



HAL
open science

Vers une correspondance entre langages maison et classification décimale de Dewey dans le cadre d'une utilisation de la norme LOM-FR

Marie Peterlongo

► **To cite this version:**

Marie Peterlongo. Vers une correspondance entre langages maison et classification décimale de Dewey dans le cadre d'une utilisation de la norme LOM-FR. domain_shs.info.comm. 2007. mem_00000566

HAL Id: mem_00000566

https://memsic.ccsd.cnrs.fr/mem_00000566

Submitted on 17 Nov 2007

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.

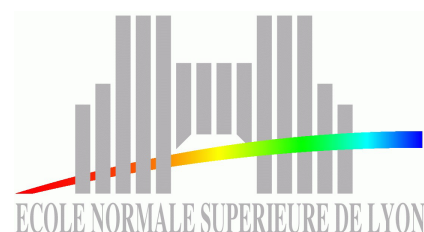


Faculté de Lettres et Civilisations

Université Jean Moulin

Lyon 3

**Vers une correspondance entre langages
maison et Classification décimale de Dewey
dans le cadre d'une utilisation de la norme
LOMFR**



Par Marie PETERLONGO

Tuteur en entreprise : Gérard VIDAL

Tuteur universitaire : Thomas KRECZANIK

Master 2 Stratégies documentaires et valorisation de l'information

Année universitaire : 2006-2007

Résumé : A l'ENS Sciences de Lyon et à l'INRP, des scientifiques et des enseignants produisent des ressources pédagogiques numériques à destination de leurs collègues. Dans ce contexte de production de ressources, le schéma de métadonnée LOMFR est utilisé. Au sein d'une des balises du schéma de métadonnée les ressources sont indexées grâce à des langages maison ou taxinomies différentes selon les disciplines. La mission de stage relaté dans ce rapport a consisté à proposer des améliorations des taxinomies maison et à effectuer des correspondances entre celles-ci et la classification décimale de Dewey comme le préconise la norme LOMFR.

Abstract : Scientists and teachers of the ENS Sciences and INRP write digital pedagogic documents for their colleagues. They use the LOMFR (Learning Object Metadata) metadata model where ressources are indexed with local (homemade) languages or taxinomies specific to each discipline in one of the tags of the schema. This paper deals with my training course where I was in charge of proposing improvements for the local taxinomies and linking them with the Dewey Decimal Classification as specified in the LOMFR standard.

Remerciements

G rard Vidal et Thomas Kreczanik, pour leurs conseils, leur pr sence et leur r activit .

Les enseignants/scientifiques avec qui j'ai eu l'occasion de travailler: Marie-Christine Artru, Fran oise Jauzein, Sandrine Heusser, Catherine Simand-Vernin, Charles Henri Eyraud, Olivier Dequinsey, Pierre Thomas, Hagop Demerdjan, Romain Barbe, Jean-Bernard Baudin.

Les biblioth caires et documentalistes qui m'ont aid    me familiariser avec le sujet et m'ont aid  sur le plan pratique de leurs conseils: la conservatrice de la biblioth que de l'ENS sciences de Lyon, Jacqueline De Condappa, le conservateur de la biblioth que de l'INRP, Philippe Marcerou, la directrice de la biblioth que de l'INSA de Lyon, Monique Joly. Leurs coll gues qui m'ont conseill  sur l'utilisation de la CDD Sophie Wagnon   l'INRP et Fran ois Deboclar   l'ENS. Les personnes que j'ai rencontr es au CNDP: Rosa Maria Gomez de Regil qui a  t  l'interm diaire pour cette rencontre, Sandrine Jambert qui a organis  cette r union, Teresa Bremand, Fran ois Feyler et Sylvie Dalbin qui m'ont beaucoup aid e par leurs conseils.

Merci aussi   S bastien Pilloz pour l'assistance technique lors de la r daction de mon m moire.

Table des matières

| | | |
|-------|---|----|
| 1 | Introduction..... | 6 |
| 2 | Présentation..... | 8 |
| 2.1 | Institutions et services | 8 |
| 2.1.1 | Présentation de l'INRP (Institut National de Recherche Pédagogique)..... | 8 |
| 2.1.2 | ACCES (Actualisation Continue des Connaissances des Enseignants en Sciences)..... | 9 |
| 2.1.3 | ENS Sciences de Lyon..... | 10 |
| 2.1.4 | Pr@tic..... | 11 |
| 2.2 | Ressources pédagogiques | 11 |
| 2.2.1 | Ressources pédagogiques..... | 12 |
| 2.2.2 | Espaces collaboratifs..... | 14 |
| 2.4 | Outils de travail..... | 17 |
| 3 | Missions..... | 19 |
| 3.1 | Contexte de la mission: l'utilisation du LOMFR..... | 19 |
| 3.1.1 | La production de documents..... | 19 |
| 3.1.2 | Les Métadonnées..... | 20 |
| 3.1.3 | Le LOM (Learning Object Metadata)..... | 21 |
| 3.1.5 | Les problèmes d'utilisation du LOM..... | 25 |
| 3.1.6 | Balise 9 du LOM mode d'indexation | 26 |
| 3.1.7 | Profils d'application et utilisation du LOMFR à l'INRP et l'ENS.... | 27 |
| 3.2 | Analyse des langages maison..... | 30 |
| 3.2.1 | Origine des taxonomies..... | 30 |
| 3.2.2 | Thésaurus/Taxonomies: définitions des terminologies "maison" | 30 |
| 3.2.4 | Proposition pour des améliorations..... | 42 |
| 3.2.5 | Taxonomies, thésaurus et ontologies : interrogations, recherches, perspectives..... | 45 |
| 3.3 | Vers des correspondances | 45 |
| 3.3.1 | Pourquoi des correspondances?..... | 45 |
| 3.3.2 | Origine du travail de correspondance avec la CDD et définition de la méthodologie..... | 46 |
| 3.3.3 | Élaboration de la méthodologie : étude du projet de classification UNIT..... | 47 |
| 3.4 | Réalisation des correspondances..... | 49 |
| 3.4.1 | Accès et version de la CDD | 49 |
| 3.4.2 | Méthodologie des correspondances | 49 |
| 3.4.3 | Correspondances entre taxonomies disciplinaires et les thésaurus de l'éducation..... | 50 |

| | | |
|---------|---|----|
| 3.4.3.1 | ETB (European Treasury Browser)..... | 51 |
| 3.4.3.2 | Thésaurus Mot Bis..... | 51 |
| 3.4.3.3 | Méthodologie..... | 51 |
| 3.5 | Bilan du travail de correspondances..... | 52 |
| 3.5.1 | Un Bilan chiffré..... | 52 |
| 3.5.2 | Validation scientifique..... | 53 |
| 3.5.2.1 | Méthodologie..... | 53 |
| 3.5.2.2 | Particularités et richesses de cette collaboration..... | 54 |
| 3.5.3 | Rencontre CNDP..... | 54 |
| 3.5.4 | Difficultés et interrogations rencontrées lors du travail de correspondances..... | 55 |
| 4 | Note de recherche..... | 58 |
| 4.1 | Analyse de la conception des taxonomies et des pratiques d'indexation | 58 |
| 4.1.1 | Caractéristiques des taxonomies disciplinaires..... | 58 |
| 4.1.2 | Une élaboration collaborative..... | 59 |
| 4.2 | Vers une interopérabilité des systèmes de descriptions de ressources?..... | 60 |
| 4.2.1 | L'interopérabilité entre les systèmes de descriptions des ressources | 60 |
| 4.2.1.1 | Définition de l'interopérabilité..... | 60 |
| 4.2.1.2 | Comment organiser cette interopérabilité? L'exemple du projet de correspondance et ses limites..... | 62 |
| 4.2.1.3 | Comment organiser l'interopérabilité? Quelles sont les solutions possibles?..... | 64 |
| 4.2.2 | Quelle est la place des usagers ?..... | 65 |
| 4.2.3 | Enjeux/Collaborations..... | 66 |
| 5 | Conclusion générale..... | 68 |
| 6 | Bibliographie..... | 70 |
| | Sur la documentation pédagogique en général :..... | 70 |
| | Sur le LOM et les métadonnées..... | 70 |
| | Sur l'interopérabilité..... | 71 |
| | Les langages documentaires..... | 72 |
| | Les ontologies..... | 73 |
| 7 | Annexes..... | 74 |
| | Exemple de Fiche LOM de Culture Science physique remplie par Catherine Simand..... | 74 |
| | Langages documentaires et ontologies..... | 78 |
| | Classification | 78 |
| | Thésaurus | 78 |
| | Ontologie | 79 |
| | Les ontologies : concepts et représentations | 79 |
| | Web sémantique | 81 |
| | Ontologies et thésaurus..... | 82 |
| | Bibliographie : | 84 |
| | Tableau de correspondance..... | 87 |

1 Introduction

A l'ENS (École Normale Supérieure) science de Lyon et à l'INRP (Institut National de Recherche Pédagogique) des équipes d'enseignants détachés produisent des ressources documentaires à destination de leurs collègues. Ces documents sont publiés sur des sites internet accessibles à tous.

Cette politique de production de ressources a amené à la constitution d'un fond important et divers. Il a été rapidement nécessaire de mettre en place un outil afin de retrouver ces documents. Après réflexion c'est en 2003 qu'une première terminologie destinée à indexer des documents en sciences de la Terre a vu le jour sous le nom de ThoKaVi. Ce premier mouvement a été suivi par l'élaboration d'outils semblables par les enseignants d'autres disciplines à savoir en sciences de la Vie avec le JaHe, en physique avec le CheBoCar et en chimie avec le TaCBaHa. Ces premières terminologies ont été élaborées par les scientifiques eux-mêmes, ils les appelaient alors thésaurus.

Ce mouvement pour une indexation des ressources a été effectué dans un contexte d'utilisation du LOMFR. Ce schéma de métadonnée comprend une balise devant être remplie par au moins une classification. De cette façon elle peut être renseignée par un mode d'indexation "maison" ou international, la Classification Décimale de Dewey est préconisée. C'est lors de journées de formation organisées par l'INRP en janvier dernier que ce nom de thésaurus a été pour la première fois mis en question. C'est aussi à cette occasion qu'est apparu l'intérêt de faire des correspondances entre le système de classification adopté par les scientifiques et la Classification Décimale de Dewey. C'est dans ce contexte que j'ai été recrutée avec deux objectifs : d'une part de proposer des améliorations aux "thésaurus" de chaque discipline que j'ai redéfinis comme étant des taxonomies après des recherches dans des revues professionnelles. D'autre part j'ai effectué un travail de correspondance entre les taxonomies disciplinaires et la Classification Décimale de Dewey afin de permettre une indexation liée: lors de l'indexation avec la taxonomie, la

correspondance effectuée avec la CDD sera automatique. J'ai aussi effectué des correspondances entre les taxonomies et des thésaurus spécialisés en sciences de l'éducation à savoir Mot Bis et l'ETB à la demande de mon tuteur Gérard Vidal.

Mon rapport de stage se présente en trois grandes parties. Tout d'abord j'exposerai le contexte de la mission et les institutions dans lequel mon travail a eu lieu. Dans un deuxième temps, je développerai le déroulement de mon travail. Pour conclure, dans le cadre de la note de recherche je m'interrogerai sur la question de l'interopérabilité des langages documentaires en essayant de remettre ce problème dans le contexte du travail effectué lors de mon stage.

2 Présentation

Mon travail s'est effectué au sein de l'INRP, cependant, dans le cadre de la mission de mon stage, j'ai eu l'occasion de travailler avec l'équipe Pratic de l'ENS et de nombreux scientifiques de l'ENS, aussi j'ai décidé de présenter les deux instituts. Le service Pr@tic¹ de l'ENS et l'ERTé² ACCES³ de l'INRP sont deux entités liées qui accompagnent la production de ressources pédagogiques à l'ENS sciences et à l'INRP.

2.1 Institutions et services

2.1.1 Présentation de l'INRP (Institut National de Recherche Pédagogique)

L'INRP⁴ est un établissement public national à caractère administratif sous la tutelle du ministère de l'éducation nationale. Son siège se situe à Lyon depuis 2005. Cette délocalisation en province de l'établissement a commencé dès 1980 avec l'installation du musée à Rouen. Ce transfert lui a permis de rationaliser son fonctionnement en renouvelant son personnel, ses missions et ses partenariats. Par ailleurs il possède des implantations à Paris ainsi que des unités mixtes de recherche à Marseille et Cachan. L'INRP a pour vocation de conduire des recherches en éducation, de mettre à disposition les résultats de ses recherches, d'assurer une veille scientifique en sciences de l'éducation, de participer à la formation des enseignants et des formateurs, de mettre à disposition des ressources à l'intention des chercheurs et des enseignants. L'INRP participe à la formation

1(Pôle de Ressources et d'Assistance pour les Technologies de l'Information et de la Communication)

2(Équipe de Recherche Technologique en éducation)

3(Actualisation Continue des Connaissances des Enseignants en Sciences)

4[<http://.inrp.fr/>]

initiale et continue des personnels de l'éducation nationale et a une activité d'expertise auprès du ministère et plus particulièrement auprès de la Direction Générale de l'Enseignement Scolaire. Pour cela, l'INRP s'appuie sur ses services: bibliothèque, musée national de l'éducation, son service des publications, et développe des partenariats avec d'autres établissements (IUFM, ENS).

2.1.2 ACCES (Actualisation Continue des Connaissances des Enseignants en Sciences)

En 2005, Naoum Salamé, qui dirigeait le service TECNE dont la mission était de concevoir et de développer des logiciels pour l'enseignement secondaire, et Gérard Vidal qui dirigeait à l'ENS Lyon (sciences) le service TIC en charge des sites DGESCO d'accompagnement des programmes, se sont réunis afin de créer une structure mixte pour l'actualisation continue des connaissances des enseignants de sciences, ACCES⁵.

Il a été choisi de constituer une ERTé, Équipe de Recherche Technologique en éducation ce qui semblait mieux satisfaire aux contraintes et aux objectifs fixés. Dans cette structure, les deux ENS de Paris et de Lyon apportent leurs compétences scientifiques et l'INRP sa connaissance du système éducatif et son savoir-faire dans la diffusion des connaissances à l'intention des enseignants. ACCES se compose de plusieurs types d'acteurs: des enseignants associés producteurs de ressources, des chercheurs et des ingénieurs INRP qui organisent et pilotent les thématiques de l'ERTé et un certain nombre de partenaires: l'IGEN (Inspection générale de l'éducation nationale) disciplinaire, les ENS, le CNRS (Centre National de la Recherche Scientifique), Météo France...

