accès à la notice par le public en trois langues.

---

<sup>45</sup> Source : version de 1992 de la matrice standard des métadonnées du CCF de l'Unesco, p.12 <http://unesdoc.unesco.org/images/0009/000924/092452eb.pdf>) de l'Unesco . Ce document prévoit la description de tous les objets Institutions, personnes, projets, livres, périodiques, rapports thèses, cartographies, licences, standards de l'Unesco. Les normes des entités bibliographiques correspondent aux zones de l'ISBD (international Standards, Bibliographic Description).

#### 2.6.2.2.2 Limites du développement

L'outil d'annotation comporte quelques limitations qui concernent pour l'essentiel les fonctions de structuration des notices et les règles d'administration.

Il ne stocke pas les documents physiques eux mêmes ( fichier image, fichier word, pdf texte, etc. ) Il sont déposés et maintenus dans un autre système et accessibles au moyen d'une url.

L'introduction de nouveaux champs de description requiert un développement spécifique

Il ne propose pas de fonction de traduction des notices. Une version traduite d'une notice apparaîtra comme une nouvelle notice sans lien avec sa version initiale.

Il ne dispose pas d'organisation de *workflow* pour la validation /modération des changements dans une notice avant édition et publication

De même il n'existe pas de gestion des accès concurrents sur une notice. La dernière modification écrase la première sans signalement.

Enfin, il n'est pas prévu de conservation de l'historique des modifications d'une notice. Tout changement intervenu est définitif.

## **2.6.3 La gestion des vocabulaires et du référentiel**

### **2.6.3.1 Caractéristiques générales de la solution**

La deuxième composante de la solution regroupe les fonctions de gestion et de maintenance des différents vocabulaires « métiers » mobilisés pour le projet.

Ce traitement est assuré au moyen de la solution propriétaire Intelligent Topics Manager (ITM) développée par l'éditeur français Mondeca.

Le déploiement de l'outil assure le traitement des vocabulaires et de leur composante ainsi que leur mise en relation. Il offre également aux équipes métiers (les documentalistes et autres utilisateurs autorisés) les environnements d'administration utiles au maintien, à l'alimentation et à l'enrichissement des vocabulaires. L'ensemble de la solution mise à disposition est manipulable grâce à un navigateur web. Nous présentons successivement les caractéristiques générales de la solution, son caractère étendu par rapport au périmètre d'exploitation défini pour le projet ainsi qu'une illustration des grands principes de traitement des vocabulaires importés en vue de leur exploitation dans le portail Plan4Learning.

### **2.6.3.2 Un domaine modélisé de connaissance**

Dans le langage de l'éditeur Mondeca, l'articulation des vocabulaires et la gestion de leur relation dans les formats standards du web sémantique ( OWL, SKOS, RDF, SPARQL) forme une base de connaissance, ou encore un référentiel. Il est aussi question d'une ontologie de domaine.

Cette ontologie décrit le cadre de relations sémantiques dans lequel vont prendre place les vocabulaires importés ainsi que leur composantes.

A l'intérieur de ce cadre, les vocabulaires et les concepts qu'ils contiennent vont être mis en relation, gérés, nommés, décrits, enrichis pour être interrogeables et exportables.

### **2.6.3.3 Un réservoir de métadonnées**

Leur représentation dans les standards RDF et SKOS permet à ce bloc fonctionnel de traiter ces vocabulaires et leur composantes (leurs attributs et leur mise en relation) comme un réservoir de métadonnées modélisées pour les rendre exploitables par les machines

(applications, moteur de recherche, etc.) identifiables et partageables sur le web au moyen d'un identifiant permanent (URI).

Ces métadonnées seront transférables vers d'autres applications pour des opérations d'annotation, d'indexation par les moteurs de recherche, d'interrogation, d'affichage, de partage, de publication dans un portail web, etc.

#### 2.6.3.4 Une solution logicielle étendue

La solution est étendue. Elle permet

- de traiter dans des standards de descriptions homogènes un grand nombre de vocabulaires différents : des vocabulaires structurés comme des thésaurus et leur structuration interne ; des listes d'autorité (institutions, pays, projets) des éléments de métadonnées eux mêmes extraits d'autres notices documentaires, etc.
- de garantir l'intégrité des vocabulaires et l'enrichissement leur description au moyen d'éléments additionnels ;
- d'organiser des relations entre ces vocabulaires et les concepts contenus dans ces vocabulaires (alignements) ;

Elle permet également d'envisager un traitement autonome du domaine de connaissance modélisé en vue d'exploitations futures. Les fonctions avancées de gestion de Thésaurus illustrent par exemple cette possibilité.

<b>Standards</b>	
Standards W3C	Standards Web Sémantique : OWL (modèle ontologie), RDF et SKOS pour la sérialisation des instances de l'ontologie (vocabulaires contrôlés, bases de connaissances etc.), URI pour l'identification des concepts, SPARQL pour l'expression des requêtes et règles.
<b>Gestion des vocabulaires terminologies et thésaurus</b>	
Import/ export	<ul style="list-style-type: none"> <li>- Import et export XML et SKOS</li> <li>- Serveur de terminologie accessible par toutes les applications via des web services</li> <li>- Traçabilité des opérations de création, modification, suppression...</li> <li>- Contrôle qualité des vocabulaires pour détecter des anomalies, manques, violations de règles d'intégrité,</li> </ul>
La gestion des vocabulaires	- Gestion complète de la définition du terme pouvant être

	<p>étendue en fonction des besoins (définition, source, exemple d'utilisation, traduction...)</p> <ul style="list-style-type: none"> <li>- Maintenance des vocabulaires,</li> <li>- Alignements inter-vocabulaires,</li> </ul> <p>Constitution de sous-ensembles de concepts pris dans différents vocabulaires,</p> <ul style="list-style-type: none"> <li>- Traçabilité des opérations réalisées sur les vocabulaires contrôlés,</li> <li>- Contrôles de conformité paramétrables : Capacité de décrire des règles de contrôle de conformité des vocabulaires contrôlés ; complétude de description des concepts, organisation des concepts entre eux, cardinalité des attributs...</li> </ul>
Identification des concepts	<ul style="list-style-type: none"> <li>- Identification des concepts par des URI.</li> <li>- Gestion d'identification multiple : URI X et URI Wikipedia ou autre, etc.</li> <li>- Gestion des synonymes et multilingue.</li> <li>- Création automatique des URI par des règles de nommage.</li> <li>- Contrôle de l'unicité des concepts</li> </ul>
Gestion de Thésaurus	<ul style="list-style-type: none"> <li>- Gestion d'un nombre illimité de thésaurus multilingues avec prise en compte des alphabets latin, cyrillique, arabe, chinois...</li> <li>- Description complète et enrichissable des descripteurs (dénominations multiples, notes d'application, image associée, niveau dans la hiérarchie...)</li> <li>- Gestion des relations de synonymie, terme générique/terme spécifique, terme associé</li> <li>- Gestion de relations métier spécifiques (...est la cause de..., ...est un composant de...)</li> <li>- Gestion des polyhiérarchies</li> <li>- Gestion de relations d'équivalence (<i>mapping</i>) entre thésaurus pour des opérations de transcodification ou rapprochement de contenus hétérogènes</li> <li>- Opérations de fusion des descripteurs</li> <li>copie, déplacement de structures hiérarchiques du thésaurus</li> <li>- Edition de listes alphabétiques et hiérarchiques</li> <li>- Fonctions de recherche, navigation, consultation</li> <li>- Représentation graphique des hiérarchies et des relations de termes associés</li> <li>- Import et export des thésaurus (XML, SKOS) - Accès par API et</li> </ul>

	Services Web - Serveur de terminologie accessible par toutes les applications via des web services
Gestion de taxonomie ad hoc	- Taxonomies utilisées pour construire des services de navigation dynamique ou de recherche à facettes dans un portail. - Ces taxonomies « ad-hoc » permettent d'organiser à façon les contenus et de présenter des sous-ensembles du contenu à des publics ciblés.

Tableau 4 Récapitulatif des standards et fonctionnalités de gestion des vocabulaires et Thésaurus dans ITM (source offre commerciale Mondeca)

### 2.6.3.5 Principes de gestion des vocabulaires dans la base de connaissance

Le modèle d'ontologie d'ITM repose sur les notions de **Classe, d'objet ou Topics** (individu ou instances), d'attributs, d'associations, de rôles et de contraintes associées à ces rôles.

La base de connaissance est ainsi constituée par l'ensemble de ces objets et des relations qui les lient :

- Les objets à représenter sont affectés à des classes. Un objet ne peut appartenir qu'à une seule classe.
- Les objets sont décrits par des informations : des **attributs** ( au minimum un nom) de différents types ( texte, url, date, pointeur, etc.)
- Les relations entre les objets sont définies par des **associations**.
- Un rôle contraint la classe d'objet qui pourra prendre part à l'association
- Les attributs et **les rôles** peuvent avoir des cardinalités. Pas les associations.

Dans le cadre du projet, tous les vocabulaires ont été affectés à une classe dans laquelle l'ontologie les traite comme **instance de classe**.

La gestion du référentiel a anticipé une exploitation qui rendait utile la création de plusieurs classes pour distinguer certaines composantes liées à certains vocabulaires. Ainsi le traitement envisagé du glossaire sur la gestion et la planification a donné naissance à une classe supplémentaire pour intégrer les commentaires que les utilisateurs du portail seront invités à soumettre en ligne sur chaque terme du glossaire.

Vocabulaires importés pour Plan4Learning	Classes de description dans l'ontologie de domaine ITM
Thésaurus de l'Unesco	Unesco Concept
	Pays de l'Unesco (liste distincte des pays membres)
	Mots clés (descripteurs)
Plan de Classement de l'IIPE	Thème de Classification
Glossaire	Termes du Glossaire
	Commentaires sur les Termes du Glossaire (remontée des contributions des utilisateurs du portail)
Institutions	Organisations
Projets	Projets
Pays (membres de l'Unesco)	Pays
Autres	<ul style="list-style-type: none"> <li>* Groupe de Listes contrôlées</li> <li>- Audience</li> <li>- Type de contributeur</li> <li>- Type de membre</li> <li>- Région</li> <li>- Type de document (article, document, monographie, site web)</li> </ul>

### 2.6.3.6 Les fonctions d'administration

Pour l'administration et la maintenance des vocabulaires l'outil fait intervenir différents espaces de travail. Seul le statut d'administrateur donne accès aux droits avancés utiles aux interventions sur la structure de l'ontologie.

Quatre profils d'utilisateurs sont prédéfinis.

Profils utilisateurs	Droits	Espace de travail
Administrateur	<ul style="list-style-type: none"> <li>- création, modification des données</li> <li>- accès à l'écran d'administration</li> <li>- gestion des droits d'utilisateurs</li> </ul>	<p>Espace Administrateur</p> <p>Espace « Ontologie », qui structure un/plusieurs espace(s) de base de connaissances (terminologies ou instances).</p> <p>Renommer un élément de l'ontologie (classe, attribut, association, rôle)</p> <p>Ajouter un attribut sur une classe</p> <p>Ajouter une association et ses rôles etc.</p>
Editeur	Consultation et création des données dans l'espace de travail	Espace édition
Utilisateur	Proposition de modifications dans l'espace de travail, qui devront être validées par un administrateur	Espace de <i>workflow</i> « contexte de proposition »
Lecteur	consultation des données dans l'espace de travail	

Tableau 5 Les quatre profils d'utilisateurs ont des droits par espace de travail dans l'environnement d'ITM



## 2.6.4 Le moteur de recherche : alimentation, requête et affichage

Le bloc fonctionnel destiné à recevoir les flux de données du gestionnaire de notice et de l'entrepôt de métadonnées est le moteur de recherche Apache SolR<sup>46</sup>

Il s'agit d'une plate forme de recherche *open source* créée par la fondation Apache et distribuée sous licence libre.

### 2.6.4.1 Les flux d'alimentation du moteur

Dans le dispositif le moteur est alimenté régulièrement (tous les soirs)

- Par l'export des notices depuis le logiciel de gestion des notices
- Par l'export du glossaire et des autres référentiels depuis ITM.

Ces flux permettent l'indexation du moteur de recherche et la constitution de son dictionnaire de synonymes à partir desquels seront traités les recherches à facette et l'extension sémantique des requêtes.

Les modalités des requêtes

Pour traiter ses requêtes, SolR lance des recherches en texte intégral<sup>47</sup> dans l'ensemble des documents indexés. Il est capable de traiter tous types de contenus et de formats textes (Word, pdf, etc.).

### 2.6.4.1 La mise en forme des résultats

La présentation des résultats des requêtes du moteur de recherche procède d'une configuration. Parmi les différentes options de mise en forme disponibles les options retenues concernent la recherche et l'affichage.

#### 2.6.4.1.1 Les options liées à la recherche

---

<sup>46</sup> Site institutionnel : <http://lucene.apache.org/solr/>

<sup>47</sup> « La recherche (en) plein texte (appelée aussi recherche en texte intégral ou recherche de texte libre) est une technique de recherche dans un document électronique ou une base de données textuels, qui consiste pour le moteur de recherche à examiner tous les mots de chaque document enregistré et à essayer de les faire correspondre à ceux fournis par l'utilisateur ». Source : [http://fr.wikipedia.org/wiki/Recherche\\_plein\\_texte](http://fr.wikipedia.org/wiki/Recherche_plein_texte))

- Des options d'affichage de résultats (10 par défaut mais modifiable dans l'interface utilisateur)

- Des options de facettes :

Facettes simples par ordre décroissant de résultat et ordre alphabétique (contrairement aux notices bibliographiques, la navigation dans le glossaire n'est permise que par ordre alphabétique) ;

Facette hiérarchiques : déployant la hiérarchie structurelle des vocabulaires ( plan de classement, et thésaurus de l'Unesco) ;

Facettes d'intervalle (Range) : permettant le choix par nombre de résultats associés aux axes des facettes ;

Facettes géographiques pour filtrer les résultats par pays sur la base d'une carte ;

Des fonctions d'auto-complétion : suggestion de termes pour la recherche dès les premières lettres à partir des vocabulaires contrôlés

Action sur la liste des résultats (options d'export / format bibliographique)

#### 2.6.4.1.2 Les choix concernant l'affichage des pages

Deux interfaces ont été prévues pour l'affichage des flux du moteur de recherche et leur premier habillage que nous illustrons par les figures 3 et 4.

La première est dédiée à la recherche et à l'affichage des notices bibliographiques.

Elle ouvre sur des options de navigation multilingue, sur les différentes facettes, sur la cartographie utilisée comme filtre géographique, les options d'exports et sur le nombre de notices correspondant aux interrogations. L'interrogation débouche sur l'affichage d'une page de second niveau affichant une notice complète.