Ses activités se regroupent autour de quatre grands axes: Santé humaine,

⁵Le service ACCES a un site sur lequel on retrouve une partie de ces informations [<http://acces.inrp.fr/acces>]

Neurosciences, Environnement et développement, Imagerie-Modélisation et bases de données. Les sites Web de ressources produits par l'équipe sont des références dans certains domaines de sciences de la vie, de la terre et de l'univers. On peut citer les quatre grands axes de travail de l'ERTé: la différenciation sexuelle et le développement, l'évolution et la biodiversité (accompagnement Phylogénie et Santo), la modélisation en sciences de la Terre (notamment le cycle du Carbone), et les Systèmes d'information géographiques EduTerre-Usages avec le BRGM (Bureau de Recherche Géologique et Minière).

ACCES a contribué à des événements académiques et nationaux qui ont permis de conforter la position de l'équipe dans son rôle de médiation et de transposition : comme la participation aux plans académiques de formation et l'organisation de journées d'étude. Cette participation a permis la mise en place des formations en ligne ré-exploitablees dans toutes les académies pour tous les enseignants du territoire.

Les formations de formateurs mises en place tirent profit des relations privilégiées qu'entretient ACCES avec ses laboratoires de recherche partenaires. Les formations sont organisées sur les axes principaux de travail de l'équipe : l'indexation, la modélisation climatologique, le cycle du Carbone, ainsi que les neurosciences et l'apprentissage.

ACCES est en train de mettre en place des dispositifs de communautés virtuelles pour accompagner les enseignants tout au long de la vie sur des thèmes évoluant rapidement du fait des avancées de la recherche. On peut prendre l'exemple de Formaterre, site élaboré pour l'accompagnement des enseignants en sciences de la Terre.

2.1.3 ENS Sciences de Lyon

L'École normale supérieure de Lyon a été créée en 1985 dans le cadre de la réorganisation des écoles normales supérieures. Celle-ci est l'héritière des sections

scientifiques des deux écoles normales de Fontenay et de Saint-Cloud dont la mission initiale était la formation des professeurs des écoles normales primaires. Devenues mixtes, les écoles normales se sont aussi réparties les disciplines. Les sections scientifiques ont été les premières à s'installer à Lyon en 1987. Les sections lettres et sciences humaines ont été transférées à Lyon seulement en 2000. L'École normale supérieure de Lyon a pour mission de former ses élèves à *"la recherche scientifique fondamentale ou appliquée, à l'enseignement universitaire et dans les classes préparatoires aux grandes écoles ainsi qu'à l'enseignement secondaire et, plus généralement, au service des administrations de l'Etat et des collectivités territoriales, de leurs établissements publics ou des entreprises. Elle exerce ses missions principalement dans les disciplines suivantes : mathématiques et informatique, sciences de la nature, de la vie et de la santé, sciences de la matière et de l'univers."* décret n°87-697 du 26 août 1987.

2.1.4 Pr@tic

Pratic⁶ est un service transversal de l'ENS science de Lyon chargé des TIC. Il comprend des enseignants, des chercheurs et des ingénieurs autour de l'élaboration de ressources et de dispositifs de formation exploitant les technologies de l'information et de la communication. Le travail des équipes se répartit en plusieurs axes : la production de ressources, la mise à disposition de ressources numériques, la veille et le développement d'outils. Pr@tic met également à disposition ses compétences informatiques (sites web dynamiques, indexation, structuration...) et en vidéo (tournage, encodage, diffusion web ou mobile).

2.2 Ressources pédagogiques

Les productions des deux services sont proposées sous forme de sites web, nous allons essayer de les énumérer et d'en faire une rapide présentation. Dans cette partie je citerai indifféremment des sites de l'INRP et de l'ENS, en effet la

6 Le service a un site sur lequel on trouve toutes ces informations. [<http://pratic.ens-lyon.fr>]

majorité ont été conçus en collaboration entre les deux établissements. Elles peuvent se diviser en 3 parties : une première regroupant la production de ressources qui sont encadrées par des conventions de collaboration entre divers établissements. La deuxième partie regroupe des espaces de formation et une troisième porte sur la participation du service aux Universités Numériques Thématiques (UNT).

2.2.1 Ressources pédagogiques

Village d'ESTE⁷ est un site de ressources pour l'Enseignement des Sciences et de la Technologie à l'École. Il a pour objectif de permettre aux enseignants du primaire d'entretenir leur niveau de connaissances en sciences ainsi que de leur proposer des travaux à effectuer en classe. Ce site a été élaboré en association entre l'école normale supérieure de Lyon (ENSL), l'Institut National de Recherche Pédagogique (INRP), l'Institut universitaire de formation des maîtres (IUFM), l'Université Claude Bernard Lyon 1 (UCBL), ainsi que le Rectorat et le Centre régional de documentation pédagogique (CRDP) de l'académie de Lyon.

Le site Biotic, site de ressources et de travail coopératif de l'INRP pour les sciences de la vie et de la Terre. Ce site propose des dossiers sur des thématiques des programmes d'enseignement du secondaire ainsi que des logiciels (exemples: RasTop: logiciel de visualisation des molécules en 3D, Phylogène: logiciel qui permet de faire des liens entre différents organismes vivants) et des bases de données dans ce champ des sciences.

Mission Santo⁸ L'objectif de ce site est de proposer des ressources scientifiques exploitables en classe à partir de recherches effectuées sur l'île d'Espiritu Santo aux îles Vanuatu. Ce site est piloté par l'INRP et l'ENS. Riche en illustration, Santo 2006 propose des ressources intéressantes à exploiter avec des élèves.

7[http://isheyvo.ens-lyon.fr/village_este||village d'Este

8 [<http://santo2006.ens-lyon.fr/>]

Relie (REseau Lyonnais d'Ingénierie Éducative)⁹ Ce site a été créé suite à une convention entre le Rectorat, l'INSA et l'ENS qui encourage la création de modules multimédias d'enseignement scientifique pour les élèves de lycées et de collèges. Il donne des liens vers des portails et d'autres sites de l'équipe : Les Fleurs Diversités et Unités¹⁰ , Les cadrans solaires¹¹, Amplication d'ADN par polymérisation en chaîne¹², Voyage au centre de la terre¹³. Ces sites, relativement anciens sont aujourd'hui un peu abandonnés.

La Direction générale de l'enseignement scolaire (DGESCO) est l'organisme chargé d'élaborer la politique éducative et pédagogique ainsi que les programmes d'enseignement. Celle-ci a mis en place une collaboration avec les ENS, pour la production de sites experts afin d'enrichir la formation disciplinaire des enseignants. L'ENS Lyon héberge deux de ces sites un en Sciences de la Terre, Planet-Terre¹⁴ et un en physique Culture-Sciences Physique¹⁵. Ces sites proposent aux enseignants d'actualiser et d'approfondir leurs connaissances par l'accompagnement aux nouveaux programmes. Les ressources sont validées et actualisées par des scientifiques afin de correspondre aux besoins des enseignants. Cette dynamique crée des liens entre l'enseignement secondaire et la recherche en introduisant des thématiques de recherche émergentes qui ne sont pas encore introduites dans les programmes. Les ressources scientifiques sont conçues pour être utilisées comme un outil d'information, mais elles ne sont pas destinées à être utilisées directement dans une classe.

Dicos (Diffusion des Connaissances Scientifiques)¹⁶ L'objectif de ce site est de regrouper les ressources de plusieurs établissements d'enseignement supérieur de

9[<http://www.ens-lyon.fr/RELIE/>||RELIE]

10[<http://www.ens-lyon.fr/RELIE/PCR/>||PCR]

11[<http://planet-terre.ens-lyon.fr/>||planet terre]

12[<http://www.ens-lyon.fr/RELIE/PCR/>||PCR]

13[<http://planet-terre.ens-lyon.fr/>||planet terre]

14 [<http://planet-terre.ens-lyon.fr/>||planet terre]

15 [<http://culturesciencesphysique.ens-lyon.fr/>]

16 [<http://dicos.ens-lyon.fr/>||Dicos]

la région. Les ressources sont réparties par disciplines, d'autre part Dicos met à disposition un dossier sur le LOM. Le LOM (Learning Object Metadata) est le schéma de métadonnée spécifiques aux ressources pédagogiques avec lequel travaillent les personnes des services Pr@tic et ACCES. Nous y reviendrons largement dans la deuxième grande partie du rapport.

2.2.2 Espaces collaboratifs

Le service Pr@tic travaille également à la création de sites web à destination des enseignants et des chercheurs de l'ENS.

Le service a travaillé à l'environnement numérique de travail, l'EPI (l'Espace Pédagogique Intégré) qui regroupe les ressources scientifiques et pédagogiques utilisées et produites au sein de l'École Normale Supérieure. Ces ressources à l'intention des enseignants, étudiants et personnels sont accessibles en extranet. Par ailleurs, l'EPI propose un espace public sur internet de diffusion vers l'extérieur d'une sélection de cours, de ressources pour l'enseignement et des actions de formation permanentes, d'actualités de la recherche et de vulgarisation scientifique.

CINEMATIC (Centre d'INdexation et d'Edition de MétA-données pour les TIC)¹⁷ est un site de ressources pédagogiques proposant une fonction de recherche et une fonction d'indexation. Il doit permettre de faire des recherches dans toutes les métadonnées des ressources proposées par le service Pr@tic.

Cinemacdm¹⁸ propose une présentation et un mode de recherche des offres de formation des universités. Le site est en construction, il manque les présentations des offres de formation des universités.

C.A.Vie-Terre La Communauté virtuelle des formateurs à l'agrégation des sciences de la Vie-Sciences de la Terre et de l'univers regroupe les personnels impliqués dans les préparations à l'agrégation en sciences de la vie et sciences de la

17[<http://cinematic.ens-lyon.fr>]|Cinematic]

18[<http://dicos.ens-lyon.fr>]|Dicos]

Terre et de l'univers. L'accès est réservé aux inscrits.

Un site de gestion de colloque, créé à l'occasion du congrès de juillet 2006 de la Federation of European Societies of Plant Biology organisé tous les deux ans dans une ville différente. Ce site, créé à l'occasion d'un événement, a eu une durée de vie limitée et il a été supprimé aujourd'hui.

EAAE European Association for Astronomy Education¹⁹ Site de l'association européenne qui propose des ressources en astronomie pour les enseignants. Le site est hébergé par un serveur du service mais celui-ci agit peu sur le contenu du site.

FormaTerre²⁰ site du réseau des formateurs en sciences de la terre. Son objectif est de faire le lien par un espace collaboratif de formation entre chercheurs, enseignants et formateurs en sciences de la terre.

Le service a également élaboré le site de laboratoire de recherche de l'ENS, comme celui du laboratoire de virologie humaine.

En projet : le site BioTICE, qui devrait permettre la diffusion des publications des élèves et enseignants de l'ENS en sciences de la vie.

2.2.3 UNT

Le service participe également à la création des Universités Numériques Thématiques, ces universités numériques ont pour fonction de proposer des formations et une mutualisation des ressources pédagogiques par thème.

UNISCIEL UNiversité des SCiences En Ligne²¹ : Unisciel a pour but de permettre la mutualisation de la production de ressources pédagogiques numériques et leur diffusion en libre accès à destination des étudiants et des enseignants. Unisciel doit permettre l'accès à des cours, exercices et documents vidéo ou

19[<https://isheyeva.ens-lyon.fr/eeae>][EAAE]

20[<http://www.formaterre.fr>][Formaterre]

21[<http://www.unisciel.fr/>][Unisciel]

multimédia, ainsi qu'un catalogue des formations offertes par les membres d'Unisciel. Unisciel est actuellement en phase de création.

UVED²² L'Université virtuelle environnement et développement durable est une Université numérique thématique (UNT). Cette UNT se fonde sur l'interdisciplinarité et un lien fort entre recherche et enseignement.

2.3 Les scientifiques contributeurs du projet

Pour chaque discipline, j'ai eu un ou deux interlocuteurs privilégiés, soit une personne utilisatrice quotidienne de ces outils soit un des concepteurs des langages d'indexation disciplinaires (étant la plupart du temps également un utilisateur fréquent) . Leur premier rôle a été une pré-validation des correspondances avant une validation collégiale. Sur les quatre disciplines que j'ai traitées à savoir les sciences de la vie, les sciences de la Terre, la physique et la chimie, trois de mes interlocuteurs sont des enseignants responsables des sites de la DGESCO; en physique, Catherine Simand-Vernin, en science de la Terre, Olivier Dequinsey, et en chimie Hagop Demerdjan ainsi qu'un chercheur de l'Ens Romain Barbe. En sciences de la Vie, les personnes référentes étaient les conceptrices de la taxonomie à savoir Françoise Jauzein, enseignante détachée à l'INRP et Sandrine Heusser, responsable de l'agrégation des sciences de la vie à l'ENS. Chaque validation a été organisée avec un groupe de scientifiques avec en plus de ceux que j'ai déjà cités: Pierre Thomas et Gérard Vidal pour les sciences de la Terre, Jean-Bernard Baudin pour la chimie et Marie-Christine Artru pour la physique.

²²[<http://www.uved.fr>][UVED]

2.4 Outils de travail

Les personnels des services Pr@tic et ACCES travaillent principalement avec des sites collaboratifs Zope/Plone. Zope est un serveur d'application Web, utilisé pour la conception d'application complexe. Différents outils ont été conçus, en s'appuyant sur Zope afin de faciliter le déploiement de portails collaboratifs de gestion de contenu. C'est le cas de Plone, un logiciel libre. Le travail collaboratif s'est effectué principalement par l'intermédiaire du site de [Pr@tic](#)²³.

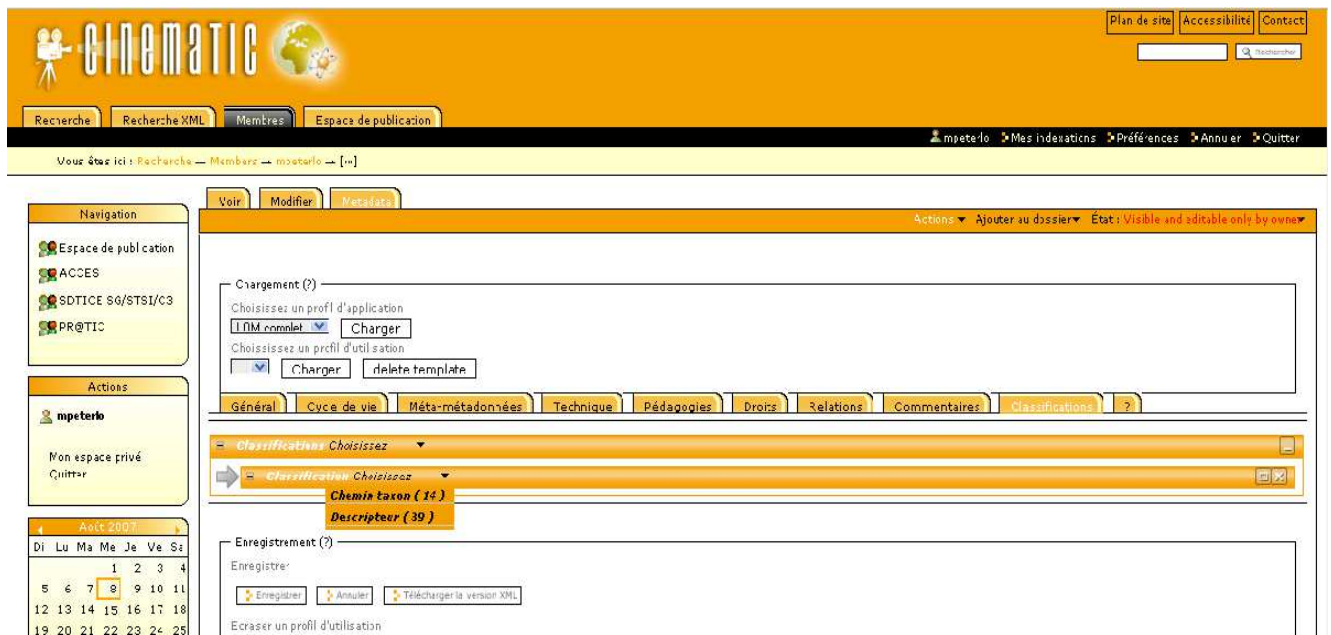
Sur ce site collaboratif chaque contributeur du projet peut déposer des documents; il existe deux niveaux de publication: un limité aux membres du groupe de travail et un second pour tous .