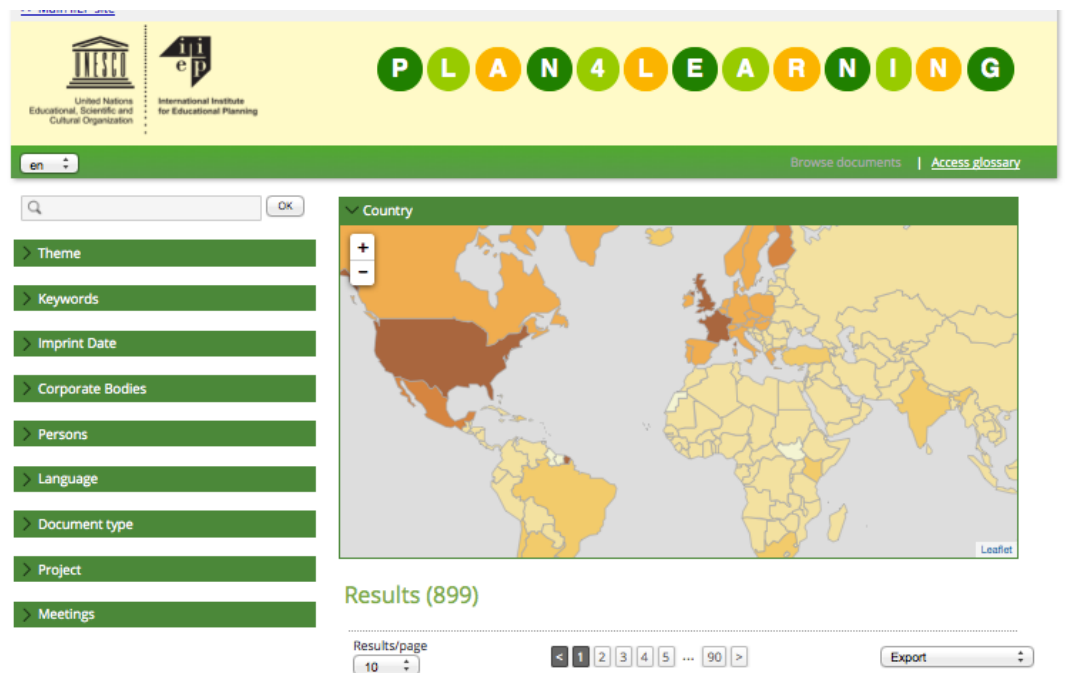


Figure 3 L'interface de recherche et de navigation à facettes dans les notices documentaires du portail d'exploitation. (Source capture d'écran <http://plan4learning.iiep.unesco.org/plan4learning/>)

La seconde interface est réservée aux seuls termes du glossaire [figure 4]. Elle est composée d'une option linguistique de navigation (fr, esp, eng), d'une barre de recherche qui permet d'interroger les termes du glossaire en trois langues, d'une barre de navigation alphabétique dans les différents termes, l'affichage d'une liste de termes cliquables accompagné d'une définition. L'interrogation d'un terme du glossaire affiche sa notice détaillée dans une page de second niveau.

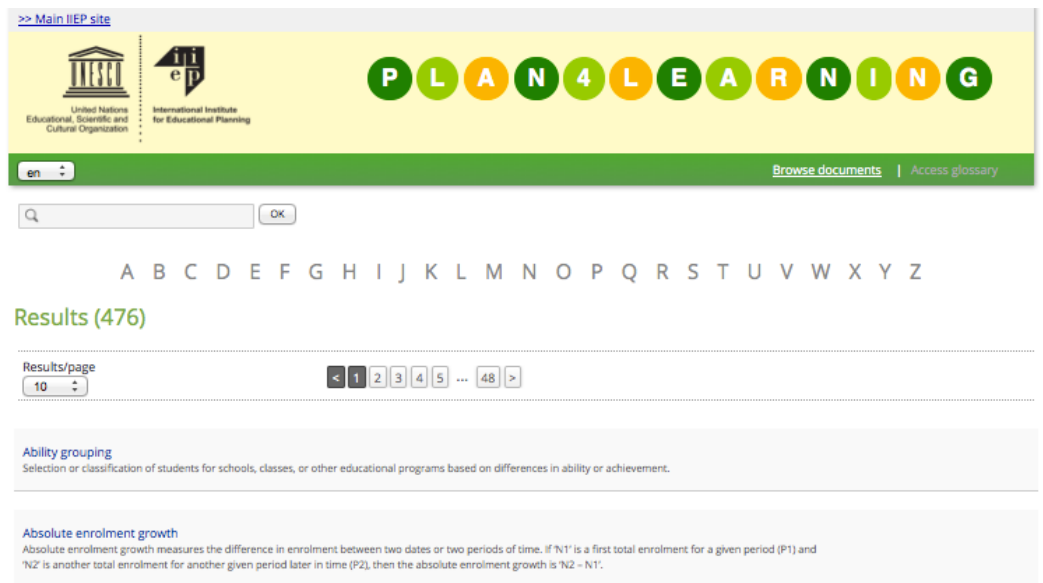


Figure 4 La page dédiée à la recherche et à la navigation alphabétique dans le Glossaire

Source : capture d'écran <http://plan4learning.iiep.unesco.org/plan4learning/>

A l'échelle du projet, l'utilité de ce portail d'exploitation est décisive. En phase initiale, avant la construction de portail final, il fonctionne comme un espace de travail pour la réalisation de tests de recherche ; il affiche rapidement (sous 24 heures) les effets des modifications introduites dans les alignements opérés entre les différents vocabulaires et les enrichissements souhaités des notices (notice bibliographique et notices des termes du glossaire) et de leur champs de descriptions.

# **Troisième partie**

## **Exploitation des vocabulaires et services rendus**

Cette partie se propose de mettre en regard les différents vocabulaires exploités par le portail thématique et les grandes familles de services destinés à faciliter la recherche et la navigation des utilisateurs finaux. Une première section décrit les principaux vocabulaires, leur origine, leur nature et les grandes lignes de leur modélisation dans la base de connaissance. Elle est ponctuée par un récapitulatif détaillé des familles de services. Dans une seconde section je restitue de façon détaillée un travail d'enrichissement de vocabulaire par alignement sur deux vocabulaires d'indexation. Je reviens sur les modalités de ce travail et ses enjeux en termes d'enrichissement. Je ponctue cette partie par une série d'observations et de préconisations

## 3 Les vocabulaires du portail et leur modélisation

---

### 3.1 Objectif et limites de la description

La présente section passe en revue ces différents vocabulaires en précisant pour chacun sa nature, le type d'exploitation qui en est fait dans le projet de portail « Plan4Learning ». Elle s'appuie pour l'essentiel sur leur description dans l'interface d'édition des notices du logiciel de gestion des vocabulaires ITM.

Chacun de ces vocabulaires pourrait faire l'objet d'une analyse détaillée, rappelant son historique, ses modes d'exploitation passés, actuels ou futurs, les politiques de maintenance ou de gouvernance qui président ou ont prévalu pour leur mise à jour, etc. Tel n'est pas l'objectif qui se concentre essentiellement sur quelques unes des propriétés qui ont trait à leur exploitation dans le cadre du présent projet de portail<sup>48</sup>. Nous introduisons néanmoins cette présentation par un rappel liminaire de l'hétérogénéité des vocabulaires et les grands principes de modélisation qui ont permis d'en tirer profit.

### 3.1.1 Retour à la modélisation

Comme on l'a vu dans l'exposé liminaire de la modélisation du référentiel, l'ontologie de domaine représente les concepts des vocabulaires dans des classes. Chaque vocabulaire ne pouvant être assigné comme instance de classe qu'à une seule classe.

La description des propriétés de chaque vocabulaire s'opère dans des champs de type attributs et association.

La modélisation permet enfin l'exploitation des concepts de chaque vocabulaire (classe de concept) par leur identification au moyen d'un URI et de format de description standards (Skos, XML, RDF) destinés aux machines (outils d'indexation, moteur de recherche, CMS).

Une illustration concrète de cette interopérabilité est fournie par l'outil d'édition des notices bibliographiques. Les champs de métadonnées retenus pour décrire les notices permettent d'importer au moyen d'une liste déroulante, les concepts du référentiel maintenus et identifiés dans ITM : mots clés du thésaurus, thèmes d'indexation, projets, pays, institutions, métadonnées descriptives (type de documents) etc.










La figure suivante illustre comment pour la description de la référence bibliographique dont le titre est « Participation and performance in education in Sub-Saharan Africa with special reference to Kenya: improving policy and practice », les mots clés d'indexation tirés des descripteurs du thésaurus sont identifiés par leur uri et sélectionnables dans une liste contrôlée à l'image du descripteur : Access to education (<http://skos.um.es/unescothes/C00019>).


Keywords

Mots clés choisis

Identifiant URI du mot-clé








Option de sélection du mot clé dans l'arborescence du Thésaurus de l'Unesco


- Educational policy (<http://skos.um.es/unescothes/C01267>) 
- Access to education (<http://skos.um.es/unescothes/C00019>) 
- Academic achievement (<http://skos.um.es/unescothes/C00008>) 
- Examinations (<http://skos.um.es/unescothes/C01455>) 
- Enrolment ratio (<http://skos.um.es/unescothes/C01372>) 
- Educational statistics (<http://skos.um.es/unescothes/C01284>) 
- Girls education (<http://skos.um.es/unescothes/C01697>) 
- Educational attendance (<http://skos.um.es/unescothes/C01210>) 
- School age population (<http://skos.um.es/unescothes/C03506>) 



 Add a new keyword


or



**Select term** ✕

-  01 - Education (14)
-  02 - Science (17)
-  03 - Culture (13)
-  04 - Social and human sciences (9)
-  05 - Information and communication (9)
-  06 - Politics, law and economics (17)
-  07 - Countries and country groupings (9)

Personal name (subject) 

Related institutions  

Related meetings 

Projects  

[www.iiep.unesco.org/project/](http://www.iiep.unesco.org/project/)

Figure 5 Extrait du champ de description des mots clés (Keywords) dans l'interface d'édition d'une notice bibliographique



## 3.1.2 Le glossaire sur la planification de l'éducation

### 3.1.2.1 Les besoins de l'exploitation

Le glossaire se présente comme une liste alphabétique de 650 termes (volume en début de mission). Il est le produit d'un travail d'harmonisation mené conjointement par les stagiaires et leurs formateurs au cours des sessions de formation (Programme de formation approfondie) de l'IIEP spécialisée dans la gestion et la planification de l'éducation. Son élaboration a été lancée au début des années 2000, à la demande des participants aux formations de l'IIEP, pour obtenir des définitions précises de termes techniques dans une langue (anglais ou français le plus souvent) qu'ils ne maîtrisaient pas bien, afin d'en faciliter la compréhension.

Les besoins de son exploitation dans le portail final peuvent se décrire comme suit :

- **Un besoin de balisage sémantique du domaine.** La publication du glossaire doit apporter une version stabilisée et actualisée du sens attribué par la communauté des chercheurs et des spécialistes de la planification de l'éducation aux termes employés et leur contexte d'utilisation. A cette fin, il est proposé en trois langues (français, anglais et espagnol). Les descriptions d'un terme du glossaire renseignent, une définition, un accès aux traductions dans les différentes langues, un contexte d'utilisation, la mention des sources utilisées pour la définition et le contexte d'utilisation, une liste de termes associés, lesquels renvoient par navigation dans le portail web à d'autres termes proches dans le même glossaire.

- **Un besoin de repérage dans le périmètre documentaire de référence,** en indiquant pour chaque terme du glossaire son inscription dans la thématique correspondante du plan de classement de l'IIEP sur lequel sont indexées les ressources documentaires. Cette fonction est assurée par l'affichage, au-dessus de la définition d'un terme, d'un fil d'Ariane cliquable qui déploie la hiérarchie des thèmes du plan de classement associés.

- **Un besoin d'accès aux notices** documentaires depuis le glossaire étant entendu que les notices bibliographiques ne sont pas indexées sur les termes du glossaire.

- **Enfin, un objectif de contribution collaborative.** Il s'agit de permettre aux publics cibles d'enrichir de façon collaborative les champs d'un terme du glossaire. En bas de

chaque page, cette option est proposée au moyen d'une question (Would you like to share a definition, a context or a translation?) ouvrant sur un formulaire en trois langues pour l'identification de l'utilisateur et le renseignement des champs proposés à l'enrichissement.

La [figure 6] ci-dessous pointe chacune des fonctions répondant à ces besoins d'exploitation (consultation, navigation, relance de recherche de notices, collaboration).



Figure 6 : Affichage d'une page d'un terme du glossaire dans le portail d'exploitation (Source : capture d'écran).

### 3.1.2.2 La description des termes du Glossaire dans le référentiel

L'édition des notices de terminologie dans ITM permet de renseigner des champs. L'interface ouvre également sur la description des associations qui relient les instances des classes et le rôle affecté à l'association (termes associés / related terms).

Le tableau qui suit donne une vue d'une notice complète d'un terme du glossaire en mode édition dans le logiciel ITM. On y voit qu'à ce stade d'avancement du projet, l'exploitation ne fait pas appel à l'intégralité des attributs disponibles. Certains ne sont pas exploités (*comments, linguistic notes, technical notes*) ; d'autres sont en phase d'alimentation et non renseignés (*used for* ; traductions manquantes etc.). En revanche cette version, intègre des attributs modifiés ou additionnels (« *example of use* » en remplacement de « *context* » et

ajout de « source of the example ») par rapport à la version initialement livrée. Ces attributs ont été suggérés, définis et administrés suite au travail d'enrichissement ajustement du glossaire que nous aborderons dans la section suivante.

Liste des entrées attributs dans l'interface d'édition de notice des termes de la Classe « Glossary terms » dans ITM	Entrée multilingue en texte libre	Exemple : notice renseignée du terme du glossaire « Access to education »
Name	oui	en - Access to education es - Acceso a la educación fr - Accès à l'éducation
Comment	oui	Non exploité
Corresponding concept	Non (choix dans liste contrôlée pour Lien vers 0 ou plusieurs descripteurs du thésaurus de l'Unesco )	« Access to education »
Definition	oui	en -Access to education refers to whether or (...) fr -L'accès à l'éducation signifie qu'un enf (...)
Definition source	oui	en - IIEP ATP training materials fr -Supports de formation de l'ATP de l'IIPE
Domain	Non (choix dans liste contrôlée pour Lien vers 0 ou plusieurs thèmes du plan de classement )	261C - Educational equality
Example of use	oui	en – As a consequence, some countries, includ (...) fr - C'est pourquoi certains pays, notamment (...)
Source of the example	oui	en - Crosier, D.; Parveva, T. 2013.

		Fundament (...) fr - Crosier, D. ; Parveva, T. 2014. Principe(...)
Linguistic note	oui	Non exploité
Technical note	oui	Non exploité
Used for	oui	Non renseigné (synonymes, termes non descripteurs du thésaurus)
Variant	oui	Non exploité
Related Glossary terms	Non (choix dans liste contrôlée des autres termes du glossaire)	Affichage par couples de termes du glossaire associés : (n=14) - Access to education Net intake rate - Access to education Enrolment - Etc.

Tableau 6 Description des attributs d'un terme du glossaire dans la fonction Edition d'ITM et exemple.

### 3.1.3 Le thésaurus de l'Unesco

#### 3.1.3.1 Périmètre et structuration

Le thésaurus de l'Unesco couvre les grands domaines de la connaissance associés aux différents champs d'action de l'Unesco, soient les sept divisions suivantes :

- 1 Education
- 2 Science
- 3 Culture
- 4 Social and human sciences
- 5 Information and communication
- 6 Politics, law and economics
- 7 Countries and country grouping

L'enrichissement et les traductions du thésaurus se sont succédés depuis sa première édition en 1977. Elles ont porté le nombre de termes à 8600 en français, anglais, espagnol et en russe. La structuration générale s'appuie sur 87 micro-thésaurus. La seule division Education – la plus importante en volume- est ainsi subdivisée en 14 microthésaurus (abrév. MT) à des fins de repérage et de recherche.