Ainsi une partie de mon travail a été publié sur le site de Pr@tic, à cette adresse: [\[http://pratic.ens-lyon.fr/projets/meta-LOM-CDM/thesaurus/thesaurus/|http://pratic.ens-lyon.fr/projets/meta-LOM-CDM/thesaurus/thesaurus/\]](http://pratic.ens-lyon.fr/projets/meta-LOM-CDM/thesaurus/thesaurus/).

Cinematic est un outil développé par la société Pentila sur une commande du service Pr@tic afin de permettre le renseignement de métadonnées suivant le LOM.²⁴

23[\[http://pratic.ens-lyon.fr/|Pratic\]](http://pratic.ens-lyon.fr/)

24<http://cinematic.ens-lyon.fr/>



Capture d'écran de Cinematic

En ce qui concerne le travail sur les langages "maison", nous nous servons d'un logiciel nommé Freemind, qui est utilisé pour faire des cartes heuristiques. Ce mode de présentation permet une structuration efficace. Il permet également de générer automatiquement une page XML qui reprend chacun des termes.

3 Missions

Le projet sur lequel j'ai travaillé s'inscrit dans une optique plus globale d'indexation des ressources pédagogiques auquel participent l'INRP et l'ENS. Cette volonté se traduit par l'adoption du LOMFR. Dans cette partie je vais présenter les premiers objectifs suivis lors du commencement du stage. Puis nous détaillerons le contexte de la mission à savoir l'utilisation du LOM et plus particulièrement son profil d'application français, le LOMFR. Puis dans un dernier temps nous détaillerons ce qui a été le coeur de la mission à savoir l'amélioration des taxonomies disciplinaires élaborées par les enseignants ainsi que la méthodologie utilisée pour les correspondances entre celles-ci et la Classification décimale de Dewey.

Gérard Vidal m'a donné deux objectifs principaux à mon arrivée :

- Proposer des voies d'amélioration des terminologies élaborées par les scientifiques appelées thésaurus au sein des groupes de travail.
- Effectuer les correspondances entre ces vocabulaires élaborés au sein de l'ENS et de l'INRP et la classification décimale de Dewey. Ce travail doit permettre le complément d'une des balises du LOM. Cette interopérabilité des taxonomies disciplinaires avec d'autres langages normalisés est aussi un des objectifs du projet.

3.1 Contexte de la mission: l'utilisation du LOMFR

3.1.1 La production de documents

Le projet se place dans le contexte d'une production importante de documents et d'objets pédagogiques par différents enseignants et chercheurs de l'ENS et l'INRP. Le service ACCES comme le service Pr@tic ont pour fonction de produire des

ressources pédagogiques. On peut faire une typologie des objets pédagogiques produits par les services: articles, images, photographies, vidéos de conférence, interviews de scientifiques.

3.1.2 Les Métadonnées

Les métadonnées sont des données permettant de décrire d'autres données. Celles-ci permettent un partage, un stockage et des recherches d'information plus faciles, par ailleurs, elles permettent des mises en relation avec d'autres ressources. En 1960, le terme "*Metadata*" est utilisé pour la première fois par Jaquie Mayer pour décrire d'autres langages. En 1986, il apparaît dans la littérature professionnelle. Ensuite il est utilisé par la NASA et s'est d'ailleurs développé par ce biais. En 1995, lors du premier congrès Dublin Core, metadata est passé dans le vocabulaire informatique. Tim Berners Lee, un des inventeurs du web, a affirmé que les ressources d'information numériques doivent être compréhensibles par les êtres humains et traitables par des logiciels.

Les métadonnées ne servent pas uniquement à retrouver plus facilement les ressources mais aussi à les évaluer. En outre elles permettent une meilleure gestion des différentes versions d'un document. Parmi les modèles de métadonnées, le plus connu et le plus générique est le Dublin Core, qui comprend quinze éléments qui reprennent la description de base d'une ressource numérique ou physique. Le Dublin Core comprend dans ses éléments de description trois aspects : formels (auteur, titre, éditeur), intellectuels (sujet, description, langage) et relatifs à la propriété intellectuelle. Il fait l'objet d'une norme internationale (ISO 15836) depuis 2003. Mais il existe d'autres formats de métadonnées que le Dublin Core. En effet, il existe un certain nombre de schémas de métadonnées spécialisés pour un domaine d'activité: IPTC Core, spécifique à la presse, MODS schéma élaboré par les bibliothèques à mi-chemin entre MARC, MARC est un format de données des catalogues de bibliothèques et Dublin Core, ou encore le LOM spécifique aux ressources pédagogiques que nous allons voir plus précisément.

3.1.3 Le LOM (Learning Object Metadata)

Le LOM (Learning Object Metadata) est un standard de métadonnées spécifiques aux ressources pédagogiques. Il est principalement utilisé pour décrire des ressources pédagogiques numériques mais il peut décrire tout objet pédagogique soit "toute entité numérique ou non, utilisée dans un processus d'enseignement, de formation ou d'apprentissage"[Gomez de Regil, 2004]²⁵

Le LOM est un standard publié en 2002 par le Learning Technology Standards Committee (LTSC) de l'IEEE (Institute of Electrical and Electronics Engineers). Un standard est un ensemble de recommandations établies par un groupe d'utilisateurs.

En pratique, peu d'établissements utilisent l'intégralité du LOM. Il est davantage utilisé selon des profils d'application. Le LOM a un profil d'application français, le LOMFR. Celui-ci a été produit sous l'impulsion du ministère de l'éducation nationale, pour les ressources numériques éducatives. Ce schéma de métadonnée est inspiré du profil ManUel ²⁶, et a valeur de norme expérimentale Afnor (Association Française de normalisation) en août 2005 et est actuellement une norme depuis décembre 2006 (Afnor Z76-040). Le LOM est un schéma de métadonnée très détaillé, il comprend plus de 50 éléments hiérarchisés.

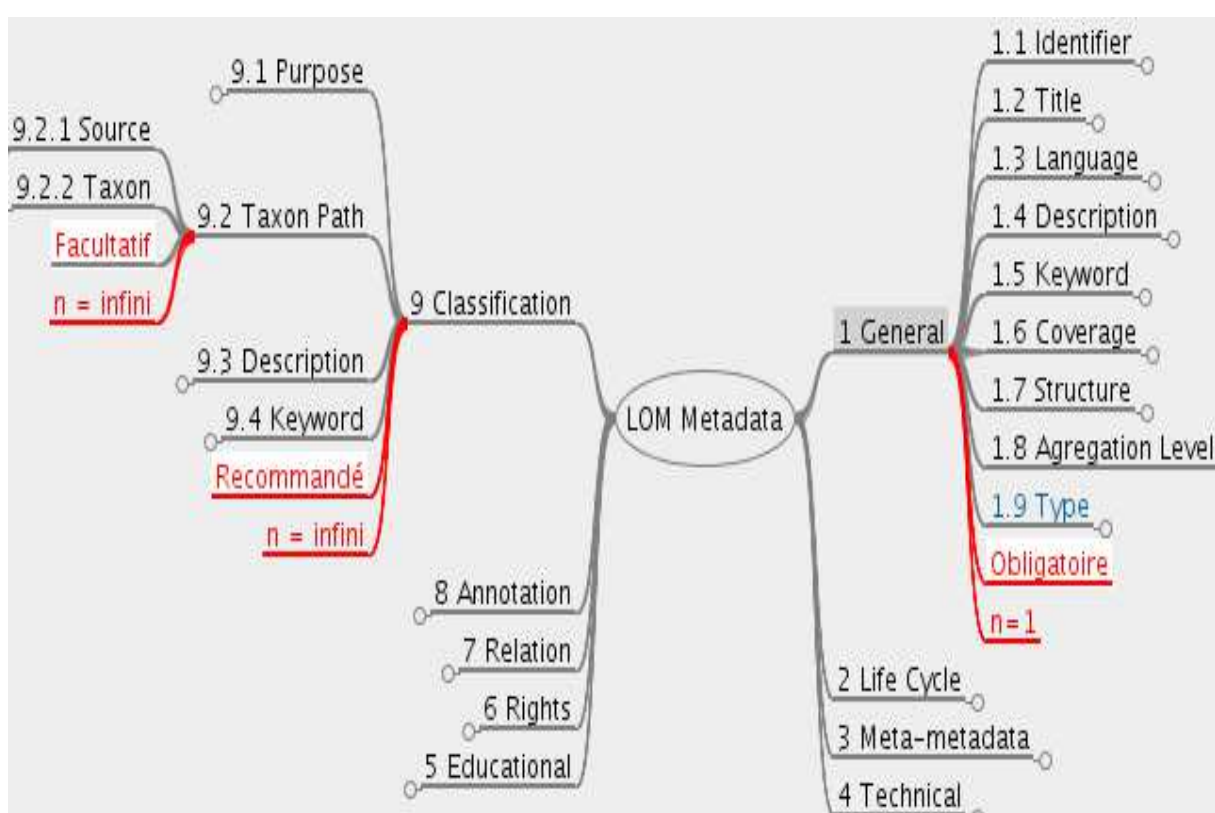
²⁵ citation d'une citation de l'article de [Gomez de Regil, 2004] note n°1 Learning Technology Standards Committee of the IEEE. Draft Standard for Learning Object Metadata. [en ligne]. New York : IEEE, 2002. Disponible sur : http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf (Consulté le 10 juin 2005).

²⁶ Le profil ManUel est un profil d'application du LOM dans le cadre du projet C@mpusSciences, projet de développement d'une université en ligne élaboré en 2004. plus de renseignement: http://sticef.univ-lemans.fr/num/vol2004/passardiere-11/sticef_2004_passardiere_11.pdf

3.1.4 Les catégories du LOM

Le LOM est structuré en neuf catégories, on distingue à l'intérieur de ces catégories plusieurs types de métadonnées : descriptives, administratives, de gestion de droit, de préservation ou d'interopérabilité. On peut voir une carte complète du LOMFR réalisée par Gérard Vidal. ²⁷

Carte du LOMFR



La quasi-totalité des éléments du LOMFR sont facultatifs, le modèle peut fonctionner même si tous les champs ne sont pas remplis, excepté pour le titre et l'identifiant de la ressource. Par ailleurs, il est possible d'ajouter d'autres éléments propres à un groupe d'utilisateurs. Selon Rosa Maria Gomez de Regil « *il est impensable que l'ensemble des métadonnées du LOM puisse être renseigné par une seule personne.* »

²⁷On peut la télécharger, parmi d'autres documents sur le LOM à cette adresse : http://dicos.ens-lyon.fr/metadata/news_events/lomall

Le LOM comprend plus de cinquante champs, dont une partie peut être renseignée automatiquement comme le format du fichier qui peut être reconnu par le logiciel. Les différentes catégories doivent être remplies par différents acteurs tour à tour, à savoir les enseignants, les documentalistes et les informaticiens²⁸.

- **Général** La fonction de cette catégorie est la description et l'identification de la ressource. Les informations contenues dans ces balises permettent d'établir des liens ou des références vers d'autres balises ou d'autres documents du système d'information.
- **Cycle de vie** Le cycle de vie permet d'avoir des informations sur l'histoire de la ressource et sur ses contributeurs. Les informations collectées ici serviront à la gestion de la ressource ainsi les ressources peuvent avoir "une date de péremption" définie par l'auteur ce qui permet de signaler le besoin de mise à jour par une vérification du contenu du document.
- **Les méta-métadonnées** permettent de garder une trace détaillée du schéma de métadonnées utilisé pour décrire une ressource. Ces informations seront utiles pour envisager l'interopérabilité des différents systèmes et pour l'échange de données. Cette balise doit indiquer qui a renseigné les différentes balises.
- **Technique** Cette catégorie présente les caractéristiques techniques qui permettront l'exécution de la ressource sur un système informatique. Les contraintes techniques d'utilisation peuvent être exprimées ici. Cette catégorie sert à décrire, évaluer et gérer les ressources.
- **Pédagogique** La pédagogie est le coeur du LOM. C'est ce qui le rend spécifique par rapport à d'autres modèles. Cette catégorie accomplit une fonction de description pédagogique de la ressource. S'y retrouvent, pour ne citer que certains des éléments, le contexte d'utilisation (scolaire, secondaire, supérieur, formation continue, etc.), le type de public cible (enseignant, auteur,

²⁸ sur ce sujet on peut consulter : [Gomez de Regil, 2004], [Kreczanik, 2004]

apprenant, gestionnaire), la difficulté, l'âge des apprenants auxquels la ressource est destinée, le temps d'apprentissage type, la langue des apprenants, le niveau et le type d'interactivité.

- **Droits** Ici figurent les conditions légales d'utilisation de la ressource : indications de coût, de licences, de droits d'auteur.
- **Relation** Cette catégorie permet de lier deux ressources physiques entre elles, en donnant une valeur à cette relation, un sens (est partie de, est version de, est format de, est référencée par, etc.)
- **Commentaire** L'annotation ou commentaire permet de faire des remarques à propos de la ressource, tout en gardant en mémoire la description de la personne qui commente et la date de ce commentaire. Cette catégorie apporte des informations complémentaires utiles à l'évaluation de la ressource, l'utilisateur peut donner son avis sur le document.
- **Classification** cette catégorie est valorisé dans le LOMFR. Il s'agit de la catégorie qui permet d'assigner des classifications et des vocabulaires contrôlés à une ressource. Des classifications peuvent être attribuées pour la discipline, les pré-requis pédagogiques, les objectifs pédagogiques, l'accessibilité, les restrictions, les compétences.