Les fonctions classiques de description univoque des termes, et d'explicitation des relations entre les termes sont assurées au moyen des symboles de relations TG (terme générique), TS (terme spécifique), TA (terme associé). Dans le portail d'exploitation actuel de l'Unesco<sup>49</sup> une notice complète d'un terme descripteur contient ainsi : le microthésaurus (MT) auquel le terme appartient ; la version anglaise, espagnole et russe ; une note explicative le cas échéant (NE) ; les relations entre les termes (EP, TG, TS, TA) ainsi que, entre crochets, le nombre de notices indexées sur chacun des termes.

Voici par exemple la notice du terme descripteur « Attitude de l'étudiant » en français dans le portail officiel du thésaurus de l'Unesco<sup>50</sup>

Terme :           Attitude de l'étudiant   [565]  
English term: Student attitudes  
Término español: Actitud del estudiante  
Русский термин : Отношения (установки) учащихся

MT 1.05 Sciences de l'éducation et milieu éducatif

<sup>49</sup> <http://databases.unesco.org/thesaurus/>

<sup>50</sup>

TG Comportement de l'élève [176]  
....TG2 Milieu éducatif [602]  
TA Adaptation de l'étudiant [48]  
TA Attitude de l'enseignant [501]  
TA Environnement de la classe [169]  
TA Participation de l'étudiant [496]



Pour l'importation du thésaurus de l'Unesco dans la base de connaissance, l'équipe projet s'est appuyée sur une version déjà représentée dans le standard SKOS<sup>51</sup>

Cette conversion a été réalisée par Juan Antonio Pastor Sánchez<sup>52</sup> docteur en science de la documentation rattaché à l'université de Murcia (Espagne) dans le cadre de ses propres recherches sur la gestion collaborative de la publication des thésaurus<sup>53</sup> permises par le standard SKOS et au terme d'une collaboration avec l'Unesco (bureau de la gestion de l'information et des organisations).

Dans l'outil de gestion, on retrouve la modélisation du thésaurus et son partage des descriptions entre attributs et association. Les relations hiérarchiques et d'association sont traitées comme des associations, tandis que les *Used for* (Employés pour) ont le statut d'attribut.

La maintenance dans la base autorise le traitement de tous les concepts. Elle permet aussi une exploitation différenciée pour les fonctions de filtres. Les mots clés des facettes sont puisés dans les seuls termes descripteurs.

---

<sup>51</sup> Son identifiant : <http://skos.um.es/unescothes>.

<sup>52</sup> Voir son blog : <http://www.um.es/docencia/pastor/investigacion/>

<sup>53</sup> Pour une présentation de ce travail par l'auteur dans <http://lists.w3.org/Archives/Public/public-esw-thes/2013Jan/0002.html> Pastor Sánchez, J.A. (2009). Diseño de un sistema colaborativo para la creación y gestión de tesauros en Internet basado en SKOS. Directores de tesis: Martínez Méndez, F.J. y Rodríguez Muñoz, J.V. Publicada en TDR (Tesis en Red). ISBN 978-84-692-1932-4. (en espagnol uniquement).

Liste des entrées attributs dans l'interface d'édition de notice des termes de la Classe « Unesco concept » dans ITM	Entrée multilingue en texte libre	Exemple : notice renseignée du terme Access to education
Name	oui	en -Access to education (Access to education... es -Acceso a la educación (Acceso a la educa... fr -Accès à l'éducation (Accès à l'éducation... ru -Доступ к образованию (Доступ к образован...
Definition	oui	
In group	Non, choix dans liste des microthésaurus	1.10 Educational policy
Notation [0,*]	Oui	Non renseigné
scope note [0,*]	Oui	Fr -Conditions, circumstances or requirement... es -Circunstancias, condiciones o requisitos... fr -Circonstances ou conditions d'admission ... ru -Условия, обстоятельства или требования, ...
Used for	Oui	Non renseigné
Broader - Narrower	Non, choix du rôle pour	<u>Broader</u> : Educational opportunities

	l'élément courant Broader /Narrower	<u>Narrower</u> : Access to education Educational priority areas Universal education
Related Concepts	Non, choix dans liste contrôlée des termes issus de tous les autres vocabulaires sauf Glossaire.	Affichage des couples de termes du thésaurus associés : (n=5) Access to education Free education Access to education Related concepts Educational priority areas Access to education etc.

Tableau 7 Description des termes descripteurs du thésaurus de l'Unesco dans la notice d'édition d'ITM et exemple.

### 3.1.4 Le plan de classement de l'IPE

Le plan de classement est le vocabulaire de référence traduit en trois langues pour l'indexation, le classement et la recherche des ressources documentaires du centre de documentation de l'IPE. Son élaboration procède d'un choix ancien. A un moment où le siège de l'Unesco tentait de faire prévaloir la Classification Décimale Universelle CDU comme outil de catalogage. L'IPE a soutenu l'option de l'emploi d'un plan de classement doublé par le thésaurus pour l'indexation. Selon Asunción Valderrama « la classification décimale universelle aurait généré des cotes trop longues pour le classement sur les rayons dans une bibliothèque aussi spécialisée que la nôtre. Le choix d'un plan de classement propre complété par des descripteurs s'est finalement imposé comme manière de pallier le manque de « finesse » dans la description<sup>54</sup>. »

Dans le portail final, les thèmes du plan de classement alimentent les facettes, autorisent la relance des recherches dans les notices documentaires à partir des notices du glossaire, et servent au *tagging* des contenus éditoriaux du site.

<sup>54</sup> Entretien

### **2.3.6.1 La structuration du plan de classement**

Il est découpé selon une logique thématique en neuf parties numérotées, soit dans la version française :

- 1- Le contexte de l'éducation
- 2 - La planification de l'éducation
- 3 - Niveaux et types d'enseignement
- 4 - Gestion de l'enseignement
- 5 - Systèmes régionaux et nationaux d'éducation
- 6 - Méthodes, techniques et contenus de l'éducation
- 7 - Coopération internationale
- 8 - Apprentissages et évaluation
- 9 - Ouvrages de référence

Chaque partie se subdivise en un nombre inégal de sections sous-numérotées, soit, par exemple, pour la partie 6 :

- 6 – Méthodes, techniques et contenus de l'éducation
- 61 - Enseignement et apprentissage
- 62 - Matériel didactique
- 63 - Technologie de l'éducation
- 64 - Informatique éducative
- 65- Curriculum

A l'intérieur de chaque section, sont listés par ordre alphabétique les thèmes de même niveau avec leur libellé suivi d'un code alphanumérique indiquant l'appartenance du thème à sa hiérarchie. Sous la section 61 on trouve ainsi :

- 61 - Enseignement et apprentissage
- Relations maître-élèves - 610A
- Stratégies d'enseignement – 610B

Liste des entrées attributs dans l'interface d'édition de notice des thèmes du plan de classement ( Themes classification )dans ITM	Entrée multilingue en texte libre	Exemple : notice renseignée du thème : 420A -Grading and grouping
Name	oui	en - 420A - Grading and grouping (420A - Grad... es - 420A - Clasificación y agrupamiento de l... fr - 420A - Classement et groupement des élèv...
Definition	oui	Non exploité
Exact name	oui	en - Grading and grouping es - Clasificación y agrupamiento de los alum... fr -Classement et groupement des élèves
Note	Oui	Non exploitée
scope note [0,*]	Oui	Non exploitée
Theme code	non	420 A
Used for	Oui	en - Ability grouping en - Nongraded schools en - Streaming
Broader - Narrower	Non, choix du rôle : Broader /Narrower	Broader : 42 - school management Narrower : 420A -

		Grading and grouping
For Reference	Non, choix du rôle : - Initial theme - Refers to this theme.	- Non renseigné
Reference note	- Oui	- Non renseigné
Related Concepts	Non, choix dans liste contrôlée des termes issus des différents vocabulaires ( sauf glossaire ): - Theme classification - Thésaurus concepts, - Organisation - Unesco Country - Country - Project	340C - Community education 420A - Grading and grouping

Tableau 8 Description des thèmes du plan de classement dans la notice d'édition d'ITM et exemple.

### 3.1.5 La liste des pays

Cette liste pays est tenue à jour par l'Unesco. Elle contient les noms des 193 Etats membres de l'Unesco ainsi que les membres associés (n= 9 en 2013)<sup>55</sup>. Elle est clairement distinguée de la liste des 320<sup>56</sup> pays gérée dans le référentiel dans les « groupe des pays » au titre de la subdivision 7 du thésaurus de l'Unesco « Pays et groupes de pays ». Dans le projet de portail, il est prévu d'exploiter cette liste comme une facette pour la navigation dynamique, le filtrage des résultats de recherche de notices, et le repérage géographique des programmes d'éducation. A la différence des autres facettes hiérarchiques permises par les vocabulaires structurés la liste pays est destinée à un traitement cartographique.

<sup>55</sup> La liste des membres est publiée en ligne <http://www.unesco.org/new/fr/member-states/countries/>

<sup>56</sup> Le volume de cette liste qui excède de loin le nombre total de pays recensés sur la planète tient compte de tous les changements d'appellation, des créations d'état, des reconnaissances officielles, etc.

La tenue à jour de la liste revêt une importance particulière pour l'IPE. La modélisation et le renseignement des attributs descriptifs le permettent au moyen d'une quinzaine de champs. Ceux –ci reprennent et complètent les entrées affichées sur les pages pays du site officiel unesco.org.

- Code pays normalisé en trois lettres (ISO 3166-Alpha 3)<sup>57</sup>
- Drapeau
- Date d'admission de l'Etat
- Statut de membre du bureau exécutif : vrai /faux
- Région de rattachement
- Nombre de sites inscrits au patrimoine mondial
- Un attribut Used for (« EP- Employé pour ») renseigne les libellés du pays en trois langues (Exemple : used for fr -Tokélaou (les) ; es –Tokelau ; en –Tokelau).

---

<sup>57</sup> [http://en.wikipedia.org/wiki/ISO\\_3166-1\\_alpha-3](http://en.wikipedia.org/wiki/ISO_3166-1_alpha-3)

Description des notices de la classe « Country » dans l'outil d'édition ITM	Entrée multilingue en texte libre	Exemple : notice du pays Burkina Faso
Name	oui	en - Burkina Faso (Burkina Faso) es - Burkina Faso (Burkina Faso) fr - Burkina Faso (Burkina Faso)
Capital	oui	Ouagadougou
Corresponding Country	Non, choix dans liste de la classe Country Unesco (n=320)	Burkina Faso
Country id	Oui	En – Burkina Faso
Country Page	Non, attribut de type url renseigne : url + titre pour lien vers page de description du pays dans Unesco.org	En <a href="http://www.unesco.org/new/en/unesco/worl...">http://www.unesco.org/new/en/unesco/worl...</a> Es <a href="http://www.unesco.org/new/es/unesco/worl...">http://www.unesco.org/new/es/unesco/worl...</a> fr <a href="http://www.unesco.org/new/fr/unesco/worl...">http://www.unesco.org/new/fr/unesco/worl...</a>
Definition	Oui	Non renseigné
Executive board group	Non	Code : Va
Executive board member	Non choix dans liste : True / False	False
Flag	Non, attribut de type url renseigne : une Url de l'image du drapeau sur Unesco.org + titre	en - <a href="http://www.unesco.org/new/fileadmin/MULT...">http://www.unesco.org/new/fileadmin/MULT...</a>
ISO 3 Code	Non, code en trois lettres normalisé	BFA
Member type	Non choix dans la liste de statuts : - Associate member - Members	Members



	- Non member state or other entity maintaining permanent	
Observation	Oui	Non renseigné
Region	Non choix dans liste de la classe Regions	Africa
Scope note	Oui	Non renseigné
Unesco date	Non, date	14/11/1960
Used for	Oui	fr -Burkina Faso es -Burkina Faso en -Burkina Faso
World heritage sites	Non, choix dans liste de codes	1.0
Broader - Narrower	Non, choix du rôle : Broader /Narrower	Non renseigné
Related Concepts	Non, choix dans liste contrôlée des termes issus des autres vocabulaires sauf glossaire	Non renseigné

Tableau 9 Description des notices pays dans le mode édition de l'outil d'ITM et exemple

### 3.1.6 Les listes projets et institutions

Les deux dernières listes que nous abordons sont les listes alphabétiques annexes : Institutions (Organisations), Projets (Project). Elles ont fait l'objet d'une normalisation préalable à leur exploitation dans l'ontologie du domaine. L'objectif était de standardiser leur libellé quand il est destiné à l'affichage dans le portail ou à la description d'une ressource bibliographique. Le portail d'annotation des notices permet le renseignement des champs Pays, Projets, institutions selon cette normalisation. Celle-ci a consisté à corriger

l'hétérogénéité des descriptions par chaînes de caractère, et de standardiser l'usage des acronymes. Au terme de cette normalisation le volume est passé à 642 institutions<sup>58</sup>.

### 3.1.6.1 Les institutions

Les institutions regroupent 1209 personnes morales de différents statuts (publics, privés, associatifs, gouvernementaux, non gouvernementaux, etc.) dont les compétences sont également hétérogènes (universités, ministères de l'éducation, consortiums voués à un programme, organismes d'évaluation, centres de recherche, etc.). Elles peuvent avoir un rôle d'auteur, d'éditeur de document, d'initiateurs d'actions et de programme dédiés à l'éducation mais aussi au développement ou être spécialisées dans d'autres problématiques connexes relevant au delà de la planification de l'éducation du périmètre d'action étendu de l'Unesco. « La liste des institutions pour le portail s'est constituée à partir d'une extraction du catalogue bibliographique général du centre de documentation (qui en contient plus de 6500) indique Asuncion Valderrama. Le catalogue ne possède pas de notices d'autorité personnes morales, ce qui ne permet pas de renseigner les différentes langues, dans le cas d'institutions internationales par exemple (World Bank/Banque mondiale). »<sup>59</sup>

Le choix de la langue s'est porté en priorité sur l'anglais :

- 1) pour les institutions internationales
- 2) pour les institutions exprimées en caractères non latins (avec la forme anglaise du catalogue UNESDOC si elle existe, sinon en se référant aux autorités de la Library of Congress). Pour les institutions exprimées en caractères latins, la forme retenue est celle de la langue nationale de localisation de la collectivité.

Actuellement, 226 notices d'ITM sur les 1209 contiennent déjà un lien vers DBpedia.

L'outil d'édition renseigne chacune de ces institutions sous la classe ( organisation). Dans la navigation à facette on les trouve dans la catégorie libellées Corporates bodies en anglais (fr –Institution ; es – institucion).

Comme l'expose le tableau qui suit, l'agence Cambodian women's development agency (Agence cambodgienne pour le développement de la femme) est décrite au moyen d'un nom en anglais uniquement, d'un acronyme et d'un label affiché retenu pour le nom développé.

---

<sup>58</sup> Selon Thomas Francart, (\* T. Francart, 2013 ppt) qui a conduit cette normalisation, l'opération a permis de ramener dans un premier temps la liste projets de 727 libellés « chaînes de caractères » à 573 objets.

<sup>59</sup> Entretien

Attributs renseignés de la notice Organisations (institution) dans ITM	Cardinalité	Entrée multilingue texte libre	Exemple
Name (Nom)	[1,*]	anglais	Cambodian Women's Development Agency
Acronym (acronyme) :	[0,*]	non	CWDA
Display label (libellé d'affichage)	[0,*]	anglais	Cambodian Women's Development Agency

Tableau 10 Description de la notice de l'IPE « *Cambodian women's Development Agency* » dans le portail d'édition.

### 3.1.6.2 La liste des projets

Les projets rassemblent les initiatives, les programmes d'actions, les plans d'éducation, et autres actions d'évaluation centrées sur la thématique de l'éducation. On trouve parmi les 642 initiatives listées des actions telles que le rapport mondial de suivi de l'Éducation pour tous (EFA Global monitoring), le programmes PISA (« Programme international pour le suivi des acquis des élèves » en français, PISA en anglais), etc.