Le LOMFR indique quels éléments sont obligatoires, recommandés, facultatifs ou implicites.[Gomez de Regil, 2004]

3.1.5 Les problèmes d'utilisation du LOM

Le LOM est critiqué pour son manque de précision sur certains aspects et de son nombre important de balises à remplir (60 champs). Le LOM reste évasif sur la définition de l'objet pédagogique, ce qui serait utile pour la description des usages

qui en sont fait. Le rôle de descripteur (du LOM) n'est pas assez poussé pour prendre en compte des ressources plus interactives comme celles mises en oeuvre dans l'évaluation des connaissances. [Arnaud, 2004] Par ailleurs, certains champs paraissent à certains auteurs comme non pertinents. C'est le cas de l'activité d'apprentissage, le public cible ou encore la difficulté. Certains champs paraissent trop subjectifs ou trop difficiles à informer comme la notion de densité sémantique. La densité sémantique peut être évaluée en fonction de la taille ou de la durée de la ressource. Autrement dit il s'agit d'évaluer le nombre de concepts présents dans le document ou l'objet décrit, ce qui est assez subjectif.

Nous avons déjà évoqué le renseignement des métadonnées par différents acteurs hétérogènes. En effet certains acteurs sont peu enclins à l'indexation, notamment les enseignants, cette tâche leur paraissant trop éloignée de leur métier. On touche ici aux questions fondamentales de l'utilisation du LOM, soit comment impliquer les enseignants-producteurs de ressources.[Kreczanik, 2006]

3.1.6 Balise 9 du LOM mode d'indexation

La balise 9 du LOM doit selon la norme [Afnor, 2006] décrire "*où cette ressource se trouve dans un système de classification particulier*". Cette balise contient quatre éléments.

- **Objectif:** précise l'objectif de la classification de la ressource, cet élément est pré-rempli par une liste fermée de source ou valeur du vocabulaire soit : la discipline (proposition dans la norme d'utilisation des nomenclatures du SCEREN)²⁹. les pré-requis pédagogiques, les objectifs pédagogiques, l'accessibilité, les restrictions, le niveau éducatif, le niveau de compétence.
- **Chemin taxon:** situe la ressource dans un système de classification. Cette balise ou sous-catégorie décrit un chemin taxinomique dans un système de

29(Services culture éditions ressources pour l'éducation nationale, ce réseau regroupe les CRDP et les CNDP.),

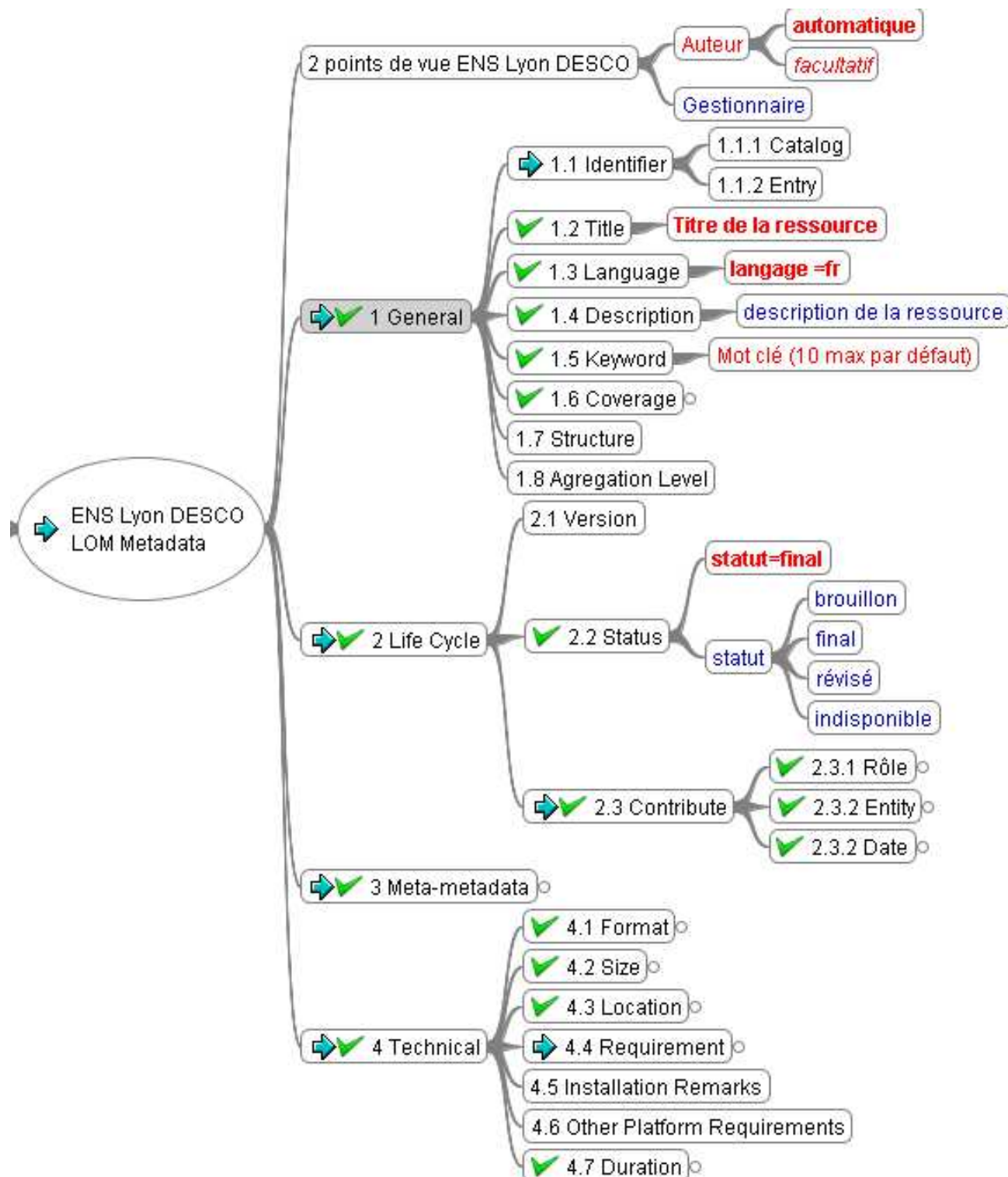
classification spécifique. Chaque niveau successif est une précision apportée au niveau supérieur. Cette sous-catégorie prend son importance par la description de la ressource cette action est à la base de la recherche documentaire. A l'intérieur de cette sous-catégorie on trouve encore trois éléments: **La Source**: le nom du système de classification utilisé. **le Taxon**: le terme de la classification. **ID**: l'identifiant du taxon le nombre ou la combinaison de lettres fournie par la source. Il est possible de répéter la sous-catégorie "chemin taxon" afin de pouvoir associer à la ressource plusieurs systèmes de classification. Ce chemin taxon doit indiquer des termes normalisés. Ceux-ci peuvent être renseignés à la fois par des langages documentaires normalisés comme la Classification décimale de Dewey ou certains thésaurus utilisés au niveau international ou national comme les thésaurus ETB et Mot Bis. Mais aussi par des classifications spécifiques comme les taxonomies disciplinaires "maison". C'est à ce niveau que se situe mon travail de correspondance que je détaillerai un peu plus loin. Cette disposition est critiquée comme "fourre tout" puisque chaque institution est libre de classer l'objet suivant son propre système. Cet inconvénient peut être vu comme un avantage, puisqu'il permet à chaque institution de conserver son propre mode d'indexation et éventuellement d'en rajouter un autre pour autoriser des échanges.

- **Description**: description textuelle qui se rapporte à la classification
- **Descripteur**: mots-clés relatifs à l'objectif de la classification choisie

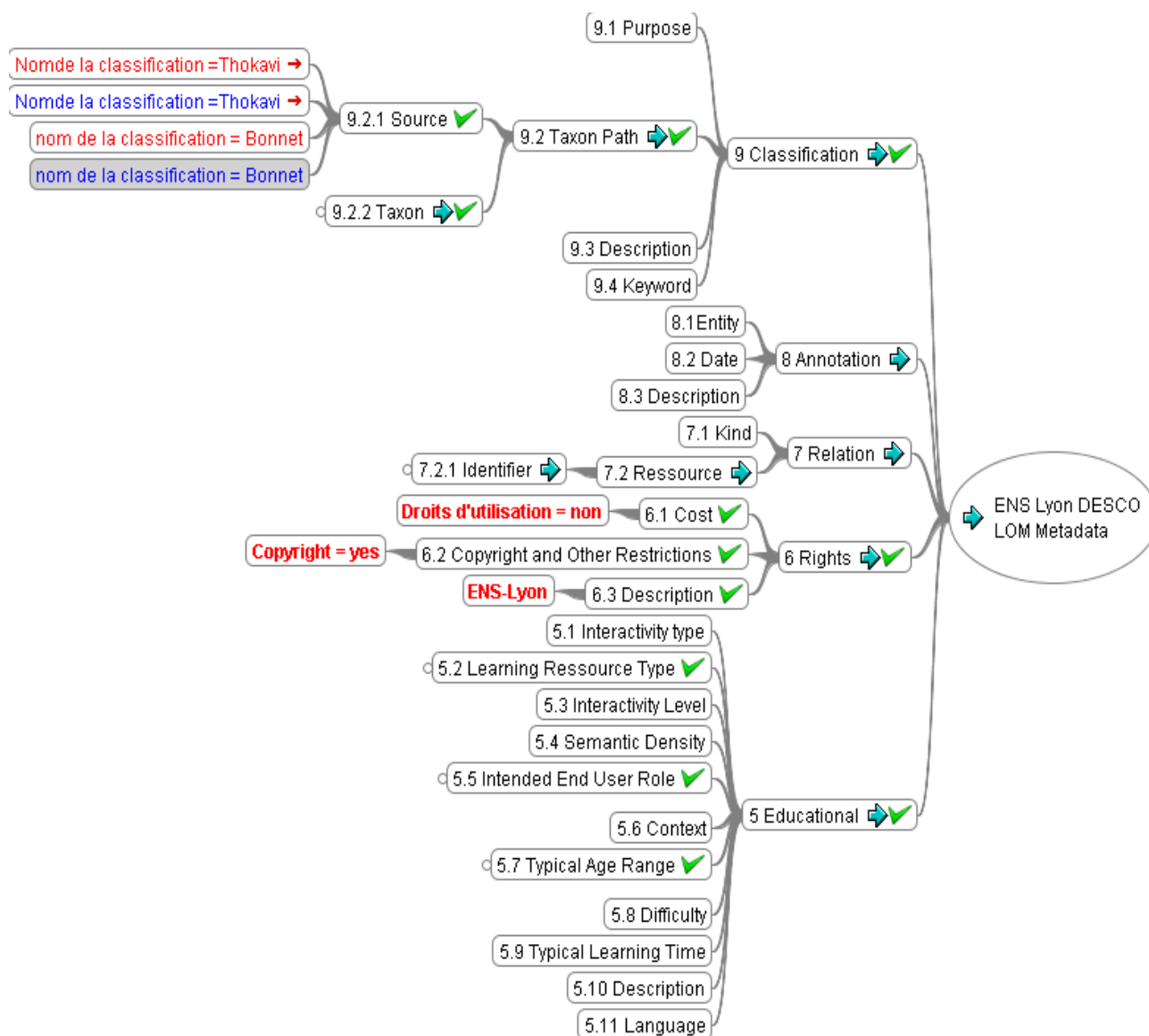
3.1.7 Profils d'application et utilisation du LOMFR à l'INRP et l'ENS

La généralisation de l'utilisation du LOM n'est pas envisageable si chaque auteur ou documentaliste doit renseigner de manière exhaustive tous les champs. Il est nécessaire de proposer une sélection d'information à fournir impérativement. Par ailleurs, il est utile dans certains contextes de pouvoir rajouter des informations. Ainsi pour les ressources mises en ligne par les enseignants responsables des sites ENS-

DGESCO, il a été mis en place un profil d'application spécifique à leurs besoins.



Profil d'application du LOM utilisé pour les sites DGESCO à l'ENS science de Lyon.



Profil d'application du LOM utilisé pour les sites DGESCO à l'ENS science de Lyon.

Les flèches bleues indiquent la présence de champs d'information dépliables. Les ticks verts indiquent que la balise est utilisée dans le profil d'application. Les étiquettes en rouge présentent les informations qui sont remplies par l'auteur et en bleu celles qui le sont par le gestionnaire. Cette carte est accessible en ligne dans sa version cliquable: [<http://dicos.ens-lyon.fr/metadata/lomall/lomfr/htmldescoENS>]

dicos]

3.2 Analyse des langages maison

3.2.1 Origine des taxonomies

La volonté de construire un langage documentaire afin d'indexer les ressources produites par les services est apparue il y a un peu plus de quatre ans. Les premières interrogations se sont posées en 2002 dans la gestion du site Planet-Terre. Après une recherche dans les thésaurus existants, les scientifiques n'ayant trouvé aucun langage documentaire correspondant à leurs besoins, ils ont décidé d'élaborer leur propre thésaurus. Le premier ThoKaVi (terminologie en science de la Terre) a été élaboré en 2003. L'originalité et la force de l'outil élaboré résident précisément dans cette création par les utilisateurs eux-mêmes sans la médiation de professionnel de l'organisation de l'information. L'originalité de cette démarche de création permet une adhésion rapide des utilisateurs. Ce travail présente un véritable défi au vu de la diversité des documents (images, textes, vidéos, documents physiques, exercices/expériences, fiches d'activités, travaux bibliographiques) et des auteurs (chercheurs, enseignants, étudiants).

3.2.2 Thésaurus/Taxonomies: définitions des terminologies "maison"

Les langages documentaires "*maison*" ont été appelés thésaurus lors de leur élaboration. Mon premier travail a été la définition exacte de ces outils. Ma première question a été : Est ce que ce sont des thésaurus ? Après des recherches sur les langages documentaires, j'ai pu donner des définitions précises à chacun des outils. Ainsi on peut affirmer qu'il manque quelques éléments pour pouvoir parler de thésaurus. Un thésaurus est un index de termes organisés de manière hiérarchique comme les terminologies existantes. Un thésaurus doit avoir des liens sémantiques entre les mots, des liens d'équivalence (qui indiquent une synonymie), des liens d'association ou encore des notes d'application. Cette deuxième partie de la

définition manque aux outils existants et présente les caractéristiques fortes du thésaurus.

Dans leur état actuel, on peut parler pour les terminologies "maison", de taxonomies, de vocabulaires contrôlés, ou encore de terminologies hiérarchisées, ces deux dernières appellations étant plus générales mais j'utiliserai au long du mémoire ces trois expressions. Les taxonomies désignent la hiérarchie ou l'arborescence qui peut être utilisée telle quelle pour classer ou indexer mais aussi servir à construire des langages documentaires plus complexes comme les thésaurus ou les ontologies. Cette définition provient de [Menon, 2007].