La liste des projets s'est constituée dès la création du catalogue bibliographique général, et s'est considérablement enrichie. Elle répond au champ "Identifier" de la base UNESDOC<sup>60</sup> Elle concerne des termes « qui reflètent le contenu d'un document mais ne sont pas issus du Thésaurus de l'Unesco. ("which reflect the content of the document but are not found in the UNESCO thesaurus.") Ce sont principalement "names of projects, programmes, international instruments, information systems, monuments, cities, mountains, rivers, deserts, etc..."<sup>61</sup>

« Pour l'IIFE, précise la directrice elle répond à la nécessité d'avoir des informations précises sur les projets menés sur le terrain dans les pays. Cette liste obéit aux mêmes choix que celle des institutions concernant la langue. De même que pour les institutions, il serait souhaitable d'envisager d'inclure un lien vers Dbpédia ou vers le site de référence du projet, afin de valoriser ce référentiel et de gagner en visibilité. »<sup>62</sup>

Les projets sont largement exploités dans le catalogue bibliographique et dans les bases de données internes, sous le champ "Identifiant".

La liste des projets est affichée dans la facette de navigation et permet la recherche et le filtrage des résultats. Elle participe également à la description des notices bibliographiques.

La fiche descriptive éditée pour le référentiel ne déclare ainsi pour le PASEC (programme d'analyse des systèmes éducatifs de la Confemen<sup>63</sup>) que les seuls attributs nom, acronyme et libellé affichés.

Tableau 11 notice du programme PASEC de la liste Projet

Attributs renseignés de la notice Project dans ITM	Cardinalité	Entrée multilingue texte libre	Exemple
Name (Nom)	[1,*]		Programme d'analyse des systèmes éducatifs de la Confemen
Acronym (acronyme) :	[0,*]		PASEC
Display label (libellé d'affichage)	[0,*]		PASEC

<sup>60</sup> Exemple de l'identifiant « identifier » PISA dans la notice Unesdoc, consultable [en ligne [http://www.unesco.org/ulis/cgibin/ulis.pl?catno=229771&set=005486CB30\\_0\\_143&gp=1&lin=1&ll=1](http://www.unesco.org/ulis/cgibin/ulis.pl?catno=229771&set=005486CB30_0_143&gp=1&lin=1&ll=1)]

<sup>61</sup> CDS/ISIS Cataloguing and indexing guide, UNESCO, 1994.

<sup>62</sup> Entretien

<sup>63</sup> Conférence des ministres de l'éducation et gouvernements de la francophonie (Confemen) site institutionnel <http://www.confemen.org>

## **3.2 Enrichissement du glossaire : retour sur la mission**

Entre le mois de juin et le mois de septembre j'ai pris part à l'enrichissement du Glossaire. Ce vocabulaire dont c'est la première exploitation, appelait une intervention sur la description des définitions multilingues, des indications de contextes destinés à être lus par les utilisateurs. Il appelait également plusieurs types de relations avec d'autres termes du glossaire et avec les termes d'autres vocabulaires de la base de connaissance. Ces liens sont destinés à la navigation interne dans le vocabulaire, à des fonctions de rebonds et à des fonctions d'expansion de la recherche. Au cours de cette phase je ne suis intervenu que sur les liens vers les thèmes du plan de classement et vers les descripteurs du Thésaurus. Pour donner sa pleine mesure à la description de cette phase du projet, j'indique également les modalités des enrichissements qui ont été faits par une autre intervenante sur les autres champs.

### **3.2.1 Périmètre de l'enrichissement**

L'intervention a porté sur la totalité des termes du glossaire, soit un volume de 650 termes. Pour chaque terme, le travail a consisté à renseigner manuellement, les entrées de type « lien » ou « texte » demeurées vides ou incomplètes dans leur notice de description administrée et éditée dans l'outil ITM. L'enrichissement a porté sur six champs descriptifs.

#### **3.2.1.1 Un état avancé de l'intégration et de l'exploitation**

Le travail s'applique à un vocabulaire importé et déjà partiellement exploité.

- Avancement de l'intégration (*back-office*)

Le glossaire est déjà importé dans le référentiel sous le libellé « *Termes du glossaire* ». (*Glossary terms*)

Les notices descriptives des termes du glossaire sont déjà structurées au moyen de 15 champs. Les modifications de structure (création et modification d'entrées) requièrent un développement ad hoc<sup>64</sup>.

---

<sup>64</sup> ou le statut d'administrateur

Les termes du Glossaire ne sont pas reliés aux autres vocabulaires même si plusieurs champs sont déjà paramétrés pour anticiper leur mise en lien (les champs attributs « *Domain* », « *Used for* », « *Corresponding concept* », et les champs d'association (« *Associated terms* » dans lesquels sont définies les relations intra et inter-vocabulaires : *Broader Narrower, Related terms* etc.).

Enfin, le glossaire n'est pas exploité pour l'indexation des notices documentaires.

#### - Avancement de l'exploitation (*front-office*)

L'intégralité du glossaire est déjà consultable dans un portail web d'exploitation opérationnel et accessible mais transitoire avant l'intégration des flux du moteur de recherche dans le CMS final.

Ce portail d'exploitation affiche les champs publiables d'un terme et autorise la recherche et la navigation alphabétique dans le glossaire.

Les champs renseignés dans la notice d'un terme du Glossaire en *back office* (ITM) sont indexés quotidiennement par le moteur de recherche qui les affiche (quand ils sont destinés à l'affichage) dans la page correspondante du portail d'exploitation.

L'emplacement et l'ordre d'affichage de champs descriptifs sont en place mais révisables.

Dans le même portail une interface distincte est dédiée à la recherche de notices documentaires et permet les options de recherche en langage naturel, par mot clés, par critères et l'emploi de filtres (facettes hiérarchiques, carte, etc.).

#### - Un portail final en cours de développement

Située en fin de phase 1 et en début de phase 2, la mission coïncide enfin avec la livraison des premières recettes de l'agence prestataire chargée du développement du portail final. La réception/validation de ces livrables a conduit à leur évaluation et à la commande d'ajustements/rectifications concernant la structuration du site, l'affichage de ses rubriques, et la navigation dans ses contenus.

Il fallait veiller à intégrer les conclusions de cette phase de test quand elles avaient un impact sur les champs de description (ordre d'affichage, libellés des champs affichés etc.) et inversement mettre à disposition des développeurs du site les identifiants des champs du glossaire éventuellement modifiés ou ajoutés au cours de l'enrichissement.

### **3.2.1.2 Les besoins et les services visés**

Trois catégories de besoins sous-tendent l'enrichissement envisagé.

Elles renvoient toutes à la définition de services destinés aux publics cibles du portail final.

- Afficher les descriptions les plus complètes des termes en trois langues (en, fr, es),

soit : le libellé du terme en trois langues, une définition, un contexte d'utilisation, les sources renseignant le contexte d'utilisation, des mots-clés.

- Naviguer par rebond entre les différents termes du glossaire.
- Afficher pour chaque terme consultable un balisage thématique correspondant au périmètre des ressources documentaires disponibles
- Permettre d'activer la recherche de notices documentaires à partir d'une page du Glossaire.
- Optimiser les résultats obtenus par le moteur de recherche suite à l'interrogation de la base de notices documentaire à partir d'un terme du glossaire.

Tableau 12 Les champs descriptifs du glossaire, leur vocabulaire lié et leur exploitation visée.

Les champs descriptifs sollicités	Mode de description dans l'outil d'édition ITM	Fonctions	Vocabulaire (s) lié (s)
« Context of use »	Texte libre	<ul style="list-style-type: none"> <li>• Affichage multilingue</li> </ul>	aucun
« Source of the example »	Texte libre	<ul style="list-style-type: none"> <li>• Affichage multilingue</li> </ul>	aucun
« Related terms »	Texte libre	<ul style="list-style-type: none"> <li>• Affichage</li> <li>• Navigation intra-glossaire</li> </ul>	Glossaire
« Domain »	Liste contrôlée	<ul style="list-style-type: none"> <li>• Affichage multilingue</li> <li>• Recherche</li> <li>• Navigation vers interface de recherche</li> </ul>	Thèmes du plan de classement
« Used for »	Texte libre	<ul style="list-style-type: none"> <li>• Recherche</li> </ul>	Termes ou EP « Used for » du Thésaurus
« Corresponding concept »	Liste contrôlée	<ul style="list-style-type: none"> <li>• Recherche</li> </ul>	Descripteurs du thésaurus « Unesco concept »

Au final on peut rattacher ces services à trois fonctions d'exploitation dans le portail final

- Affichage multilingue à vocation descriptive des termes et de leur contexte

- Navigation dédoublée en navigation interne au glossaire d'une part et navigation externe d'autre part allant d'une page Glossaire vers l'interface de recherche de notices documentaires.
- Recherche (expansion de la recherche de notices à partir d'un terme du glossaire).

La figure suivante propose un schéma des relations entre les champs sollicités, les vocabulaires autres que le glossaire auxquels ils sont liés<sup>65</sup> ainsi que le type de service visé en *front office*.

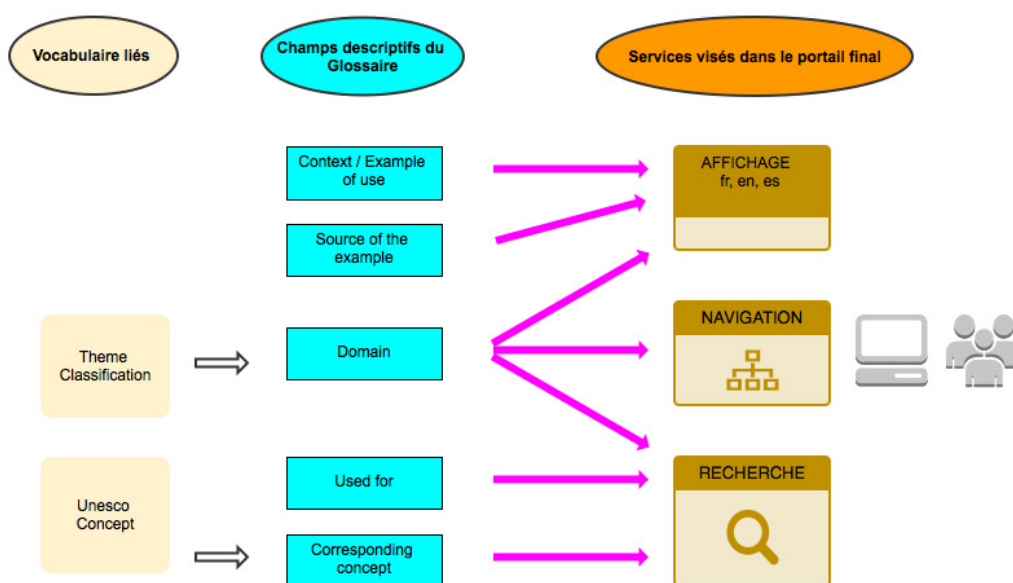


Figure 7 Figure 1 : Schéma des champs de description du périmètre d'enrichissement indiquant les vocabulaires liés et les services rendus en *front office*

<sup>65</sup> D'où l'absence dans la figure du champ *Related terms* qui permet des liens internes au glossaire et non vers d'autres vocabulaires.



## 3.2.2 Ciblage des champs descriptifs et *modus operandi*

Cette étape de la réalisation devait répondre précisément à deux séries de questions préalables à la phase opérationnelle. Dans l'outil d'édition, quels champs descriptifs correspondent aux enrichissements visés ? Ce point touche à l'identification des entrées et de leur conformité aux besoins de la mission. Par ailleurs, comment procéder à l'alimentation des descriptions, dans quel délai les réaliser, comment les valider ? Ces questions relèvent du mode opératoire. Nous les abordons successivement.

### 3.2.2.1 Le choix des champs descriptifs

Trois entrées pouvaient être ciblées s'agissant des descriptions de type texte :

- La description des contextes,
- La traduction des définitions existantes en anglais et en espagnol et le renseignement des termes associés (*Related terms*).
- Par ailleurs, le champ de type attribut « *Domain* » était déjà paramétré de façon à relier dans le référentiel un terme du glossaire à un ou plusieurs thèmes du plan de classement sur lequel sont indexées les notices documentaires.

Cet attribut « *Domain* » a été retenu pour générer le fil d'Ariane cliquable. Celui-ci remplissant simultanément une fonction de navigation hors des pages du glossaire et une fonction d'interrogation en activant une recherche de notice documentaire à partir du segment de l'arborescence du plan de classement exposé.

L'affichage des « *Related terms* » positionnés en bas des pages glossaire pouvait être assuré par le renseignement de l'entrée « *related terms* » de type « Association » dans l'outil d'édition.

Quoiqu'assez simple, cette procédure a du en cours d'exécution prendre en compte quelques re-définitions du périmètre d'enrichissement, et être assortie de vérifications complémentaires.

Ainsi, le champ initial « *Context* » a été renommé « *Example of use* » à l'affichage mais maintenu dans son libellé dans l'outil de *back office*.

Un champ « *Source of the example* » a du être créé dans la notice pour intégrer la citation des sources documentaires ayant permis de renseigner les « *Example of use* ».

Enfin, il a fallu veiller avec le prestataire à ce que le moteur intègre bien les termes (non-descripteurs) entrés dans le champs « *Used for* » pour traiter ses requêtes. (voir infra résultats)

### **3.2.2.2 L'extension de la recherche par les descripteurs du thésaurus**

Les liens à opérer entre les termes du glossaire et les descripteurs du thésaurus ont pareillement été précédés de deux types de questions.

#### **3.2.1.2.1 Un objectif d'expansion sémantique**

Il se rapporte au souci de garantir un accès aux notices documentaires y compris en interrogeant avec les termes du glossaire. Ceux-ci n'étant pas des descripteurs. L'expansion sémantique consiste à relier ces termes à des concepts du thésaurus et permettre au moteur de recherche d'enrichir son dictionnaire de synonyme avec ce nouveau terme lié.

Un exemple illustre l'écart du nombre de résultat obtenu selon qu'on interroge au moyen d'un terme du glossaire ou d'un descripteur du thésaurus :

*Part time teacher* (terme du glossaire) / *Teacher conditions of employment* (descripteur)

L'interrogation des notices documentaires par chacun de ces deux termes générant respectivement 3 résultats pour le terme du glossaire et 28 résultats pour le descripteur du thésaurus.

Dans l'outil, le lien s'opère en renseignant le champs « *Corresponding concept* » de la page des notices du glossaire.

Le choix s'est entouré toutefois d'un point de vigilance quant à la façon dont le moteur de recherche pouvait avec ou sans développement supplémentaire intégrer ces « *corresponding concepts* » pointant vers le thésaurus dans son traitement des requêtes.

### **3.2.2.3 L'évitement des liens de type « Association »**

Cette solution a évacué l'alternative de la construction d'une relation plus structurelle entre les deux vocabulaires permise par le recours à d'un champ de type association dans la notice.

Dans ITM les associations sont des liens bi-directionnels qui affectent les notices cibles. Elles sont exploitées en particulier pour créer des rôles entre les termes à l'intérieur des vocabulaires structurés. C'est par ce moyen que sont renseignées et maintenues les relations qui structurent les termes au moyen des relations d'association (Related terms) ; de hiérarchie (Broader/Narrower).