Je la cite intégralement : *"Une taxonomie est un cadre d'organisation (et pas seulement documentaire - en cela les taxonomies ne sont pas toujours des langages documentaires) pour les ressources de toutes natures destiné à en permettre une représentation ordonnée et y donnant accès par une navigation hypertextuelle."*

Dans cet article il donne la définition que j'ai adoptée mais fait aussi un point sur les différentes définitions présentes dans la littérature. Je reprends ici mes notes prises à ce sujet qui résument celles-ci :

« Taxonomies et taxinomie sont des équivalents. Les taxonomies sont davantage le fait de praticiens ou d'industriels. » Thesauroglossaire: taxonomie est un terme spécifique de langage hiérarchisé mais un synonyme de systématique et en donne comme définition *« classification des formes vivantes »*

J . Maniez *« La taxinomie est au sens général, la science des classifications, mais on nomme aussi taxinomie les classifications de structure monohiérarchique. »*

Dans son lexique en ligne la société Lingway donne cette définition: *« Réseau sémantique dans lequel la seule relation est la relation hiérarchique »*

Le Montague institute définit une taxinomie comme *« un système pour nommer et organiser des objets en groupe qui partagent des caractéristiques similaires »* En précisant que cette définition n'implique pas que le seul mode d'organisation d'une taxonomie soit une hiérarchie thématique.

Dans le glossaire du Dublin Core: *« classification systématique selon des principes ou des lois généraux »* et encore que *« un système de classification comme la classification de la bibliothèque du congrès est un exemple de taxonomie »* ou encore *« un thésaurus est une taxonomie »*

Ces définitions très hétérogènes passent de l'acception restreinte et originelle de classification biosystématique à une définition élargie aux systèmes classificatoires exclusivement hiérarchiques puis à tout langage documentaire doté d'une organisation hiérarchique.

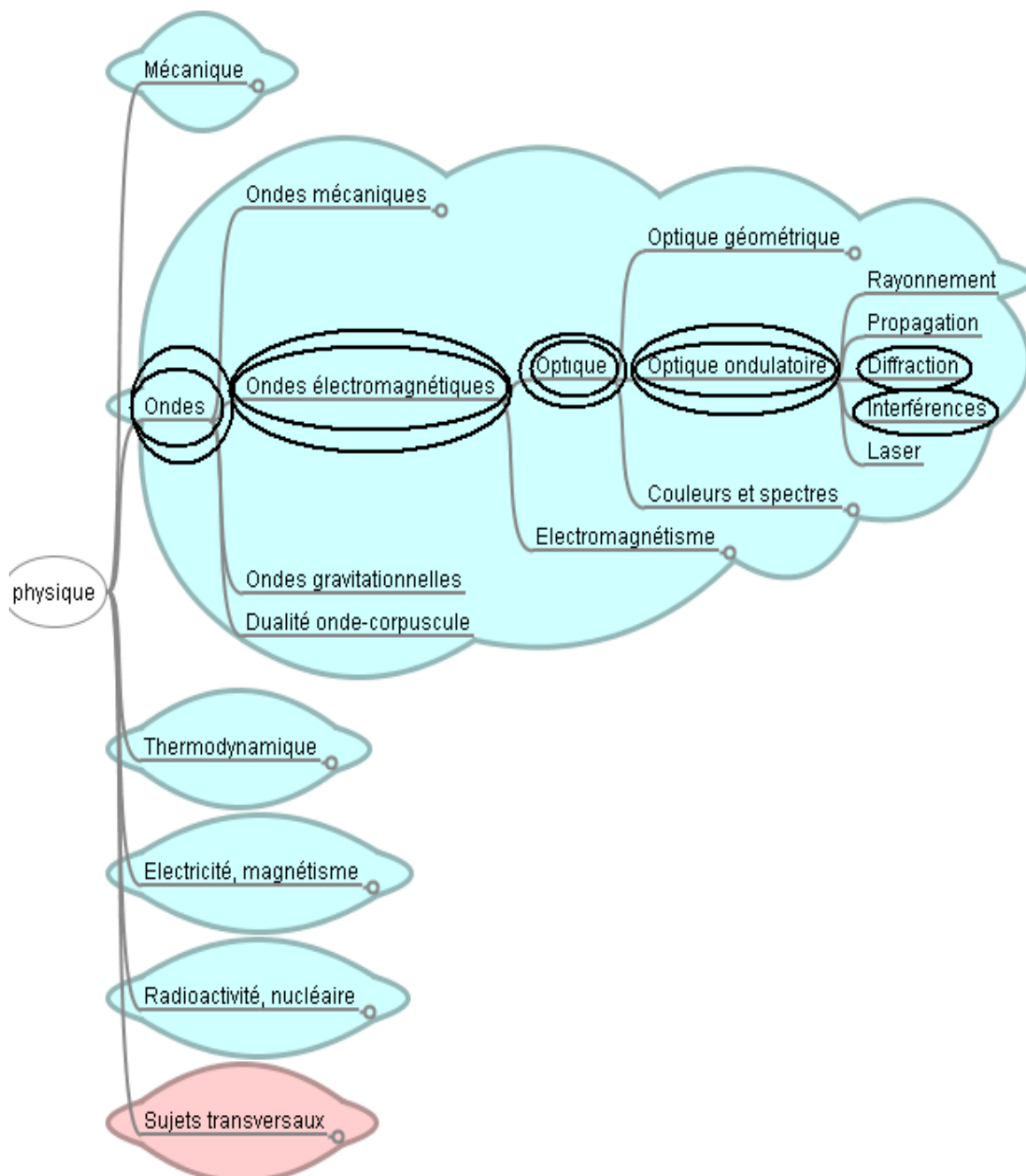
J'ai été tentée de les appeler classification ou plan de classement, dont la définition ne correspond pas vraiment, en particulier en ce qui concerne leurs règles d'usages. Les classifications sont des outils fondés sur la structuration hiérarchique de connaissances. La limite d'un classement de type réside dans l'absence de développement de plusieurs aspects ou point de vue.

Les terminologies "maison" dans leurs règles d'usage permettent l'usage de plusieurs "chemins", et de plusieurs termes pour un même objet à décrire. J'insiste beaucoup sur ce point, les taxonomies malgré leurs ressemblances avec un plan de classement, n'en sont pas. Elles sont utilisées comme des langages combinatoires mais pas comme des langages classificatoires.

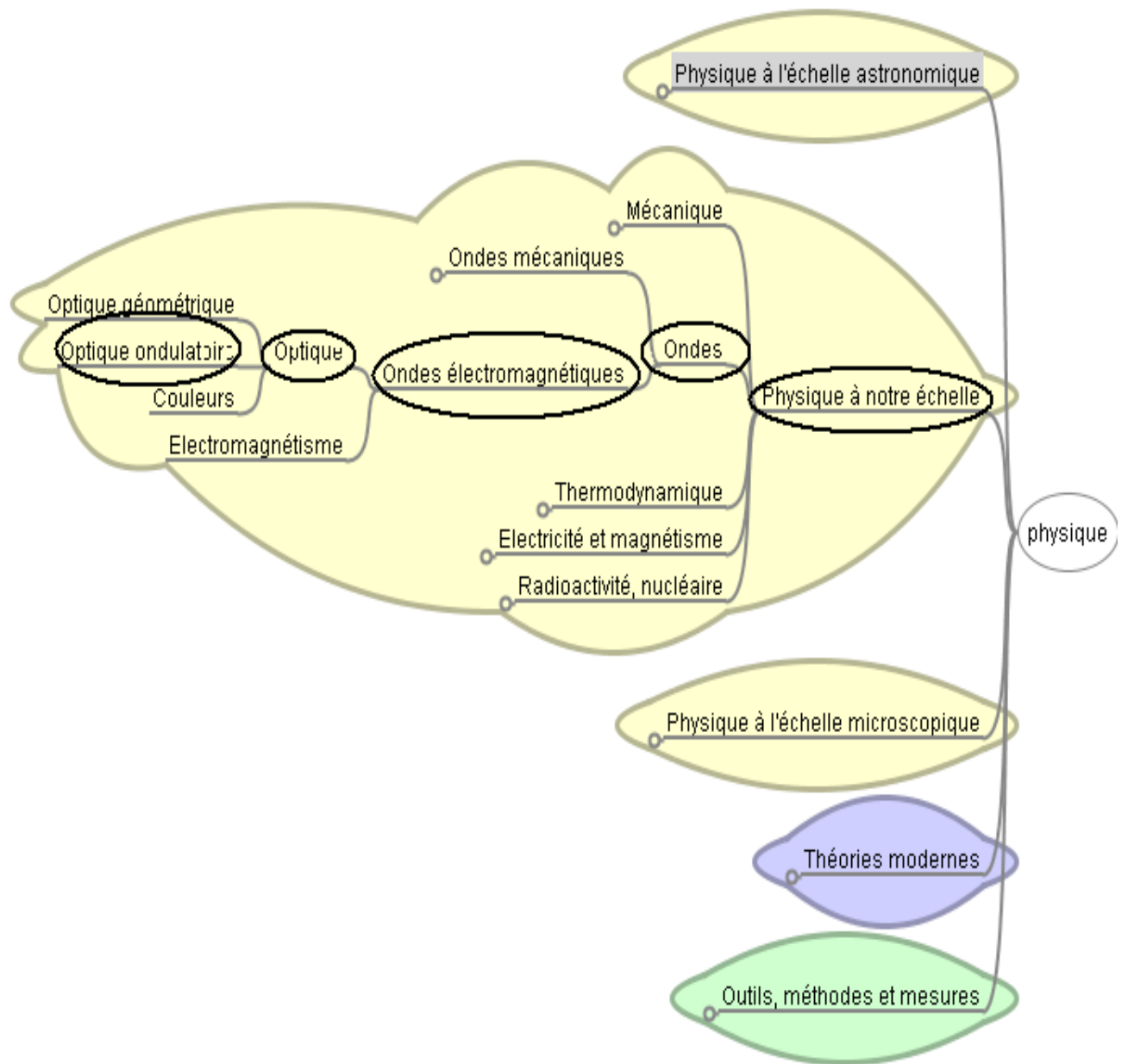
Pour appuyer cette affirmation, il est nécessaire de rappeler les caractéristiques de chaque langage. Les langages classificatoires représentent le contenu des documents de manière synthétique alors que les langages combinatoires donnent une représentation analytique du contenu des documents et des questions. Cette distinction est fondée sur plusieurs critères ; le plus important dans le cas qui nous intéresse est la coordination. Il en existe deux approches. Un langage pré-coordonné, comprend des expressions déjà composées (ex: tectonique des plaques). Un langage post-coordonné quant à lui comprend des expressions recomposées selon les besoins. Dans le cas des langages post-coordonnés, on associe plusieurs mots-clés aux documents. Cette distinction s'effectue au moment où a lieu la représentation des sujets, c'est à dire au moment de la construction du langage lui-même, ou au moment de son utilisation par un indexeur. C'est cette règle d'usage qui est la plus courante dans l'utilisation des thésaurus mais c'est aussi celle-ci qui a été adoptée dans l'utilisation de nos taxonomies. C'est cet usage particulier et original qui fait l'intérêt de ces outils. Pour illustrer ces propos voici une

carte du CheBoCar,³⁰ taxonomie en physique pour laquelle nous avons entouré les différents descripteurs utilisés pour l'indexation d'un document sur le site de culture sciences physiques. La fiche LOM dont je tire ces informations est en annexe.

30 La version du CheBoCar n'est pas celle qui a été revue lors de mon stage, qui n'est pas encore utilisée. Ici c'est la version précédente (V-0.2) qui est présentée.



Exemple d'utilisation du CheBoCar v-0.2



Exemple d'utilisation du CheBoCar v-0.2

3. Amélioration de ces outils

Les deux premières semaines de mon stage, j'ai essayé de faire une synthèse de toutes mes observations sur les taxonomies disciplinaires. Parallèlement j'ai rencontré une grande partie des acteurs du projet ce qui m'a aidé à compléter cette première vision. Pour cela, j'ai pris une à une chaque taxonomie et préparé des questions aux scientifiques référents de chaque discipline. Gérard Vidal m'a présenté un ou deux (pour les sciences de la vie et la chimie) référents scientifiques pour chaque discipline. Certains font partie de l'équipe des scientifiques ayant élaboré les " thésaurus ", d'autres sont arrivés plus récemment. Ils ont répondu à mes questions sur la conception et l'organisation des taxonomies et m'ont aidé à me familiariser avec le vocabulaire scientifique.

On peut résumer rapidement comment le travail de conception des taxonomies s'est effectué. Le groupe de travail acteurs de la conception, a été dès l'origine réduit à deux ou trois personnes ayant une vision globale de leur discipline. Le thésaurus est ensuite validé et corrigé par des pairs de la disciplines. Ensuite il est proposé à des utilisateurs pour un test. Nous avons repris la même méthodologie lors de l'amélioration effectuée lors de mon stage, à la différence que la phase de test n'a pas été effectuée.

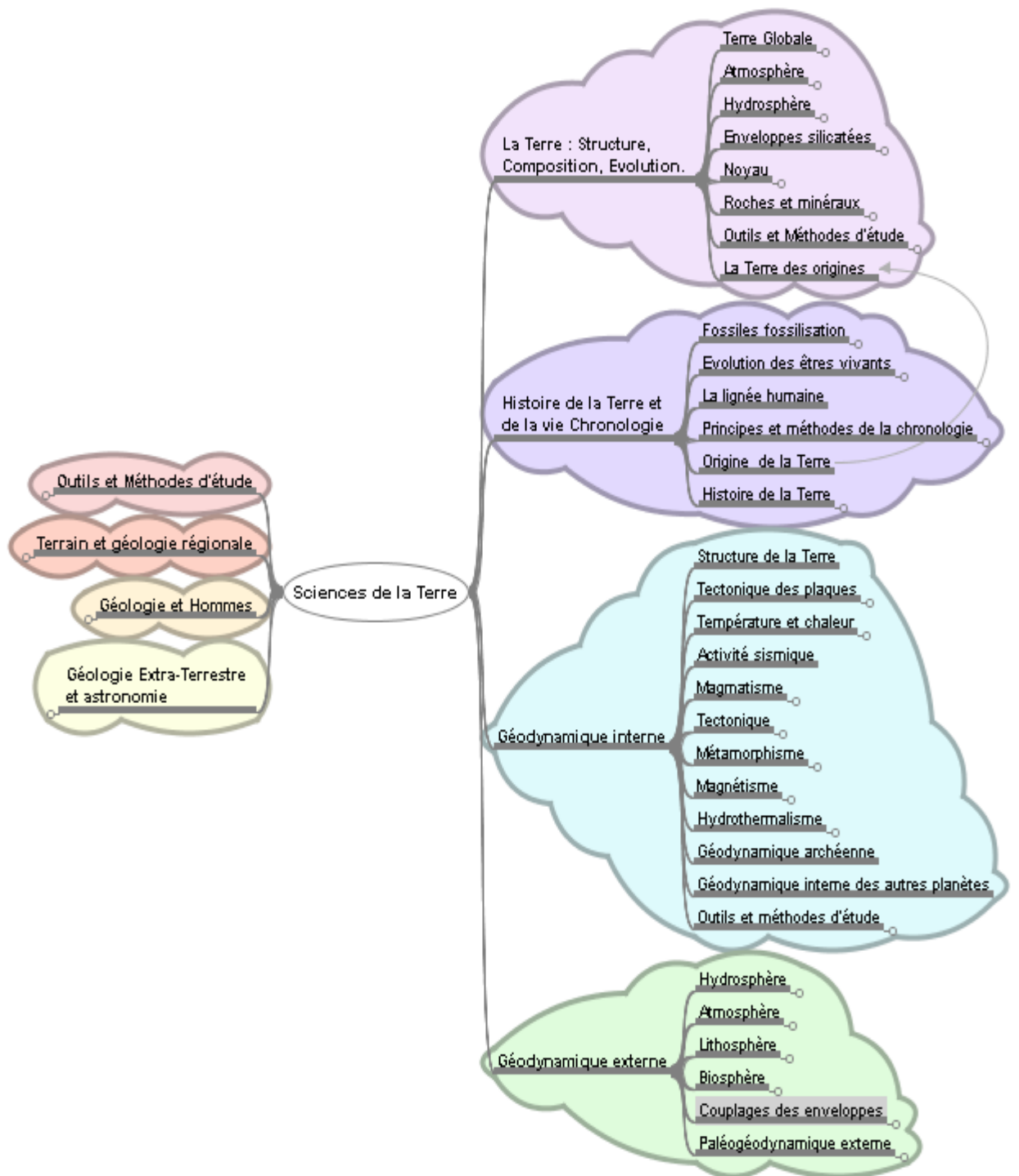
Cette première étape d'observation et de rencontre m'a permis de faire des propositions d'amélioration avec un regard neuf. Les premières ont principalement porté sur : des rajouts de termes, des demandes de précisions, des interrogations quant à la cohérence des enchaînements et de la structuration ainsi que des modifications de présentation.

Dans un premier temps, mes remarques se sont portées au niveau des termes en eux-mêmes, afin de les préciser. En effet n'étant pas scientifique de formation, je portais un regard neuf sur la discipline, aucun terme n'allant vraiment de soi selon mon point de vue. Par exemple, dans le JaHe, le nom de la taxonomie en

Sciences de la Vie, une des grandes sous-parties était nommée "Fonction". Pour un biologiste, il allait de soit que cette partie portait sur les grandes fonctions physiologiques (reproduction, nutrition...), ce qui n'est pas le cas pour un non-spécialiste de la discipline. Nous l'avons modifié en "fonction physiologique".

Dans un deuxième temps, j'ai fait des propositions d'amélioration de la structure et de la présentation. Notamment dans le ThoKaVi, la taxonomie en sciences de la Terre, il existait une grande classe "Terrain et géologie régionale" dans laquelle les ressources étaient divisées en deux régions géographiques: la "France" et le "Reste du monde". Ces deux sous-parties étant elles mêmes divisées respectivement en régions et en continents. Cette structuration m'a paru incohérente et peu rigoureuse, elle correspondait en fait à la réalité du fond documentaire. Nous avons donc décidé de diviser cette partie en géologie par continent puis de rajouter les océans.

Mais en revenant sur la taxonomie, les scientifiques ont eu beaucoup plus de modifications que moi à faire. Leurs remarques étant complètement différentes en fonction de leur utilisation et de leur pratique d'indexation, ou de leur vision et de leur spécialité au sein de la discipline. Les "indexeurs" avaient des remarques assez pratiques de part leur expérience de l'utilisation de l'outil et de la recherche documentaire. Les "non-indexeurs" qui étaient présents lors des réunions avaient davantage de remarques sur des aspects de la discipline manquant dans la taxonomie. Ce travail d'amélioration a été effectué en concertation grâce à la réunion d'une équipe de scientifiques.



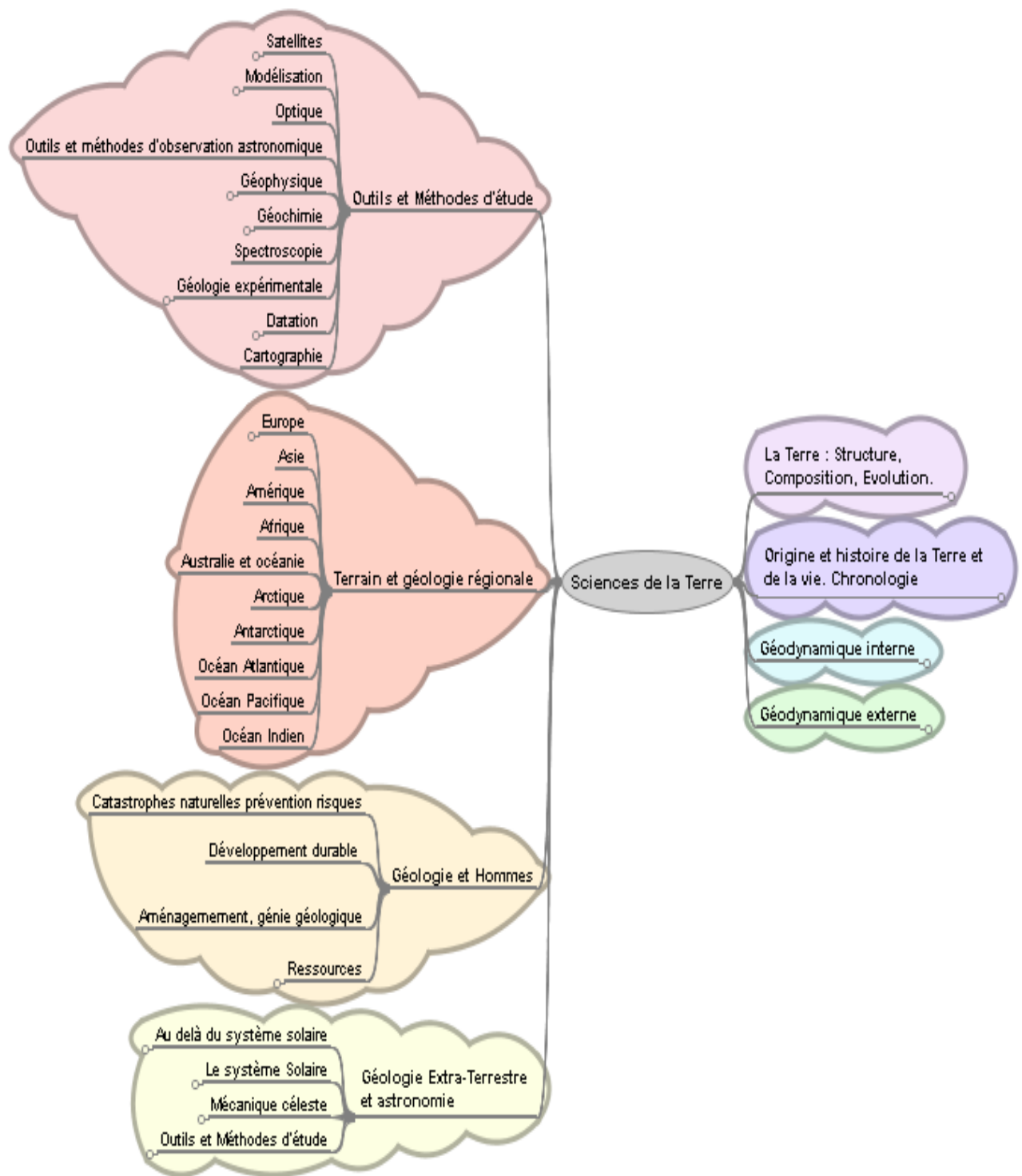
Carte du ThoKaVi 0.7 tel qu'il était à mon arrivée



Carte du ThoKaVi 0.7 tel qu'il était à mon arrivée



Carte du ThoKaVi 1.0 tel qu'il était après les modifications effectuées lors de mon stage



Carte du ThoKaVi 1.0 tel qu'il était après les modifications effectuées lors de mon stage

Par ailleurs nous avons pu remarquer que la façon d'envisager la terminologie et la discipline étaient très différentes en fonction des différentes sciences. Ces remarques peuvent trouver plusieurs explications. La première cause peut être tout simplement liée à la spécialité des scientifiques associés au projet. La seconde peut provenir de la façon dont la discipline est enseignée et de la façon dont elle conditionne la façon de penser. Cette dernière remarque ne peut pas servir de généralité, le panel étant trop étroit, mais c'est une question intéressante. Ces différentes conceptions rendent difficiles les liens inter-disciplinaires, si ceux-ci ne sont pas le but premier de ces taxonomies, ils pourraient avoir leur intérêt pour certains sujets interdisciplinaires présents dans les programmes comme l'écologie ou l'astronomie. Nous avons effectué quelques essais de liens entre les taxonomies comme je l'expliquerai dans la partie suivante.

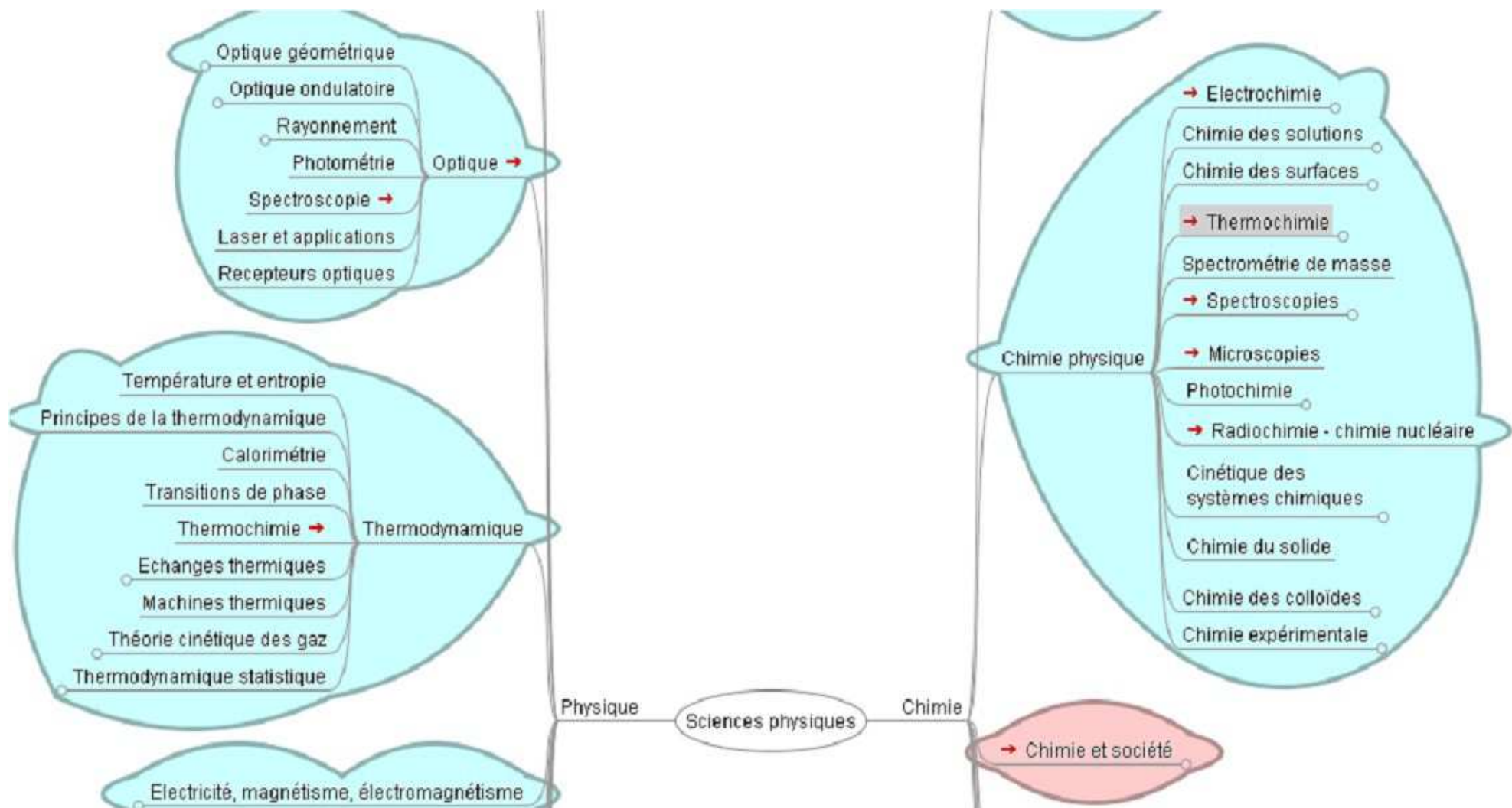
3.2.4 Proposition pour des améliorations

Grâce au travail effectué, les taxonomies disciplinaires ont été publiées dans une version 1.0. J'ai préconisé une remise à jour régulière tous les 2 ans. Dans certaines disciplines notamment en sciences de la vie et en chimie, je pense que le besoin de remise à jour va se faire sentir d'ici moins de deux ans. En sciences de la vie, ce problème provient de la discipline même. En ce qui concerne la chimie, cette taxonomie est la plus récente et les personnes l'ayant élaborée avaient peu d'expérience de l'indexation, il est donc probable que les chimistes se rendent compte des problèmes à l'usage.

Je me suis demandée quels pouvaient être les freins au travail avec les taxonomies par de nouveaux utilisateurs. J'en ai recensé plusieurs dont la crainte de voir un allongement important des tâches de production du document ainsi qu'une mauvaise appréhension de celles-ci. Un effort pédagogique important est à faire en direction des nouveaux auteurs-utilisateurs des taxonomies pour les sensibiliser à l'intérêt pour l'utilisateur final. Cette tâche peut s'avérer difficile, en effet certains enseignants producteurs de ressources voient ce travail comme une perte de temps et ne correspondant pas à leur mission première. Ce travail à effectuer au sein des

équipes est important si on veut obtenir une utilisation plus étendue des taxonomies.

J'ai fait des propositions et amorcé un travail de liens internes au sein de la taxonomie. Mais aussi entre elles, particulièrement celles de disciplines enseignées conjointement dans le secondaire. Ainsi, j'ai fait des ébauches de liens entre les sciences de la Terre et de la vie puis entre la physique et la chimie. Ces liens devraient permettre un rapprochement et une unification des terminologies sur certains sujets proches et interdisciplinaires, c'est le cas du développement durable, de l'environnement ou de l'astronomie. J'ai effectué quelques essais que j'ai ensuite soumis aux scientifiques, par exemple j'ai fait un lien entre "chimie quantique" et "physique quantique" ou encore "chimie et société" et "physique et société" mais les scientifiques m'ont fait remarquer que ces concepts ne sont pas directement liés. Les liens sont signalés par de petites flèches rouges dans le schéma ci-dessous. ces flèches permettent de faire des liens dynamiques dans Freemind. Ce travail a été seulement amorcé, ces liens internes ou non doivent être continués par des scientifiques de la discipline.