Ces entrées sont évidemment indispensables pour restituer la structure du thésaurus et celle du plan de classement dans l'ontologie et dans l'outil de maintenance du référentiel. Comme on l'a vu, le glossaire lui-même bénéficie de telles descriptions puisque un lien d'association relie bien un terme à ses termes associés dans le glossaire.

En revanche, relier de cette manière le glossaire aux descripteurs du thésaurus aurait outrepassé le but recherché.

**Illustration capture d'écran ITM** : Exemple d'utilisation d'un lien d'association pour les « *related terms* » dans la notice du terme du glossaire /Choix dans une liste contrôlée : « *Educational deprivation indicator* »

#### **3.2.2.4 Le traitement des équivalences**

Deux règles de mise en équivalence ont été définies en vue de la mise en lien du glossaire avec les thèmes du plan de classement d'une part et avec les descripteurs du thésaurus d'autre part.

Un terme du glossaire doit être relié à un thème du plan de classement et un seul à partir du champ « Domain » en vue de l'affichage d'un seul fil d'Ariane. L'alignement devant se faire sur un terme exactement équivalent ou proche.

Un terme du glossaire peut être relié à un descripteur du thésaurus et éventuellement à un ou deux autres au maximum à partir du champ « Corresponding concept »

### **3.2.3 Modalités d'exécution**

Les enrichissements à effectuer ont été répartis entre deux intervenants distincts placés tous deux sous la responsabilité de la directrice du centre de documentation, l'affectation des tâches est récapitulée dans le tableau ci-dessous.

#### **3.2.3.1 Les ressources mises à disposition**

- Formation à la prise en main d'ITM (organisation générale, fonctions édition, contexte de proposition, validation)
- Ouverture d'un droit d'utilisation ITM en mode utilisateur et en mode édition.
- Assistance permanente de l'équipe projet dans le cadre de réunions de validation, de cadrage et de points d'avancement informels.

### 3.2.3.3 Objectifs

- Volume : Traitement de l'intégralité des termes du glossaire par ordre alphabétique
- supports : Renseignement des champs « Domain » et « Corresponding concept » et édition dans l'outil ITM.
- Renseignement en double des enrichissements dans un tableur Excel pour contrôle, suggestion et éventuel traitement statistique des entrées.
- validation : par lots de 200 termes + 1 lot reliquat (250 termes) échelonnés sur la période 15 juillet- 1<sup>er</sup> septembre 2014.
- Date de validation : validation du dernier lot au 1<sup>er</sup> septembre 2014.

Modalités d'exécution	Lien vers le plan de classement de l'IPE	Lien vers le Thésaurus de l'Unesco
Volumétrie	650 termes	650 termes
Outil d'édition du lien	Gestionnaire de vocabulaires ITM	Gestionnaire de vocabulaires ITM
Nature du lien	Unidirectionnel de type Attribut : « Domain » dans la notice d'édition du glossaire dans ITM	Unidirectionnel de type Attribut : « Corresponding concept » dans la notice d'édition du glossaire dans ITM
Niveau d'alignement sur vocabulaire cible	Au thème équivalent le plus détaillé ou générique	Au terme descripteur équivalent exact ou proche
Mode de saisie	Semi-automatique par choix dans un vocabulaire contrôlé	Semi-automatique par choix dans un vocabulaire contrôlé
Nombre de concepts cibles à renseigner	n = 1 dans champ de saisie ITM (+ 1 suggestion sur fichier Excel)	1 ≤ n ≤ 3 dans champ de saisie ITM et tableur Excel
Support livrable	Notices « <i>Glossary Terms</i> » dans ITM + tableur Excel	Notices « <i>Glossary terms</i> » dans ITM + tableur Excel
Affichage dans le portail final	Automatique 24h après saisie = fil d'Ariane	Pas d'affichage
Workflow de validation	Non-automatisé	Non-automatisé
Validation par lots	Ex-post : par lot de 200 termes liés	Ex-post : par lot de 200 termes liés

Tableau 13 Récapitulatif des modalités d'exécution de la mise en lien du glossaire avec le plan de classement et le thésaurus de l'Unesco.

## 3.2.4 La réalisation des liens : bilan et retour d'expérience

### 3.2.4.1 Les liens vers le plan de classement

Au terme de la mission, chacun des 650 termes du glossaire étaient dotés d'un lien unique vers un thème équivalent ou proche dans le plan de classement.

En cas d'absence d'équivalence stricte entre le terme du glossaire et les thèmes du plan de classement, ont été privilégiés les thèmes sémantiquement proches ou des thématiques proches ou englobantes. Deux exemples illustrent ces cas de figure.

**Exemple d'équivalence exacte** : Basic education = Basic education 311A

- Terme du glossaire : « Basic education » « Education de base » « Educación básica »
- Thème lié : Basic education 311A
- Place du thème dans le plan de classement (fil d'Ariane) : LEVELS AND TYPES OF EDUCATION > Basic education > Basic education >

**Exemple d'un choix de thème générique** : Absolute enrolment growth > Educational indicators 231C

- Terme du glossaire : « Absolute enrolment growth » « Croissance absolue des effectifs scolarisés » « Crecimiento absoluto de la matrícula »
- Thème lié : 231C - Educational indicators
- Place du thème dans le plan de classement (fil d'Ariane) : EDUCATIONAL PLANNING > Statistical methods in educational planning > Educational indicators >

### 3.2.4.2 Les liens vers les descripteurs du thésaurus

Tous les termes du glossaire ont été reliés à au moins un « Corresponding concept » sélectionné dans les descripteurs du thésaurus de l'Unesco.

Contrairement, à l'opération précédente, plusieurs termes pouvaient être renseignés dans la notice. Nous avons limité ce nombre à trois.

Le traitement des équivalences a respecté les règles d'alignement choisi :

Privilégier les équivalents exacts, à défaut sélectionner des termes sémantiquement proches.

### 3.2.4.3 Modalité et durée d'exécution

Le travail d'alignement manuel qui a été réalisé condense plusieurs manipulations imbriquées depuis la consultation de chaque terme du glossaire à aligner pris un à un jusqu'au renseignement final des liens dans l'outil d'édition ITM. A la lumière de la mission effectuée et du temps effectivement passé on peut estimer le nombre de journée/ homme requis pour cette tâche à J/H = 35. Cette estimation ne porte que sur les seuls alignements du glossaire avec le plan de classement d'une part et les descripteurs du thésaurus de l'Unesco d'autre part. Elle ne prend évidemment pas en compte les enrichissements qui ont portés sur les champs textes (définition, contextes, etc.) réalisés parallèlement.

Par ailleurs, l'estimation intègre des phases de préparation inhérentes à l'état d'avancement du projet au moment de la mission :

- Identification certaine des champs de descriptions à cibler dans l'outil d'édition des notices du glossaire ITM ( en particulier le choix du champs « Corresponding concept » ainsi que les conditions et les modalités précises de son exploitation par le moteur de recherche a nécessité plusieurs échanges préalables au sein de l'équipe projet).
- Un temps de prise en main de l'outil d'édition ITM. Au-delà d'une demi-journée de formation s'est ajoutée deux ou trois jours pour se familiariser avec les différents interfaces.
- Un travail préparatoire visant à harmoniser l'alimentation et la validation progressive ( usage de l'espace de travail Workflow de validation dans ITM) des enrichissements menés parallèlement sur le glossaire par une seconde intervenante.
- les séquences d'échanges et de validation

Par ailleurs, le temps passé prend également en compte

- Une absence de connaissance initiale de l'environnement métier et documentaire spécifique de l'IPE et des outils d'indexation à l'œuvre dans la bibliothèque de l'IPE.
- La manipulation croisée et simultanée de plusieurs outils dans le but de repérer les termes et de capter le sens des termes à aligner : portail de publication du thésaurus de l'Unesco ;

Portail d'exploitation du projet Plan4Learning ; portail d'annotation des notices bibliographiques *back office* ; espace de travail ITM ; Plan de classement.

Enfin, au-delà du renseignement des champs descriptifs dans ITM, l'alignement s'est doublé d'un recensement systématique du nombre de notices accrochées aux termes dans le portail d'exploitation tels qu'ils apparaissent à partir d'une interrogation sur les termes descripteurs sélectionnés dans le thésaurus et les termes de départ du Glossaire. Le renseignement des champs ITM étant ainsi augmenté d'un tableur Excel lequel a également facilité les séances de validation / rectification par lots. Le tableau ci-dessous illustre l'enregistrement des alignements et des statistiques de résultats consignés dans le tableur Excel.

<b>Termes du glossaire</b>	<b>Nombre de réf. dans P4L avant lien</b>	<b>"Corresponding concept" Thésaurus Unesco</b>	<b>Nombre de réf. Dans P4L avant lien</b>	<b>Code Thème du plan de classement</b>	<b>Thème Plan de classement renseigné dans "Domain"</b>	<b>Autres Thème classification identifié</b>
Alternative education strategies	2	Alternative education	19	610B	610 B Teaching-learning strategies	810A Educational alternatives
Books development	7	teaching materials	13	660B	660 B books and reading	620 A Educational materials
Conditions of schooling	16	Educational quality	431	250B	250B - Quality of education	
School timetable	0	Educational timetables	3	420E	420E - School calendars	

Tableau 14 Extrait du tableau d'enregistrement des mises en lien du glossaire avec le thésaurus et le plan de classement de l'IIPE.

### **3.2.5 Validation**

L'alimentation quotidienne du moteur de recherche a permis de vérifier au fur et à mesure de la saisie que les fils d'Ariane correspondants étaient bien affichés dans le portail d'exploitation, qu'ils étaient actifs et qu'ils permettaient une relance de la recherche dans les notices documentaires.

L'enregistrement des saisies dans le gestionnaire des vocabulaires apporte toutes les garanties utiles en matière de possibilité de mise à jour des liens (modification, suppression) et d'éditations de rapports.

Parallèlement, un tableur a été livré. Il double ce traçage en renseignant pour tous les termes du glossaire les équivalences retenues et leurs libellés exacts. L'équipe projet dispose ainsi d'une base stable et validée - mais nécessairement datée - pour analyser le cas échéant, les niveaux de recouvrement des termes, les occurrences des thèmes du plan de classement et des concepts du thésaurus dans l'alignement à ce moment précis d'avancement du projet.



### 3.2.5.1 Les champs en cours d'enrichissement

L'enrichissement s'est poursuivi jusqu'à l'automne 2014. La progression de ce travail au cours de la période de référence autorise deux remarques touchant à l'extension de l'objectif de départ et à son impact sur le volume du vocabulaire.

Cette intervention touchait davantage à la profondeur sémantique des termes du glossaire. Il était le plus à même de détecter des anomalies (définitions initiales non renseignées, libellés manquants dans certaines langues), des redondances, des incohérences dans l'existence même de certains termes.

Ce travail a eu pour effet bénéfique d'affiner les champs descriptifs comme on l'a évoqué plus haut. Le champ initialement affiché comme « *context* » a été remplacé par la plus formulation plus circonstanciée « *Example of use* » ; un champ descriptif « *source of the example* ») a été ajouté dans la notice en back-office pour déclarer les sources auxquelles puisent ces exemples ; enfin plusieurs propositions ont pu être formulées pour l'intégration d'autres synonymes.

C'est également au fil de cette intervention qu'a été décidée la suppression de termes jugés non pertinents. Plus de 150 termes ont été effectivement supprimés faisant passer le volume du glossaire de 650 termes à 480.

### 3.2.5.2 Points de vigilance

Au terme de cette opération deux points de vigilance sont à verser en vue d'une évaluation éventuelle des effets de cette mise en lien.

#### 3.2.5.2.1 Contrainte de réalisation et *workflow* de validation

Les liens ont été faits à partir d'un Glossaire qui dans le même temps était lui-même en phase d'enrichissement/ajustement. Comme on l'a vu ce travail très utile a de surcroît permis de détecter des doublons ou des termes dont la pertinence et l'utilité ont été jugées discutables ou insuffisantes.

Ce décalage n'a pas eu d'incidence préjudiciable sur la mise en lien. Reste que les interventions ultérieures sur le glossaire (maintenance, collaboration) gagneront à intégrer systématiquement le travail accompli sur tous les champs descriptifs.

	Opérations d'enrichissement		
	Liens vers plan de classement	Liens vers descripteurs du thésaurus	Enrichissement des entrées
Etat d'avancement	Achevé et validé	Achevé et validé	En cours
Suite à donner	Evaluation des pratiques utilisateurs à configurer	Confirmation de prise en compte des « correspondant concepts dans l'indexation du moteur de recherche SolR. requête pour le moteur SolR + évaluation de requêtes à configurer	

Tableau 15 Récapitulatif de l'avancement et des suites à donner.

## **3.3 Les services rendus : synthèse**

A la lumière de la description des vocabulaires il est permis d'esquisser sinon une typologie des services rendus aux utilisateurs du futur portail du moins un récapitulatif par grandes familles de services qu'autorise la base de connaissance sous-jacente.

Comme on l'a vu ces services sont permis par l'alimentation de l'index du moteur de recherche et son dictionnaire de synonymes, et en amont par la structuration des vocabulaires, par leur mise en relation et par leur regroupement dans des groupes spécifiques. Signalons que ce recensement est provisoire. En particulier il ne prend pas en compte les ultimes arbitrages en cours concernant l'exploitation finale des vocabulaires dans le portail final. Nous nous basons pour l'essentiel sur la configuration stabilisée au moment de la publication du portail d'exploitation intermédiaire.

### **3.3.1 Des accès simplifiés aux ressources**

Les ressources peuvent être appelées depuis une simple barre de recherche dès la page d'accueil du portail et sans qu'il soit besoin d'utiliser une grammaire spécialisée d'interrogation ( opérateurs Booléens) ni de connaître à l'avance le libellé précis du ou des termes qui correspondent aux documents attendus.

- Dans cet environnement simplifié l'assistance proposée s'appuie sur

La suggestion automatique de dix termes par auto-complétion dès la saisie de la première lettre d'un mot dans la barre de recherche. Pour générer ses suggestions, le moteur de recherche s'appuie sur l'ensemble des vocabulaires contrôlés qu'il a préalablement indexé ( mots clés du thésaurus, non descripteurs, thème du plan de classement, nom de pays, etc.) L'auto-complétion est étendue aux trois langues et intègre la permutation des termes.

- La seconde aide apportée est l'interrogation multilingue. Elle permet à l'utilisateur de formuler une recherche dans une langue et d'obtenir des ressources mêmes si elles sont indexées dans une langue différente. Les trois langues qui font l'objet de ce traitement croisé des langues pour le portail sont le français, l'anglais et l'espagnol.

### **3.3.2 La démultiplication des points d'accès aux ressources**

La multiplication des points d'accès aux ressources participe autant de la stratégie de valorisation (visibilité) de celles-ci que d'une catégorie de services utilisateur complémentaire. Elle englobe les possibilités de relancer les recherches et de découvrir de nouvelles ressources. Trois types d'accès sont envisagés.

➤ Accès depuis les notices bibliographiques

Une notice bibliographique autorise ainsi le rebond depuis tous les champs indexés sur les vocabulaires administrés dans la base de connaissance.

A titre d'exemple la notice documentaire du rapport de la direction de la banque mondiale intitulé « Incentives to improve teaching: lessons from Latin America »

Permet de relancer la recherche sur les champs cliquables renseignés dans l'outil d'annotation des notices sur la base des vocabulaires correspondants suivants : Institution, Pays \*(Liste pays étendue au sens de l'Unesco), Theme \*(ie thème du plan de classement) et mots clés \*(descripteurs du thésaurus de l'Unesco).