Extrait de la carte des sciences physiques

3.2.5 Taxonomies, thésaurus et ontologies : interrogations, recherches, perspectives

Gérard Vidal avait de nombreuses interrogations sur les langages documentaires, après mon travail de définition de l'existant, il m'a demandé de faire une synthèse d'articles sur les ontologies et de donner un avis sur l'opportunité d'étendre le travail déjà effectué sur les langages documentaires à un outil intelligent de ce type. A la suite de mon travail de recherche et de synthèse à ce sujet, je suis arrivée à la conclusion, que ce n'est pas quelque chose d'opportun que de se doter d'un outil aussi complexe lorsque les utilisations sont simples. Ce travail a donné lieu à une synthèse que j'ai élaborée tout au long de mon stage. (en annexe)

3.3 Vers des correspondances

3.3.1 Pourquoi des correspondances?

Les correspondances ont deux origines. La première provient simplement de la recommandation du LOM d'utiliser la Classification Décimale de Dewey. Cette recommandation voit son intérêt du fait de l'utilisation massive dans les bibliothèques universitaires de cette classification. Il a paru évident et pertinent d'élaborer des correspondances entre la classification décimale de Dewey et les terminologies disciplinaires déjà utilisées. La seconde provient d'une idée qui est apparue lors de discussions à l'occasion de journées de formation organisées par l'INRP. En janvier dernier, Gérard Vidal et son équipe ont organisé deux jours de formation sur l'indexation et le LOM qui s'intitulaient: "*Description, indexation, partage et mutualisation de ressources pédagogiques numériques*"³¹

31Des enregistrements ont été fait lors de ces journées de formation disponibles sur le site internet de l'INRP [<http://www.inrp.fr/INRP/formation-de-formateurs/catalogue-des-formations/formations-2006-07/description-indexation-partage-et-mutualisation-de-ressources-numeriques/description-indexation-partage-et-mutualisation-de-ressources-pedagogiques-numeriques>][Formation INRP]

Une présentation des taxonomies disciplinaires encore appelées "thésaurus" a eu lieu et c'est lors d'une discussion à la fin de cette présentation que l'idée d'élaborer des correspondances avec d'autres langages documentaires est apparue.

3.3.2 Origine du travail de correspondance avec la CDD et définition de la méthodologie

L'origine des correspondances avec cette classification provient d'une recommandation du LOM. En effet la Classification Décimale de Dewey étant une classification massivement utilisée et connue de tous les professionnels, cette mise en place est plus aisée. A propos de l'interopérabilité au sein du LOM, on peut citer [Noël, 2007] "*L'interopérabilité nécessite des vocabulaires communs pour que les échanges puissent se faire. Le problème est alors de concilier cette interopérabilité avec des vocabulaires qui, pour être utiles à la communauté concernée, doivent être suffisamment précis et détaillés, et doivent être partageables et partagés. Or, il y a peu de vocabulaire vraiment communs, hormis la classification Dewey ! Par exemple, si, dans le cycle primaire, il est possible et utile de détailler les tranches d'âges, le LOM ne le permet pas pour les adultes, ce qui n'a plus de sens pour la description de ressources pédagogiques destinés au public universitaire.*" La question d'une mise en équivalence bijective a été soulevée à plusieurs reprises lors de la présentation de ce travail de correspondance entre la CDD et les taxonomies "maison". Ces correspondances n'ont pas été conçues comme une mise en équivalence bijective mais comme injective.

Par ce terme d'injective nous entendons une projection ou une traduction de la taxonomie vers la CDD. Autrement dit, les correspondances effectuées entre les taxonomies disciplinaires et la CDD ne fonctionnent que dans un seul sens, de la taxonomie vers la CDD. L'inverse ne peut pas fonctionner, il est impossible de passer de la CDD aux taxonomies: plusieurs indices de la CDD correspondent à différents et parfois à de nombreux termes de la taxonomie.

Les premières semaines du stage j'ai rencontré de nombreuses personnes qui

m'ont aidé à m'approprier ce sujet. Ainsi les conservateurs des bibliothèques de l'ENS, Jacqueline de Condappa et de l'INRP, Philippe Marcerou, m'ont fourni une aide précieuse en me présentant leur personnel ayant une connaissance de la Classification Décimale de Dewey. Ils m'ont aidée dans mon utilisation de la classification. Celle-ci étant utilisée à la bibliothèque de l'ENS exceptée pour le fonds en mathématique et en informatique. François Deboclar m'a donné des conseils à ce propos. La bibliothèque de l'INRP utilise la Classification Décimale Universelle, mais Philippe Marcerou m'a présenté une bibliothécaire connaissant bien la Classification Décimale de Dewey, Sophie Wagnon ayant travaillé au sein de bibliothèques où celle-ci est utilisée.

Parallèlement j'ai rencontré Monique Joly, directrice de l'INSA de Lyon qui a mené un projet semblable par l'élaboration pour l'Université numérique en sciences et techniques de l'ingénieur (UNIT). Ce projet consistait en l'élaboration d'une classification à deux niveaux construite en relation avec la classification décimale de Dewey afin de permettre au sein de l'UNT un échange des ressources. La classification UNIT est limitée à 200 disciplines et à deux niveaux à l'intérieur de celles-ci afin de faciliter l'accès à l'information dans le portail par un nombre de "clic" le plus réduit. L'objectif du portail UNIT est de permettre un accès simple aux ressources numériques avec des interfaces utilisant la classification UNIT. Je vais détailler la méthodologie de ce projet dans la partie suivante. Ces rencontres m'ont aidée à mettre en place une méthodologie et à avoir une meilleure connaissance de la Classification Décimale de Dewey.

3.3.3 Élaboration de la méthodologie : étude du projet de classification UNIT

Pour élaborer notre méthodologie et suite à la rencontre et aux documents fournis par Monique Joly, nous nous sommes appuyés sur le travail qui a été effectué à l'INSA de Lyon. Nous allons mettre en perspective la méthodologie d'UNIT et ce que nous avons décidé d'utiliser au sein de celle-ci.

Le premier travail effectué pour l'élaboration de la classification UNIT a été de

cerner le champ de la connaissance à classifier par la mise en évidence des grandes disciplines d'après l'étude des programmes et des classifications utilisées par les sociétés savantes. Dans chacune des disciplines un certain nombre de sous-disciplines ont été identifiées. Afin de créer un outil d'aide à l'indexation, des mots-clés ont été associés à chaque discipline et sous-disciplines retenues. Toute cette première partie d'élaboration de la classification a été effectuée à l'ENS et à l'INRP par les scientifiques eux-mêmes, comme je le décris dans la partie précédente.

La deuxième grande phase a consisté en la validation par des groupes d'enseignants et de chercheurs pour la validation d'un nombre restreint de sous-disciplines, dix au maximum pour chaque grande discipline. Ce travail a été suivi de celui de correspondance entre la classification UNIT et la CDD et de la validation de celles-ci. Pour certains libellés de la classification UNIT, un seul indice Dewey est retenu mais pour d'autres libellés plus globaux, plusieurs indices ont été associés. D'autre part un même indice peut se retrouver à différents niveaux de la classification. Cette phase de travail a été celle que j'ai effectuée lors de mon stage. Nous avons repris en partie de la méthodologie d'UNIT concernant l'utilisation de la CDD mais pas en ce qui concerne la limitation du nombre de disciplines et sous-disciplines. Dans nos tableaux de correspondance, on retrouve l'utilisation de la CDD d'UNIT. De la même façon plusieurs indices Dewey peuvent correspondre à un même terme de la taxonomie et les mêmes indices peuvent apparaître à différents niveaux.

En ce qui concerne l'utilisation de la classification, chaque ressource doit être décrite avec un indice de la classification UNIT. À chaque terme de la classification a été attribué un indice et des mots-clés libres. Ensuite chaque ressource est décrite par un bibliothécaire par un ou plusieurs des indices de la CDD présents dans le tableau de correspondance. Les taxonomies produites à l'ENS et à l'INRP ne sont pas quant à elles utilisées exactement de la même façon. Il existe des mots-clés libres attribués par l'auteur de la ressource mais la description par la taxonomie entraîne automatiquement l'indexation en Dewey, il n'y a pas l'intermédiaire d'un bibliothécaire ou d'un documentaliste. La taxonomie possède également un système d'indice.

Les ressources d'UNIT, seront accessibles sur leur portail grâce au moteur de recherche ORI, développé dans le cadre de l'inter-UNT. Ce moteur de recherche permet de rechercher dans les ressources avec la classification UNIT mais aussi à partir d'indice de la CDD.

Les éléments repris de cette méthodologie ont constitué le socle du travail de correspondance mais la classification UNIT et les taxonomies "*maison*" n'ont pas du tout été conçues de la même façon. Ces dernières comprenant de trois à six niveaux alors que la classification UNIT est limitée à deux niveaux. Cette différence de niveaux a rendu le travail assez différent, ne nous limitant pas aux seules grandes disciplines et sous-disciplines, nous nous sommes retrouvés avec des impossibilités de correspondances et les limites de la CDD.

3.4 Réalisation des correspondances

3.4.1 Accès et version de la CDD

J'ai eu des difficultés d'accès à la dernière version de la CDD, soit la vingt deuxième édition. J'ai tâtonné pendant un mois, j'ai d'abord travaillé avec la version abrégée de cette dernière édition, la seule version traduite en français, que Jacqueline de Condappa, conservatrice de la bibliothèque de l'ENS m'a permis de consulter. Je me suis rendue compte des limites du travail avec une version abrégée, et des manques pour une utilisation à un niveau scientifique. Aussi j'ai travaillé avec la version intégrale en anglais de la dernière édition en ligne, d'abord dans sa version d'essai gratuite puis dans sa version payante.

3.4.2 Méthodologie des correspondances

J'ai choisi de me limiter aux trois premiers niveaux des classifications pour ma correspondance dans un premier temps, c'est à dire de me limiter aux niveaux les plus généraux mais je me suis vite rendue compte pour que certains domaines leur

précision était plus importante que celle de la CDD. J'ai attribué un code à chacun des termes de ces trois premiers niveaux pour faciliter leur traitement. J'ai ensuite commencé à effectuer les correspondances avec la version abrégée, puis lorsque j'ai eu accès à la version intégrale, j'ai complété ou effectué directement les correspondances à partir de la version intégrale comme pour le TaCBaHa, la taxonomie de chimie. Dans mes tableaux de correspondance j'indique : l'indice Dewey correspondant, l'intitulé en français suivi de celui en anglais³². Toutes mes correspondances ont ensuite été revues avec un scientifique de la discipline, afin de corriger les contre-sens ou de préciser en effectuant à nouveau des recherches dans la CDD permettant ainsi des correspondances plus précises grâce à une meilleure connaissance du sujet. J'ai à ce moment étendu mon travail à l'ensemble des classifications sans plus me limiter au troisième niveau. Mon travail a été ensuite validé par une équipe de scientifiques de la discipline.

³² voir annexe extrait d'un tableau de correspondances

3.4.3 Correspondances entre taxonomies disciplinaires et les thésaurus de l'éducation

Nous avons décidé de faire des correspondances entre les taxonomies disciplinaires et des thésaurus utilisés par différents acteurs de la documentation scientifique pédagogique. L'offre de thésaurus dans le domaine de l'éducation est assez important.

| Thésaurus | | Organisme | Dernière remise à jour |
|-----------|--|---|------------------------|
| TESE | Thésaurus des systèmes éducatifs en Europe | Réseau Eurydice | 2006 |
| TEE | Thésaurus européen de la formation | Commission Européenne et le conseil de l'Europe | 1998 |
| ETB | European Treasury browser | European Schoolnet | 2002 |
| TEF | Thésaurus Européen de la Formation | Cedefop | 2006 |
| Mot Bis | | CNDP | 2007 |

Notre choix s'est arrêté sur l'ETB et Mot Bis. Gérard Vidal avait entendu parler de l'ETB par des homologues allemands dans le cadre du projet Formation collaborative en Ligne-Pairform@nce. Ce projet est soutenu en France par la SDTICE du ministère de l'éducation nationale et Intel Éducation afin de permettre l'échange de documents et de métadonnées au niveau européen par une utilisation du LOM. Mot Bis est un thésaurus largement utilisé par les centres de documentation de l'éducation nationale. Nous allons présenter ces deux thésaurus puis la méthodologie utilisée pour effectuer des mises en équivalence entre ceux-ci et les taxonomies disciplinaires.

3.4.3.1 ETB (European Treasury Browser)

Le thésaurus ETB a été créé par l'European Schoolnet pour un classement des matériels portant sur les contenus éducatifs et particulièrement pour les ressources pédagogiques multimédias. La particularité principale de ce thésaurus réside dans son multilinguisme. Il a été élaboré en cinq langues puis étendu à huit. Ce thésaurus a été mis à jour pour la dernière fois en 2002.

3.4.3.2 Thésaurus Mot Bis

Mot Bis est un thésaurus pour l'éducation française, il a été produit et est maintenu par le Centre National de Documentation Pédagogique, qui fait paraître une mise à jour tous les ans. Ce thésaurus est massivement utilisé par les centres de documentation de l'éducation nationale: les Centres de documentation et d'information (CDI) des collèges, lycées et de l'enseignement agricole, les IUFM Institut Universitaire de Formation des Maîtres, les CAFA Centre Académique de Formation de l'Administration et l'ensemble du réseau SCEREN CNDP Centre National de Documentation Pédagogique, CRDP, Centre Régional de Documentation Pédagogique, CDDP Centre Départemental de Documentation Pédagogique, CLDP Centre Local de Documentation Pédagogique.

3.4.3.3 Méthodologie

La méthode utilisée pour effectuer les correspondances entre les taxonomies et les deux thésaurus présentés plus haut a été simple. J'ai repris chaque terme des taxonomies disciplinaires en essayant de trouver un équivalent dans les thésaurus. L'ETB ne se place pas au même niveau de précision que les taxonomies "*maison*" son objectif n'étant pas une classification de ressources scientifiques mais pédagogiques, aussi je suis restée à un niveau très général. En ce qui concerne Mot Bis, j'ai pu trouver davantage de détails mais la fonction de ce thésaurus étant l'indexation de ressources encyclopédiques je suis également restée à un niveau

général.

3.5 Bilan du travail de correspondances

3.5.1 Un Bilan chiffré

Les taxonomies en sciences de la Vie et en Physique sont celles qui ont trouvé le plus de correspondances dans les autres langages documentaires. La chimie est celle qui a trouvé le moins de correspondance particulièrement avec les thésaurus spécialisés dans l'éducation. Il est probable dans le cas des correspondances avec les thésaurus de l'éducation que l'ont puisse expliquer la lacune des thésaurus par le fait que la chimie et les sciences de la terre ont une place moins importante dans les programmes scolaires, au sein desquels elles sont moins représentées que la physique et les sciences de la Vie avec lesquelles elles sont enseignées. On peut aussi se demander si ces différences ne proviennent pas de méthodes de travail différentes en fonction des disciplines. En physique et en science de la vie, le travail a été effectué ensemble ligne à ligne alors que en sciences de la terre et en chimie le travail s'est davantage effectué par échange de mails ou par l'intermédiaire du site de travail collaboratif.