(ndbp : en phase d'alimentation des notices, le nombre de valeurs à retenir pour l'indexation sur le plan de classement et sur les descripteurs du thésaurus est un choix d'indexation. Il a été arrêté à : inférieur ou égal à 5 pour les thèmes et à 10 pour les descripteurs.)

Champs de la notice	Valeur affichée dans la notice	Valeurs correspondantes importées depuis la base de connaissance
Institution	World Bank	Word bank
Pays	Peru Honduras Bolivia Chile Mexico Nicaragua Latin America Caribbean	Peru ( <a href="http://skos.um.es/unescothes/C02955">http://skos.um.es/unescothes/C02955</a> ) Honduras ( <a href="http://skos.um.es/unescothes/C01826">http://skos.um.es/unescothes/C01826</a> ) Bolivia ( <a href="http://skos.um.es/unescothes/C00421">http://skos.um.es/unescothes/C00421</a> ) Chile ( <a href="http://skos.um.es/unescothes/C00584">http://skos.um.es/unescothes/C00584</a> ) Mexico ( <a href="http://skos.um.es/unescothes/C02501">http://skos.um.es/unescothes/C02501</a> ) Nicaragua ( <a href="http://skos.um.es/unescothes/C02727">http://skos.um.es/unescothes/C02727</a> ) Latin America ( <a href="http://skos.um.es/unescothes/C02221">http://skos.um.es/unescothes/C02221</a> ) Caribbean ( <a href="http://skos.um.es/unescothes/C00512">http://skos.um.es/unescothes/C00512</a> )
Theme	Studies of	Studies of achievement

	achievement Teacher salaries	( <a href="http://www.iiep.unesco.org/Ontology/ThemeClassification#870E">http://www.iiep.unesco.org/Ontology/ThemeClassification#870E</a> ) Teacher salaries ( <a href="http://www.iiep.unesco.org/Ontology/ThemeClassification#470H">http://www.iiep.unesco.org/Ontology/ThemeClassification#470H</a> )
Mots clés	Educational finance Educational quality Academic achievement t Etc.	Educational finance ( <a href="http://skos.um.es/unescothes/C01232">http://skos.um.es/unescothes/C01232</a> ) Educational quality ( <a href="http://skos.um.es/unescothes/C01275">http://skos.um.es/unescothes/C01275</a> ) Academic achievement ( <a href="http://skos.um.es/unescothes/C00008">http://skos.um.es/unescothes/C00008</a> ) ) etc. (n=10)

Tableau 16 La notice documentaire du rapport de la direction de la Banque mondiale intitulé « *Incentives to improve teaching: lessons from Latin America* »

- Accès depuis les notices des termes du glossaire

Une notice d'un terme du glossaire offre deux rebonds.

Un premier sur les autres termes du glossaire à la signification proche. Ils sont décrits comme des termes associés (Related Terms) et sont affichés en bas de notice, le nombre envisagé est compris entre 3 et 5.

Un second sur les thèmes du plan de classement défini comme étant sémantiquement proche. Comme on l'a vu ce thème du plan de classement s'affiche comme un fil d'Ariane au dessus de la définition du terme. Il a l'avantage d'enclencher une recherche dans les notices documentaires.

- Accès depuis le contenu éditorial

Dans les spécifications données aux prestataires chargés du développement du portail final, au moins deux autres points d'accès sont envisagés.

Le premier par l'annotation par l'administrateur du site des contenus rédactionnels au moyen de thèmes sélectionnés dans le plan du classement.

Le second, optionnel par l'annotation des contenus rédactionnels par les termes du glossaire.

### **3.3.3 Optimisation des requêtes**

L'optimisation des requêtes vise à corriger le silence engendré par la non indexation des notices documentaires sur certains vocabulaires et notamment le Glossaire qui n'est pas un outil d'indexation. On en a donné une illustration en décrivant l'expansion sémantique de la requête du moteur au moyen de l'alignement des termes du glossaire avec au moins un terme descripteur du thésaurus.

### **3.3.4 Recherche à facettes**

Dans le projet de l'IIPE, la recherche à facettes forme l'éventail le plus large d'exploitation des vocabulaires pour la recherche et la découverte.

Les facettes sont des filtres qui permettent à l'utilisateur de sélectionner les résultats d'une recherche initiale sur un ou plusieurs axes proposés, ou encore d'interroger directement la base de ressources à partir d'une facette et de naviguer ensuite.

Le portail d'exploitation affiche dans ce sens un total de 10 facettes de nature différente à partir de vocabulaires de la base de connaissance et de listes contrôlées.

Facettes alphabétique, facettes hiérarchiques et extraction de métadonnées descriptives (date de publication, type de document) et facette géographique dont la particularité est d'être représentée dans le portail par une cartographie.

Toutes les facettes affichent le nombre de ressources correspondantes à l'interrogation de l'utilisateur, le moteur de recherche remettant à jour le calcul statistique des résultats au fil de la navigation et des sélections successives.

Une option de recherche par « critère » permet enfin d'interroger la base de notices en cumulant les filtres proposés.

Le tableau ci-dessous récapitule les propriétés des différentes facettes dans leur libellé en français et leur vocabulaire ou liste sources

Facettes	Type	Vocabulaire ou source
Thématique	hiérarchique	Intégralité du plan de classement avec sa structuration hiérarchique
Mots clés	hiérarchique	Extraction des seuls descripteurs du thésaurus de l'Unesco et leur structuration hiérarchique ( sauf classe pays et région)
Date de publication	date	Extraction de métadonnées
Institutions	Alphabétique	Liste contrôlée
Personnes	Alphabétique	Extraction noms d'auteur
Langues		Extraction de métadonnées
Type de document		Liste contrôlée dans ontologie
Projet	alphabétique	Classe Project
Conférence	alphabétique	Liste contrôlée
Pays	géographique	Classe Pays de l'Unesco

Tableau 17 Propriétés des différentes facettes dans leur libellé en français et leur vocabulaire ou liste sources

Au total, la configuration des facettes permet de filtrer sur la base de métadonnées qui font sens à l'utilisateur. L'ensemble des facettes donne de ce point de vue une représentation à la fois étendue et fine de la base des ressources disponibles. L'affichage du nombre de documents associés aux éléments des différentes rubriques permet de visualiser rapidement la répartition globale des ressources dans la totalité des rubriques. Assez logiquement les rubriques les plus documentées restituent très bien la prime donnée au portail thématique aux problématiques de qualité de l'éducation, d'évaluation, d'apprentissage.

Ainsi, sur la base du premier lot de notices bibliographiques importées ( n=900 ) la thématique (plan de classement) Apprentissages et évaluation en accroche 647 et le mot-clé ( descripteur du thésaurus) « Evaluation de l'éducation » 875. La répartition des ressources sur de nombreuses autres rubriques fonctionnant comme autant de possibilité données d'élargir, de changer de perspective, de cibler un pays, une période de publication etc.

Sans doute à ce stade d'avancement du projet est –il prématuré d'envisager une anticipation plus approfondie des besoins des utilisateurs qui pourrait après une période significative de mise en route inspirer un regroupement des filtres sur des bases ou des appellations différentes.

## **3.4 Bilan et préconisations**

Quel bilan tirer de cette expérience d'enrichissement des vocabulaires ? et plus généralement de la stratégie choisie par le centre de documentation de l'IPE pour valoriser ses contenus sur la qualité de l'éducation ? Quels points de vigilance souligner à la veille de la mise en ligne du portail ? Comment envisager sa pérennisation ? Quelles pistes d'enrichissement peuvent être envisagés qui tireraient pleinement parti des potentialités du dispositif en place ? Je distingue ici quelques éléments de réponses sous forme de remarques sur l'expérience menée d'une part, et de préconisations d'autre part.

### **3.4.1 Remarques**

#### **3.4.1.1 Sur la conformité du dispositif aux objectifs de l'IPE**

Le portail thématique apporte une réponse complète et innovante aux besoins formulés. Le niveau d'achèvement du dispositif d'accès aux ressources donne une vue déjà précise de la complétude de la solution mise en place et de sa pertinence. Les utilisateurs auront accès à un fonds de ressources au moyen de notices renvoyant au texte intégral des documents en phase avec la thématique de la qualité de l'éducation. Le périmètre multilingue est déjà en phase d'être élargi au-delà de l'objectif fixé des trois langues fr, en es.

Par comparaison avec la publication organisée dans les portails précédents, celui sur les plans et les politiques d'éducation Planipolis ou celui sur VIH Health and ClearingHouse, le projet introduit des innovations audacieuses.

- L'articulation d'un accès optimisé aux ressources et d'un contenu éditorial original et renouvelé (étude de controverse, focus, news, espaces de collaboration (blogs, forums, enrichissement collaboratif des termes du glossaire).
- La mobilisation après repérage, sélection, normalisation de plusieurs vocabulaires contrôlés en phase avec la thématique définie et les objectifs ambitieux formulés en matière de visibilité, d'accès facilité, de navigation, de découverte.
- La mise en place d'un dispositif logiciel ad hoc articulé autour d'un moteur de recherche capable d'exploiter les standards innovants du Web de données pour



enrichir le traitement de ses requêtes, l'affichage des résultats et donner des options de filtrages et de navigation nombreux à des utilisateurs déchargés de l'obligation de recourir aux grammaires d'interrogation pour trouver et découvrir les documents pertinents.

#### **3.4.1.2 Sur la conduite du projet : opportunité et contraintes**

La conception du portail a été rendue possible par la formulation d'un projet transversal aux différentes unités de l'IIFE. Cet ancrage dans la stratégie de moyen terme de l'organisation est à l'évidence une opportunité et un soutien qui manque parfois à tant d'autres centres de documentation. Dans le cadre de Plan4Learning, cet appui a été un stimulant de l'innovation. Pour le centre de documentation, l'opportunité a été saisie pour tenir ensemble la réponse aux attentes du programme de l'IIFE et d'autre part, l'enrichissement de ses questionnements et hypothèses relatives à son propre périmètre de compétence de gestionnaire de la collecte, du maintien et de la mise à disposition des ressources documentaires. Pour le centre de documentation Plan4Learning permet d'expérimenter ainsi une modalité nouvelle d'exploitation et de valorisation d'un segment du catalogue bibliographique. Il introduit un outil avancé de maintien des vocabulaires alignés sur des normes (Skos et Norme thésaurus ISO 25964 1-2) pleinement inscrites dans les l'environnement du web de données. Il s'adosse à un référentiel modélisé pour l'exposition et le référencement des métadonnées par les moteurs de recherche sur le web. Il s'ouvre ainsi les perspectives d'enrichissement et de partage avec d'autres sources de données à l'échelle globale.

### **3.4.2 Préconisations**

Le projet a l'avantage de s'appuyer sur un référentiel dont les standards de modélisation relèvent des technologies conçues pour l'échange, le partage, l'enrichissement des connaissances dans le web de données. Les préconisations que j'esquisse vont dans le sens de cette ouverture et de cet enrichissement.

#### **3.4.2.1 Poursuivre l'enrichissement du référentiel avec les données DBpédia**

L'enrichissement du portail par l'exploitation des données structurées de l'encyclopédie multilingue Wikipédia est une première piste. DBPédia est la plate forme qui maintient les données extraites de l'encyclopédie collaborative Wikipédia et qu'alimentent aussi bien des particuliers que des organisations. Tous les articles de l'encyclopédie font l'objet d'une extraction de métadonnées qui décrivent différents compartiments des pages. Ces métadonnées sont maintenues dans le standard RDF par DBPédia pour les rendre interrogeable par requête Sparql, et récupérable, et ré-utilisable de plusieurs manières.

#### 3.4.2.1.1 Pour l'enrichissement des contenus du portail

La solution d'annotation du service DBPédia spotlight est la plus immédiatement opératoire pour enrichir le contenu éditorial du futur site. L'annotation ou taggage des mots ( nom propre, organisation, etc.) contenus dans un texte en ligne permet la reconnaissance automatique des ressources correspondantes dans les données DBPédia et permet le renvoi et la récupération dynamique du contenu associé dans le texte.

L'éventail d'exploitation de ces contenus Wikipédia est large : affichage de définition, de traduction, de notices complètes décrivant une organisation etc. , génération dynamique de contenu etc. <sup>66</sup>

Dans le cadre du projet ces options d'enrichissement doivent être précédées d'une phase pour l'identification, la sélection, l'évaluation des correspondances recherchées entre les notices internes et les ressources cibles.

La perspective de pouvoir mettre à disposition d'un public étendu et multilingue des contenus enrichis disponibles dans un nombre supérieur aux trois langues prévues pour exploitation initiale du portail encourage à systématiser la démarche.

Le projet Joconde Lab<sup>67</sup> développé au sein du Ministère de la culture donne une illustration convaincante de ce type d'exploitation étendue de DBPédia dans une perspective de valorisation multilingue d'un fond de 300 000 notices d'œuvres d'arts alignés sur DBPédia et disponible en 14 langues.

#### 3.4.2.1.2 Pour l'exposition des référentiels.

---

<sup>66</sup> Sur son blog, Thomas Francart, met à disposition plusieurs démonstrations qui illustrent l'éventail des exploitations possibles des liens avec DBPédia pour enrichir une page html. <http://www.sparna.fr/linked-data-demo.html>

<sup>67</sup> Le projet permet au portail Joconde Lab de proposer son fonds en 14 langues. <http://www.iri.centrepompidou.fr/experimentations/joconde-lab/>

Dans le cadre de son engagement dans le projet linking open data<sup>68</sup> la plate forme collaborative offre un service de valorisation par la publication et la mise en relation de plusieurs sources de données structurées reliées. Des sources de données externes liées aux ressources de wikipédia bénéficient d'une exposition propice à la valorisation de référentiels issus de différents domaines.<sup>69</sup> Y contribuent des organisations telles que Cordis<sup>70</sup>, les programmes de la chaine britannique BBC, L'agence européenne de statistiques Eurostat, l'Insee, l'observatoire de la santé mondiale Global health etc.

En vue de l'élaboration du portail l'équipe projet a déjà amorcé cet alignement. 242 projets de la liste de projets sur 1209 pointent vers déjà vers Wikipédia. La liste des institutions est identifiée comme la seconde source de ressources qui mériterait selon la directrice d'être valorisée par ce canal de publication et de partage.

#### **3.4.2.2 Valorisation pérennisation du référentiel**

Cette perspective fait écho à plusieurs échanges qui ont porté sur les contraintes de publication actuelles du Thésaurus de l'Unesco. Bien que l'administration du Thésaurus ne relève pas de la compétence de l'IPE mais du siège de l'Unesco son exploitation dans le cadre du portail a mis en lumière l'utilité de disposer d'une version du thésaurus en SKOS. Comme on l'a indiqué plus haut [Supra 3.1.4.1] la version exploitée dans le portail est une version convertie en Skos et publiée par l'université de Murcia dont elle garde dans le référentiel l'identifiant. La prise en charge par l'Unesco d'un chargement en propre du thésaurus en Skos servirait la pérennisation du référentiel de l'IPE gérée dans ITM et ouvrirait des opportunités nouvelles quant à sa maintenance et à sa publication.

---

<sup>68</sup> [Interlinking](#)

<sup>69</sup> <http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/>

<sup>70</sup> CORDIS est le principal référentiel et portail public de la Commission européenne dédié à la diffusion des informations sur tous les projets de recherche financés par l'UE l'UE et leurs résultats au sens le plus large. [http://cordis.europa.eu/news/rcn/122038\\_fr.html](http://cordis.europa.eu/news/rcn/122038_fr.html)

### 3.4.2.3 Fiabiliser et interconnecter les identifiants auteurs et organisations

La perspective d'une exposition accélérée sur le web de données conduit à une seconde recommandation : mesurer l'avantage que l'IIPÉ aurait à adosser l'identification des auteurs et organisations gérés dans ses référentiels sur le standard international ISNI (International Standard Name Identifier, ISNI). Ce jeune standard repose sur la norme ISO 27729 :2012. Son objectif est de généraliser à l'échelle internationale l'adoption d'un identifiant (numéro) fiable et pérenne pour tout contributeur (auteur ou organisation) de contenus culturels, scientifique et technique.<sup>71</sup> Il étend la couverture des besoins qu'on attend de tout identifiant en général<sup>72</sup> au contexte contraintes et opportunités propres à l'environnement numérique et aux échanges dans le web de données.

Deux apports principaux méritent d'être soulignés

- Une identification stable<sup>73</sup> et fiable des auteurs publics : l'identifiant ISNI permet ainsi de décrire de façon unique, stable et fiable les auteurs contributeurs définis de façon large (auteurs ou organisations). La fiabilité des informations d'auteurs associés à l'identifiant est contrôlable en cas d'homonymie au moyen d'un schéma de métadonnées de référence permettant de décrire : le nom de l'identité publique,

---

<sup>71</sup> [www.isni.org](http://www.isni.org) L'exposé des caractéristiques principales de la norme s'appuie pour l'essentiel sur l'analyse qu'en donne Anila Angjeli dans [Angjeli, A. 2012, p102] et [Angjeli A., 2014]

<sup>72</sup> Comme le rappelle Anila Angjeli, [ANGJELI, 2012.] l'élaboration d'un identifiant d'une manière générale couvre des besoins et fonctionnalités bien établies : la protection des droits, le besoin des utilisateurs, la définition de la nature de l'identité qui fait l'objet de l'identification, l'universalité, l'opérationnalisation dans un environnement en réseaux permis l'actionnabilité, l'interopérabilité et la pérennité.

<sup>73</sup> La feuille de route stratégique du Ministère de la culture insiste tout particulièrement sur cet avantage d'ISNI comme une solution adéquate pour garantir le pérennité des URI qui décrivent toutes ressources sur le web de données, et dont la longévité mérite dans certains cas être supérieure à la vie de la ressource elle-même. Quel sort réserver à l'identifiant d'une ressource culturelle qui passe d'une autorité à une autre ?

les variantes du nom et les variantes d'écritures, le secteur dans lequel l'identité publique a ou a été active, son rôle dans la création, des URI pointant vers les sources de référence.

- La description des identifiants dans le cadre d'interopérabilité et d'échanges du web de données. L'attribution des ISNI alimente la base de données centralisée de tous les identifiants internationaux formalisés en RDF qui en contient plus de 8 millions. Le standard autorise ainsi l'échange et les interconnexions entre différentes sources de données.

La gouvernance centralisée de l'identifiant est assurée par l'agence ISNI. Elle s'appuie sur une agence d'attribution d'identifiants qui est OCLC Online Computer Library Center<sup>74</sup>) Une agence qualité assurée par la Bnf et la British Library et d'agences d'enregistrements. L'infrastructure s'appuie sur une base de données qui contient déjà plus de 8 millions d'identifiants en RDF et interrogeable en XML ou via une interface de programmation (API)

---

<sup>74</sup> L'OCLC est basé à Dublin aux Etats-Unis ce centre a également en charge la gestion du fichier d'autorité virtuel des auteurs VIAF.  
[http://fr.wikipedia.org/wiki/Online\\_Computer\\_Library\\_Center](http://fr.wikipedia.org/wiki/Online_Computer_Library_Center)



# Conclusion

L'objectif de cette étude était de proposer une mise en lumière d'un projet original de valorisation d'un fonds de ressources documentaire. Plus spécialement, il s'agissait de mettre l'accent sur la façon dont l'exploitation de vocabulaires (thésaurus de l'Unesco, Plan de classement, listes d'autorités) pouvait au moyen d'un portail web multilingue, optimiser la mise à disposition, la visibilité, la recherche et la découverte des ressources documentaires issues du catalogue de la bibliothèque de l'IPE- UNESCO.

Par contraste avec les portails thématiques dont la bibliothèque a déjà une solide expérience, le portail Plan4Learning expérimente la mobilisation de ressources inédites dont il tire son originalité. En matière de gestion (*back-office*) le dispositif s'adosse à un système avancé de gestion de thésaurus et de vocabulaires ; il s'appuie sur la modélisation des ressources dans les standards du web sémantique et l'organisation d'un échanges de données structurées entre un outil d'indexation de notices documentaires et un moteur de recherche sémantique.

En *front-office* le projet recouvre une dimension éditoriale conséquente dans laquelle prennent place les services élaborés pour la recherche, la navigation, la découverte, la visibilité et l'enrichissement collaboratif des ressources au sein de contenus d'information renouvelés.

L'intérêt de traiter d'un segment même partiel de ce projet en cours de développement se mesure à la prise concrète qu'il nous a donné sur les problématiques, les contraintes pratiques autant que sur la part fortement novatrice du dispositif retenu.

En effet, à l'heure où l'interrogation sur les contours de métiers de l'information et de la documentation en prise avec la concurrence des technologies numériques continue d'animer les institutions professionnelles et académiques, le projet Plan4Learning tel que nous l'avons abordé apporte quelques éclairages utiles. Nous les résumons sous la forme de deux séries d'observations.

La première est rétrospective. La genèse du projet nous a permis de souligner tout d'abord comment ce dispositif procède d'une prise en charge audacieuse d'une priorité formulée à titre transversal par l'IPE-Unesco dans le but de proposer un apport facilité de contenu et de ressources documentaires ciblées sur la thématique de la qualité des apprentissages. A ce titre, il illustre à une échelle certes mesurée un cas d'intrication étroite des stratégies d'information et de documentation dans les objectifs généraux d'une institution internationale spécialisée. Cette synergie a pu fonctionner comme un précieux levier d'innovation pour explorer une mobilisation différente de ses ressources et de ses outils documentaires (notices bibliographiques, référentiels, métadonnées descriptives) et l'introduction de technologies avancées. Cette mise en cohérence participe du même coup de l'élaboration de nouveaux services allant au-devant des besoins des utilisateurs. Elle n'est pas sans effet sur la conduite opérationnelle du projet et l'extension des compétences métiers mobilisées. La conduite du projet s'est tout d'abord manifestement insérée dans la



gestion ordinaire des nombreuses tâches et missions routinières que doivent assurer les personnels de la bibliothèque. Dans cette configuration, le travail des documentalistes a également exploré des territoires nouveaux et des allers et retours inévitables entre la préparation et la sélection des ressources, l'importation et la validation de vocabulaires et l'anticipation et la préparation de leur exploitation optimale dans un cadre éditorial. L'articulation des échanges organisés entre les référentiels et le moteur de recherche sémantique a introduit un dispositif technique d'accès à l'information qui tranche avec un environnement dans lequel les documentalistes disposent d'une complète maîtrise des ressources, de la manière de les interroger et d'optimiser les résultats de requêtes. Ces remarques permettent d'attirer l'attention sur la renégociation diffuse des frontières des compétences et des expertises métier induite par les technologies numériques en matière d'accès à l'information.

Le projet appelle enfin une observation quant à ses prolongements. Au-delà du portail Plan4Learning, le dispositif introduit pour la gestion et l'interopérabilité des référentiels ouvre les voies du partage, de l'enrichissement, de la publication et de la ré-utilisation des jeux de données au sein du web de données. Une littérature de plus en plus nourrie documente les perspectives d'enrichissements et de renouvellement des dispositifs documentaires ouvertes par le recours aux identifiants pérennes et à la structuration sémantique des données. Cette piste recouvre l'inscription de données sur le web à des fins de visibilité des catalogues bibliographiques et de référencement par les robots des moteurs de recherche, voire de navigation entre jeux de données issues de sources différentes. Au-delà, elle ouvre, par leur interopérabilité et leur ré-utilisabilité la voie à des enrichissements et à la conception de services nouveaux. Comme le souligne Emmanuelle Bermès, « l'intérêt de l'exposition des données dans le Web de données ne réside pas seulement dans la navigation en parcourant les liens d'un jeu de données à un autre, mais dans la ré-utilisation de ces données pour créer de nouvelles applications. » [6, BERMES, p 133].

A l'échelle du projet, une première amorce de lien vers l'entrepôt de métadonnées Dbpédia atteste de l'intérêt de l'IIPÉ-UNESCO pour cette démarche. Elle a porté pour l'instant sur une liste contrôlée d'organisations et l'examen d'autre référentiel candidat et notamment la liste des projets également centrés sur la gestion et l'évaluation de l'éducation. Cette démarche esquisse ce qui pourrait donner lieu à un projet de publication et d'enrichissement de données à part entière. Il intégrerait l'évaluation des référentiels candidats, la nature des enrichissements souhaités, l'examen des sources cibles ainsi que les besoins en termes de services nouveaux.

L'objectif de cette étude était de proposer une mise en lumière d'un projet original de valorisation d'un fonds de ressources documentaire.

L'accent placé sur l'exploitation des vocabulaires a permis de saisir les modalités concrètes d'un dispositif innovant qui participe du renouvellement en cours de l'offre de médiation et de prestations des bibliothèques qu'encourage par la généralisation du web et des moteurs de recherche généralistes.

Le périmètre de l'étude a délibérément été restreint aux phases de finalisation d'un projet en cours de déploiement. Il donne à voir les conditions institutionnelles d'émergence du projet ainsi que les contours concrets de sa mise en oeuvre à l'échelle d'un service d'information.

Il donne aux professionnels de l'information et de la documentation une prise directe sur les aspects organisationnels, logiciels et documentaires d'une architecture retenue pour

# Bibliographie

---

La présente bibliographie analytique et sélective contient trente références réparties en cinq sections :

1. Sur le Web sémantique
2. Le web sémantique et les politiques documentaires
3. Langages documentaires et référentiels
4. Accès à l'information et nouvelles médiations
5. Travaux sur l'IIPE –UNESCO

A l'intérieur de chaque section les références sont présentées par ordre alphabétique d'auteurs. Chaque référence est augmentée d'un commentaire mentionnant l'apport pratique ou de réflexion que nous avons retenu pour conduire le présent travail.

## 1. Sur le web sémantique

1. BERMÈS, Emmanuelle. Convergence et interopérabilité : l'apport du web de données. 77e Congrès de l'IFLA, San Juan de Porto Rico, 13-18 août 2011 [en ligne]. <http://conference.ifla.org/past-wlic/2011/149-bermes-fr.pdf>

*L'interopérabilité du web de données donne aux institutions culturelles (bibliothèques, archives, musées) un levier pour renouveler la mise à disposition des ressources et anticiper les besoins des utilisateurs allant bien au-delà des protocoles d'échanges de notices en vigueur ( Z 39.50) et de mutualisation de données (OAI PMH). Dans cette contribution, l'auteur insiste sur la place nouvelle conférée aux référentiels (valeurs ou métadonnées) dans la constitution et la mise en relation de cet espace global d'information.*

2. GANDON F. ; FARON-ZUCKER C. ; CORBY O. Le web sémantique : comment lier les données sur le web ? Paris : Dunod, 2012. 206 p. ISBN 978-2-10-057294-6

*Cet ouvrage est une entrée générale dans les couches technologiques du web de sémantique. Il assume un statut intermédiaire entre le manuel d'introduction et le guide technique à destination des développeurs. Une contextualisation et de*

*nombreux exemples d'applications dans les sphères publiques et privées illustrent l'utilisation des principaux standards.*

3. VATANT Bernard. Des métadonnées à la description des ressources : les langages du web sémantique. In CALDERAN Lissette, HIDOINE Bernard, MILLET Jacques dir. Métadonnées : mutations et perspectives. Paris : ADBS éditions, 2008. Chapitre 5, p.163-194. ISBN 978-2-84365-104-5

*Cette contribution dense offre une mise en regard des standards du web de données avec les notions connexes de ressources et de métadonnées. La force didactique du propos tient au rappel qui est fait des discussions qui ont jalonné les débats du W3C autour de l'adoption et la compréhension des principaux standards du web sémantique (URI / HTTP ; RDF, RDFS, OWL etc.). L'auteur nourrit ses démonstrations de préconisations pratiques inspirées d'une expérience professionnelle à la tête de la société Mondeca.*

## **2. Web sémantique et politiques documentaires**

4. ANGJELI Anila. ISNI : un identifiant passerelle. Documentation et Bibliothèques, Asted Inc, 2012, vol. 58, n° 3, p. 101-108. ISSN 0315-2340

*Cet article retrace la genèse et les enjeux de la mise en place d'une gouvernance centralisée de ce nouvel identifiant ainsi que son inscription délibérée dans l'environnement nouveau des données liées. L'article détaille l'articulation prévue entre cet identifiant international des agents avec le fichier d'autorité des personnes (VIAF) ainsi que les modalités concrètes de son déploiement.*

5. ANGJELI Anila. La normalisation en bibliothèque à l'heure du web sémantique. Documentaliste - science de l'information, 2011, vol.48, n°4, p. 53-54

*Les standards du web sémantique intègrent et refaçonnent le travail de normalisation à l'œuvre dans le monde des bibliothèques. L'auteur livre une synthèse des normalisations telle que Dublin Core, RDA et ISNI dont les résonances avec les horizons de partage et d'interopérabilité du web de données illustrent la recherche de nouveaux consensus l'échelle internationale.*

6. BERMÈS, Emmanuelle. Le web sémantique en bibliothèque. Antoine Isaac, Gautier Poupeau, collaborateurs. Paris : Electre éditions du cercle de la librairie, 2013. 171 p. ISBN 978-2-7654-1417-9

*Dans cet ouvrage sont rappelés les enjeux et promesses d'interopérabilité et d'ouverture des technologies du web sémantique pour les stratégies d'ancrage des institutions culturelles et de leurs ressources dans le web. L'exposé des technologies s'accompagne de l'examen détaillé sous forme de cas pratiques des différentes modalités d'application aux fins de publication, d'enrichissement et de réutilisation des données par lesquelles les bibliothèques et leur responsable peuvent engager la mise en place de nouveaux services aux utilisateurs.*

7. BOURDONET Philippe. L'espace documentaire en restructuration : l'évolution des services des bibliothèques universitaires. 2013. 533 p. Thèse, Science de l'information et de la communication, CNAM, 2013

*A la lumière de l'observation des services des bibliothèques universitaires, l'auteur décrit la crise de désaffection que connaît la fréquentation de leurs catalogues concurrencés qu'ils sont par d'autres points d'accès à l'information. La démonstration s'élargit par l'examen de l'ensemble de la médiation numérique qu'affecte la généralisation des standards du web et les nouvelles pratiques des utilisateurs. Cette thèse aide notamment à comprendre in concreto comment la nouvelle modélisation des notices bibliographiques et l'interopérabilité du web sémantique constituent une issue privilégiée mais encore naissante dans laquelle les professionnels s'emploient, par l'exposition et la ré-utilisabilité de leur ressources à reprendre la main sur la maîtrise de la valeur ajoutée de l'information bibliographique et regagner la confiance des utilisateurs au moyen d'une visibilité plus grande et l'élaboration de nouveaux services.*

8. CHARTRON, Ghislaine. La valeur des services documentaires en prise avec le numérique. Bulletin des bibliothèques de France [en ligne], n° 5, 2012 [consulté le 11 janvier 2015] <<http://bbf.enssib.fr/consulter/bbf-2012-05-0014-003>>. ISSN 1292-8399.

*L'auteur recense les grands enjeux associés à l'évolutivité des technologies introduites par le numérique et ses effets sur les contours et la nature des activités*

*des professionnels de l'information pris comme médiateurs de plus en plus investis dans les objectifs généraux de leur organisation.*

9. MINISTERE DE LA CULTURE ET DE LA COMMUNICATION. Feuille de route stratégique, Métadonnées culturelles et transition Web 3.0. [en ligne]. Paris, version 7 novembre 2013. <<http://cblog.culture.fr/projet/2013/11/07/groupe-de-travail-metadonnees-culturelles>>

*Ce document donne une vue actualisée sur l'apport des technologies du web sémantique dans la définition de quelques actions prioritaires en matière d'inscription des politique culturelle sur le web : interconnexion des référentiels culturels dans un « Graphe Culture », renouvellement des échanges avec le public, soutien d'une R&D dédiée, formation interne, etc. Parmi les neuf actions décrites, sont en particulier détaillés l'enjeu et l'utilité de pouvoir appuyer les référentiels sur des identifiants uniques et pérennes tels que permis s'agissant des auteurs le nouvel identifiant international ISNI.*

10. W3C. Rapport final du groupe d'incubation « Bibliothèques et web de données » Groupe d'incubation W3C, du 25 octobre 2011, version traduite en français, Septembre 2012 [En ligne] <<http://mediatheque.citemusique.fr/MediaComposite/ARTICLES/W3C/XGR-Ild-fr.html>>

*Un état des lieux complet et richement documenté des apports des technologies du web sémantique à la valorisation des ressources des bibliothèques sur le web. L'intérêt de ce travail tient à la nature collective professionnelle et internationale de son élaboration ainsi qu'à son périmètre circonscrit à l'environnement des bibliothèques de leur prestataires (développeurs et éditeurs de logiciel) et de leur utilisateurs. Un état des lieux des limites et défis rencontrés par les bibliothèques accompagne l'exposition claire des avantages clés des technologies sémantiques pour l'identification pérenne, le partage, l'enrichissement et la ré-utilisation des jeux de données formalisées par ces standards.*

### **3. Langages documentaires et référentiels**

11. COTTE Dominique, Les nouvelles perspectives du web sémantique pour les professionnels de l'information. Documentaliste - science de l'information, 2011, vol.48, n°4, p. 58-59

*Cet article prend la mesure des effets du paradigme du web de données sur les modalités d'exercice des tâches « classiques » (indexation, structuration, normalisation, qualification des ressources) de l'information et de la documentation et la recomposition des frontières professionnelles que cette mutation induit.*

12. FAVIER Laurence ; MUSTAFA EL HADI Widad. L'interopérabilité des systèmes d'organisation des connaissances : une nouvelle conception de l'universalité des savoirs ? In PAPY Fabrice, dir. Recherches ouvertes sur le numérique, approches pratiques en information-communication. Paris : Hermès Science publication, 2013. Chapitre 1, p.23-38. ISBN 978-2-9462-4535-8

*Dans cette contribution, l'exploitation de liens entre des langages documentaires hétérogènes est présentée comme un enjeu grandissant en matière d'aide à la recherche à l'heure de la convergence des sources d'information sur le Web et les besoins en information pluridisciplinaire et multiculturelle. Les auteurs rappellent les contributions de la recommandation SKOS et de la norme thésaurus ISO 25 964 à l'interopérabilité sémantique et aux passerelles entre différents langages documentaires. Des dispositifs expérimentaux comme les visual catalog donnant à l'utilisateur la possibilité de consulter plusieurs catalogues en manipulant différents langages (Rameau, MotsBis, CDD ) et l'extension culturelle de la notion d'interopérabilité illustrent des modalités nouvelle d'accès unifié aux ressources. L'extension de la notion d'interopérabilité des langages documentaires à une dimension culturelle pourrait ainsi rejoindre à l'heure du numérique l'ambition des concepteurs des langages universels.*

13. HUDON, Michèle. ISO 25964 : pour le développement, la gestion et l'interopérabilité des langages documentaires. Documentation et Bibliothèques, 2012, vol. 58, n° 3, p. 130-140. ISSN 0315-2340

*Un article utile pour comprendre la genèse de la nouvelle normalisation sur les thésaurus ISO 25964-1 et ISO 25964-2. L'auteur en rappelle les enjeux et retrace les différentes initiatives et discussions qui ont conduit à la remise sur le chantier de cette normalisation.*

14. HUDON Michèle. Guide pratique pour l'élaboration d'un thésaurus documentaire. Danièle Degez, Dominique Ménillet, collaborateurs. Québec : Les éditions Asted, 2009. 274 pages. ISBN 978-2-923563-17-6

*Dans ce guide, l'auteur rappelle les règles fondamentales qui régissent la structuration des thésaurus. Il offre une entrée de référence dans la singularité de ces instruments d'organisation des connaissances et une méthodologie pour leur élaboration et leur maintenance. La présente édition n'intègre pas en revanche les évolutions introduites par la norme ISO 25964.*

15. ISAAC Antoine. Entre thésaurus et ontologie : une affaire d'interopérabilité et d'alignement. Documentaliste - science de l'information, 2011, vol.48, n°4, p. 48-49

*Une contribution éclairante sur le rôle d'intermédiaire entre les ontologies et les référentiels joué par la représentation de concepts Skos. Au sein du web sémantique Skos est présenté comme le maillon essentiel pour la publication, l'échange et l'interconnexion des référentiels documentaires.*

16. ISAAC, Antoine. Les référentiels : typologie et interopérabilité. Séminaire IST Inria : le document numérique à l'heure du web de données, Oct 2012, Carnac, France. ADBS, pp.85-104, Sciences et techniques de l'information; Le document numérique à l'heure du web de données. <hal-00740282v1>

*L'auteur clarifie la notion peu stabilisée de référentiel à l'heure du web. Il brosse une typologie des objets de description qui sont susceptibles de tenir ce rôle. IL distingue les éléments de métadonnées ( ontologies RDFS, OWL, FOAF, FRBR, Dublin Core etc.), les vocabulaires de valeurs et les systèmes d'organisation de connaissances, autres jeux de données (thésaurus, schémas de classification, listes d'autorités, etc.) et les jeux de données de référence ( graphes RDF de Dbpédia, etc.). L'auteur introduit ensuite les opérations nouvelles que l'environnement du web de données permet d'envisager pour chacune de ces familles de référentiels (ré-utilisation ; édition distribuée ; alignement). Dans le cas particulier des Kos de type thésaurus, l'auteur rappelle l'emploi souvent complémentaire d'alignement manuel et automatique pour gérer les aspects sensibles que sont le multilinguisme et l'hétérogénéité des libellés en présence.*



17. ISO 25964-2 Mars 2013. Information et documentation – Thésaurus et interopérabilité avec d'autres vocabulaires – Partie 2 : Interopérabilité avec d'autres vocabulaires. Paris : AFNOR, 2013. 96 p.

*Le second volet de la norme retenu ici pour ses préconisations en matière de mise en correspondance avec d'autres formes de vocabulaires.*

18. RUIZ LEPORES Domingos. Des grandes classifications au Web de données et l'émergence de l'indexation sémantique : le cas du tagging sémantique dans le portail Histoire des Arts. 2011. 118 pages. Mémoire, Ingénierie de la documentation, INTD-CNAM, 2011.

*Sur la base d'un travail d'annotation collaborative conduit au sein du portail sur l'histoire des arts HDA Lab, l'auteur met en lumière l'apport de cette indexation sémantique comme prolongement des instruments historiques d'indexation documentaires.*

19. VATANT Bernard, «Porting library vocabularies to the Semantic Web, and back. A win-win round trip ». IFLA 76e Congrès, August 2010, Göteborg [en ligne]. Consulté le [30 octobre 2014] <<http://www.ifla.org/files/hq/papers/ifla76/149-vatant-en.pdf>> <http://conference.ifla.org/past-wlic/2010/149-vatant-en.pdf>

*Une contribution qui vise à réconcilier les approches longtemps opposées des vocabulaires d'indexation « traditionnels » et des logiques autonomes des moteurs de recherche. Les standards du web sémantique sont présentés comme un nouveau terrain d'échange et d'enrichissement croisés entre le monde du web et celui des professionnels de l'information-documentation.*

#### **4. Accès à l'information et nouvelles médiations**

20. DELAHOUSSE Jean, Apports d'une ontologie de domaine aux services d'accès aux contenus. In Grivel, Luc Directeur. La recherche d'information en contexte. Paris : Hermès Science publications, 2011.186 p. ISBN 878-2-7462-2581-7

*Dans ce chapitre l'auteur introduit la notion de taxonomie de publication comme instrument d'accès aux contenus générés à partir d'ontologies de domaine. il donne des recommandations de mise en oeuvre et de maintenance des référentiels*

*terminologiques. La recherche à facettes est abordée comme mode d'exploitation fondamentalement éditoriale de ces taxonomies.*

21. DELAHOUSSE Jean. L'extension des capacités des moteurs de recherche par l'utilisation de terminologies métier ou comment rendre les moteurs de recherche plus efficaces? [Blog Mondeca - Leçons de Choses]. [consulté le 12 sept 2014] <<http://mondeca.wordpress.com/2007/10/06/l%e2%80%99extension-des-capacites-des-moteurs-de-recherche-par-l%e2%80%99utilisation-de-terminologies-metier/>>

*Une approche pratique de l'extension sémantique des recherches au moyen de terminologies et un rappel de l'importance de conjuguer terminologies métiers et expressions des utilisateurs dans l'alimentation et la mise à jour des index des moteurs de recherche. L'auteur présente également la question de la visibilité à donner aux extensions effectuées par le moteur comme un choix à graduer selon que les extensions programmées modifient profondément ou non le sens de la requête de l'utilisateur.*

22. MARKEY, Karen. The Online Library Catalog : Paradise Lost and Paradise Regained ? [En ligne]. D-Lib Magazine, 2007, vol. 13, 1/2, janvier 2007. <<http://www.dlib.org/dlib/january07/markey/01markey.html>>.

*L'intérêt de cette tribune datée est de marquer un moment de la réflexion des professionnels des bibliothèques sur la crise de désaffection qui touche les catalogues en ligne à l'heure de la généralisation de la numérisation des ressources et de l'ascension triomphale des moteurs de recherche de type Google ainsi que les premières voies de sorties esquissées.*

23. LALLICH-BOIDIN Geneviève ; MARET Dominique. Recherche d'information et traitement de la langue. Fondements linguistiques et applications. Villeurbanne : Presses de l'enssib, 2005, 288 p. ISBN 2-910227-60-X

*Le chapitre 9 de cet ouvrage en particulier expose les principes généraux de la recherche d'information en langage naturel et les modalités de traitement de la recherche intégrale par les moteurs de recherche et les langages d'interrogation sous-jacents par les opérateurs booléens.*

24. MINEL, Jean-Luc ; POUYLLAU, Stéphane. ISIDORE : l'accès unifié aux données numériques des sciences humaines et sociales. Journées ABES [en ligne]. 17 mai 2011, Montpellier [En ligne]

[http://m.abes.fr/content/download/1188/5453/version/1/file/Session2-1-isidore\\_Jabes11.pps](http://m.abes.fr/content/download/1188/5453/version/1/file/Session2-1-isidore_Jabes11.pps), consulté en 2013

*Ce document propose une présentation complète et détaillée de l'architecture et du fonctionnement de la plate forme sur les sciences humaines et sociales Isidore.*

25. POUYLLAU Stéphane, Les moteurs de recherche profitent aussi de la sémantique. Documentaliste - science de l'information, 2011, vol. 48, n°4, p. 36-37

*A la lumière de l'expérience de la plate-forme de recherche en SHS Isidore, l'auteur revient sur l'apport de la structuration des données et des données liées sur la performance des moteurs de recherche ainsi que sur les promesses de ce couplage dans le domaine des données publiques.*

26. SIMMONOT Brigitte, L'accès à l'information en ligne. Moteurs, dispositifs et médiations. Cachan : Hermès Science Lavoisier, 2012. 249 p. ISBN 978-2-7462-3829-9 ISSN 1967-8409

*Une analyse stimulante des dispositifs d'accès à l'information, de leur histoire, de leur transformation et de leur relation avec les pratiques elles même changeantes d'accès et de recherche d'information en ligne. Les chapitres 7 à 8 traitent en particulier de la nature hybride des médiations humaines et techniques à l'œuvre dans les processus de mise en relation pour l'accès à l'information. Les formes de médiations nouvellement introduites par les moteurs de recherche commerciaux s'expriment au moyen d'interface, d'ergonomie qui interrogent autant les pratiques d'informations que le statut et l'autorité éditoriale pas franchement assumée des dispositifs, leur rapport à la transparence de leur traitement et la forme que pourrait prendre le réinvestissement des utilisateurs sur ces outils et sur l'évaluation de leur pertinence.*

## **5. Travaux sur l'IPE**

27. BERGONNIER SIGNOLES Aurélie. Mise en œuvre d'une nouvelle structuration des données pour valoriser le fonds documentaire de l'IPE UNESCO : l'expérience FRBR. 2011. 152 pages. Mémoire, Ingénierie documentaire, INTD-CNAM, 2011.

*L'auteur livre une évaluation serrée de l'opportunité et de la pertinence de la mise en œuvre de la modélisation FRBR dans la stratégie de valorisation des ressources documentaires de l'IPE. L'expérimentation s'inscrit dans le contexte du renouvellement des pratiques d'accès et de recherche d'information des utilisateurs et de la concurrence des canaux d'information stimulée par le web.*

28. SOBRAL BOYER Maria. Etude de solutions dans le monde du logiciel libre pour la ré-informatisation d'un centre de documentation tourné vers les pays en développement : Cas du centre de documentation de l'IPE-UNESCO. 2010. 146 p. Mémoire, Titre professionnel de niveau 1 « Chef de projet en ingénierie documentaire », INTD-CNAM, 2010.

*A la faveur d'une étude sur la ré-informatisation du centre de documentation de l'IPE-UNESCO, ce travail apporte un éclairage complet sur les problématiques documentaires du centre de documentation et l'utilité du remplacement de l'outil CDS-ISIS de gestion utilisé pour la gestion des bases de données. Les recommandations s'appuient sur une évaluation comparée des outils open source de remplacement.*

29. POLEWKA Anna. Étude des enjeux de la virtualisation des collections du centre de documentation de l'IPE-Unesco. 2012. 109 p. Mémoire pour obtenir le Titre professionnel « Chef de projet en ingénierie documentaire », INTD, 2012.

*La politique documentaire du centre de documentation est analysée dans cette étude à la lumière de l'étude des publics locaux et distants du centre. L'opportunité de la virtualisation des collections est mesurée à l'aune des enjeux en terme de gestion, de préservation et de renouvellement des accès aux ressources.*

30. TRAYNOR, Mary. 2013. La contribution d'un centre de documentation à la gestion de la mémoire organisationnelle : le cas de l'IPE-UNESCO. 2013. 157 p. Mémoire pour obtenir le Titre professionnel « Chef de projet en ingénierie documentaire », INTD, 2013.

*Ce travail original porte sur l'étude de la contribution spécifique que le centre de documentation et son équipe peuvent verser à l'organisation et la gestion des connaissances de l'IIPÉ et à leur transmission à l'échelle de l'organisation. Un cas pratique de Knowledge Management (KM) dans lequel le centre de documentation pourrait également verser son expérience et ses savoirs faire accumulés au fil du temps.*