Tableau comparatif des correspondances des taxonomies disciplinaires

| Taxonomies | nombre de termes | nombre de correspondances CDD | nombre de correspondances ETB | nombre de correspondances Mot Bis |
|------------|------------------|-------------------------------|-------------------------------|-----------------------------------|
| ThoKaVi | 506 | 165 | 44 | 105 |
| CheBoCar | 327 | 157 | 35 | 135 |
| TaCBaHa | 460 | 120 | 18 | 70 |
| JaHe | 453 | 235 | 32 | 225 |

3.5.2 Validation scientifique

3.5.2.1 Méthodologie

La validation scientifique a eu lieu à plusieurs étapes de mon travail. Les premières validations ont eu lieu lors de modifications que j'ai apportées en collaboration avec un ou deux scientifiques. Pour certaines disciplines l'intégralité de la taxonomie a été reprise et modifiée comme en science de la vie (JaHe) ou en chimie (TaCBaHa), alors que pour les sciences de la terre (ThoKaVi) et la physique (CheBoCar), les modifications n'ont pas touché la structure principale mais celles-ci se sont limitées à des ajouts, des précisions ou des modifications de certains termes par des synonymes plus explicites.

La deuxième phase de validation a été effectuée pour les correspondances entre les taxonomies et la classification décimale de Dewey, je soumettais mon travail à un scientifique "référent" de la discipline pour la taxonomie, étant utilisateur quotidien ou concepteur de celle-ci. Ce travail nous amenait à vérifier les correspondances une à une et souvent à refaire des recherches dans la classification décimale de Dewey afin de trouver l'indice le plus proche du terme de la taxonomie. Prenons un exemple, en ce qui concerne la taxonomie en sciences physiques, nous avons du rechercher les différentes sortes d'ondes à associer et sans une connaissance de la discipline, il était difficile de faire ce travail. La troisième phase de validation avait lieu avec un groupe de scientifiques par discipline pour valider l'ensemble des correspondances.

3.5.2.2 Particularités et richesses de cette collaboration

La familiarisation avec le vocabulaire scientifique, au moins avec celui des taxonomies, a pris quelque temps mais a été enrichissante. D'autre part, mon regard neuf sur la taxonomie et plus généralement sur les différentes disciplines, n'étant pas scientifique de formation, a permis un travail de précision ou d'explicitation des

termes utilisés. La collaboration avec les scientifiques a été quelque chose d'assez facile, mis à part les difficultés de vocabulaire. Cette question de vocabulaire a eu deux aspects : un aspect positif par la précision nécessaire au dialogue entre personnes ne provenant pas d'une même discipline ; un aspect négatif par les difficultés et/ou incompréhensions que cela a pu générer. Prenons un exemple concret, dans la taxonomie en sciences de la vie, les scientifiques ont créé une classe "Fonctions". Pour un biologiste il est évident qu'il s'agit des grandes fonctions biologiques, pour un novice dans cette discipline, cela n'a rien d'évident. J'ai aussi pu faire un certain nombre de remarques sur les différences de langage et de vision de leur propre discipline par les scientifiques.

3.5.3 Rencontre CNDP

A la fin du mois de mai, j'ai cherché qui pourrait valider le travail ou simplement avoir connaissance de ce que j'avais fait avec Mot Bis. J'ai pris contact par le réseau de relations de Gérard Vidal avec des personnes du CNDP. Une rencontre a été organisée le 22 juin, soit deux semaines avant la fin du stage, pour une présentation de mon travail. La préparation de cette présentation m'a permis de faire un bilan sur mon travail. Les personnes présentes lors de cette rencontre : Rosa Maria Gomez de Regil, et d'autres personnes en charge du thésaurus Mot Bis: Sandrine Jambert, Teresa Bremant, François Feyler ainsi que Sylvie Dalbin, experte en langages documentaires qui effectue actuellement une mission de consultante pour l'amélioration du thésaurus.

La rencontre a été très riche, j'avais eu jusqu'à ce moment peu de regards de professionnels de la documentation sur mon travail. Aussi cette occasion a été profitable à l'amélioration de mon travail mais il m'a aussi permis de me rendre compte de ses limites. Sylvie Dalbin et François Feyler m'ont fait remarquer les problèmes que posaient la correspondance d'une classification et d'un thésaurus.

La pertinence des correspondances entre les taxonomies disciplinaires et les thésaurus a été remise en cause. Mais pas celles avec la Classification Décimale de Dewey qui ont toute leur légitimité. A propos de ces dernières, Sylvie Dalbin a

souligné les problèmes qui tiennent à l'utilisation de la CDD aujourd'hui et particulièrement pour des sujets très pointus ou interdisciplinaires. J'ai effectivement rencontré des problèmes pour effectuer des correspondances sur des sujets interdisciplinaires comme l'environnement et le développement durable comme déjà expliqué un peu plus haut. Les personnes présentes lors de cette rencontre m'ont donné des conseils pour l'amélioration de mon travail notamment la mise en place de tests de celui-ci.

3.5.4 Difficultés et interrogations rencontrées lors du travail de correspondances

Lors du travail de correspondances, des questions de plusieurs ordres se sont posées. En premier lieu, nous avons rencontré des problèmes que l'on peut qualifier de techniques, il s'agissait d'utilisation et de limites de la Classification décimale de Dewey. Dans un deuxième temps, les correspondances avec les thésaurus ont posé des problèmes d'ordre théoriques et conceptuels. C'est lors de la rencontre du CNDP(Centre National de Documentation Pédagogique) que l'on a souligné et amené à réfléchir à ces problèmes.

Nous avons parfois rencontré des différences de niveau de précision entre la CDD et nos taxonomies disciplinaires: impossibilité de trouver une correspondance pour un aspect très général ou trop spécifique. Les classifications bibliographiques comme la classification décimale de Dewey ont été les premiers outils de représentation du contenu thématique des documents. Sa conception en dix grandes classes principales limitant une extension importante donne une mauvaise image de la classification décimale de Dewey auprès des scientifiques qui y voient un outil trop ancien et dépassé. J'ai du faire un effort de pédagogie afin de leur expliquer le fonctionnement de la CDD et ses particularités, certains étant surpris que Dewey ait mis sur un même plan des disciplines et des problèmes pratiques. Je cite un exemple tiré de: [Maniez, 1994] 6 sciences appliquées, technologies 61 sciences médicales (division logique) 616 Maladies (partition) 616.8 Maladies du système nerveux (division logique)

Pour citer des problèmes précis rencontrés lors du travail de correspondances, ils ont été particulièrement nombreux avec la taxonomie en sciences de la vie. En effet celle-ci prend les termes avec un point de vue fonctionnaliste, il a été vraiment difficile de trouver une correspondance. Par exemple dans la taxonomie : le terme "Locomotion" pris sous sa forme nominale et qui implique une idée de mouvement, n'a pu être traduit en Dewey que par "573.7 Système musculosquelettique" qui traduit seulement ce qui permet la locomotion. Les problèmes rencontrés lors de ce travail ont aussi été liés à la conception même de la classification, à la fréquence de ses mises à jour qui ne suit pas la vitesse des découvertes scientifiques, par une mise à jour tous les sept ans. Des sujets de préoccupation assez récents comme l'environnement et l'écologie sont représentés dans des classes très diverses de la CDD. Aussi bien en science avec des classes générales (500) qu'avec des classes plus spécialisées en 577, classe qui traite d'écologie que dans les classes traitant d'économie (300) que de techniques et sciences appliquées (600).

Les correspondances entre les taxonomies disciplinaires et les thésaurus posent davantage de problèmes et de difficultés que celles avec la Classification Décimale de Dewey. En effet classification et thésaurus ne se basent pas sur le même mode d'accès à l'information. Les thésaurus permettent de répondre à des questions posées sous forme de requête par mot-clés. Chaque terme représentant un concept alors que dans le cas d'une classification les termes sont représentés sous forme de classe. La concordance entre classe et concept n'est pas évidente et peut apparaître comme un appauvrissement par rapport à une indexation complète avec des concepts. Tout ceci donne lieu à de nombreuses interrogations.

Nous avons décidé de reprendre l'idée de départ du travail de correspondance fait par Monique Joly et son équipe à l'INSA de Lyon, dans lequel chaque classe de leur classification peut correspondre à plusieurs classes de la classification décimale de Dewey. Nous avons suivi cette même méthode. Conceptuellement, les correspondances entre les taxonomies disciplinaires et la Classification décimale de Dewey fonctionnent bien. Mais par cette utilisation de plusieurs classes pour une seule de la taxonomie, le résultat et l'utilisation même de la CDD est entièrement

différent. La Classification décimale de Dewey est utilisé de cette façon comme un langage pour l'indexation et pas uniquement comme un langage pour classer.

En effet si un terme correspond à une classe ou une thématique dans une classification et dans une taxonomie ce n'est pas le cas dans un thésaurus. Un terme appelé descripteur dans un thésaurus correspond à un concept. Un terme dans une classification correspondant à une classe ou à un champ sémantique plus important puisqu'il comprend un certain nombre de terme-fils et de terme-père, la relation est verticale et descendante et souvent suit la logique tout/partie. A l'inverse dans un thésaurus un terme ne correspond qu'à un seul concept, qu'il prend comme unité élémentaire. Autour de ces éléments, il est construit un "*entrelacs de relations verticales et transversales dans le cadre d'un champ sémantique prédéfini*" [Maniez, 1994] dont l'organisation se fait à posteriori.

4 Note de recherche

Avant de rentrer dans le vif de la note de recherche portant sur l'interopérabilité des langages documentaires, il m'a paru nécessaire de revenir sur les raisons qui ont présidé à l'élaboration des taxonomies et aux pratiques d'indexation.

4.1 Analyse de la conception des taxonomies et des pratiques d'indexation

4.1.1 Caractéristiques des taxonomies disciplinaires

La création pour chaque besoin d'un nouveau langage documentaire est une attitude assez commune, d'après cette constatation, il a été décidé de créer un langage spécifique au regard des nombreux thésaurus déjà existants. C'est pour cette raison qu'un certain nombre de règles d'usages ont été adoptées afin que les taxonomies aient certaines originalités.

- Bien qu'ayant la présentation et certaines caractéristiques d'une classification ou d'un vocabulaire hiérarchique, les taxonomies sont utilisées comme un langage d'indexation.
- Les redondances sont acceptées. Ainsi pour chaque document les indexeurs peuvent reporter plusieurs termes de la taxonomie. Cette règle d'usage permet une indexation précise : les descripteurs se complètent les uns les autres.
- Le travail avec le logiciel Freemind permet l'organisation d'idées sous forme de carte heuristique ou carte conceptuelle. Cet outil permet uniquement un mode hiérarchique d'organisation des termes et a limité la conception de la terminologie à un plan hiérarchique sans liens entre les termes ou les disciplines.

- Le nombre de ramifications doit être raisonnable, nous l'avons limité à six niveaux de profondeur, le choix des grandes orientations s'est effectué de manière évidente du général au particulier.
- Le JaHe (taxonomie en science de la vie) privilégie les objets biologiques par rapport aux disciplines, ainsi que les fonctions et les descriptions des êtres vivants, ce qui présente une certaine originalité de la vision sur la discipline.

4.1.2 Une élaboration collaborative

La conception et l'élaboration collaborative des taxonomies disciplinaires a permis d'impliquer une partie de la communauté des scientifiques producteurs de ressources pédagogiques. Le fait que ces outils soient construits par et pour des personnes d'une même discipline a pour objectif une meilleure adoption par les usagers. On entend parler ou on utilise des folksonomies qui sont un autre mode d'indexation des ressources par les utilisateurs eux-mêmes. Une folksonomie est un néologisme qui désigne un système de classification collaborative décentralisée spontanée. Les utilisateurs ne doivent pas respecter une terminologie prédéfinie mais ils peuvent adopter les termes qu'ils souhaitent pour classer leurs ressources. Leur intérêt et leur usage par une communauté consiste au classement d'une ressource par plusieurs contributeurs et c'est l'union de leurs classifications qui constitue la folksonomie. Les termes sont souvent appelés mot-clés, tags ou étiquettes. C'est un concept qui fait partie intégrante du web 2.0 selon l'article de Wikipédia cité dans [Francis, Quesnel, 2007]. Ces mode de conceptions sont destinés à impliquer davantage les utilisateurs. Mais dans le cas des outils élaborés ici, on ne peut pas parler de folksonomie.

4.2 Vers une interopérabilité des systèmes de descriptions de ressources?

4.2.1 L'interopérabilité entre les systèmes de descriptions des ressources

La question de la compatibilité des langages documentaires et de l'interopérabilité n'est pas une question nouvelle, dès les années soixante ces problèmes se sont posés. Dans un contexte où la communication intersystème, l'interopérabilité, la réutilisabilité et le partage sont des valeurs dominantes, le grand nombre de langages documentaires paraît quelque peu paradoxal. L'idée d'interopérabilité a été désignée par plusieurs types de termes par le passé. Nous allons essayer de donner une définition à ce concept en faisant un panorama des différentes acceptions du terme. Nous nous interrogerons ensuite sur les enjeux et la place de l'interopérabilité dans un contexte plus général ainsi que les limites de notre projet.

4.2.1.1 Définition de l'interopérabilité

Si l'interopérabilité n'est pas un souci récent, au fil des années, ce concept n'a pas toujours été défini de la même manière. Dans son article Jacques Maniez [Maniez, 1997] parle de compatibilité ou d'intégration, d'harmonisation, de concordance ou encore de conciliation, en considérant que tous ces termes suggèrent une idée de convergence. Pour cet auteur "*la compatibilité des langages n'est pas une qualité naturelle, mais plutôt un objectif à atteindre, et qu'elle comporte plusieurs degrés*". Danièle Degez [Degez, 1998] dans un article écrit quelques mois plus tard rajoute les termes, de traductibilité, de convertibilité au concept de conciliabilité. Elle donne comme définition "*ce qui peut s'accorder avec autre chose, exister en même temps*". Michèle Hudon [Hudon, 2005] dans son article définit l'interopérabilité "*comme la capacité pour un système de travailler de concert avec*

un autre système, et ce sans que leurs opérateurs ou utilisateurs respectifs aient à fournir un effort particulier. L'interopérabilité repose sur l'ouverture, le partage, l'adaptation, la réconciliation des composantes, des politiques et des pratiques". Elle souligne également que l'interopérabilité n'est pas limité à un aspect technique mais a aussi une dimension politique, légale, disciplinaire et sémantique. Dans la documentation plus récente comme dans le grand dictionnaire terminologique ³³ Cette définition est donnée : "Capacité que possèdent des systèmes informatiques hétérogènes à fonctionner conjointement, grâce à l'utilisation de langages et de protocoles communs, et à donner accès à leurs ressources de façon réciproque. Note : L'interopérabilité implique qu'un programme tournant sur un système ouvert fonctionnera également sur un autre système. L'interopérabilité a besoin de plus qu'une bonne connectivité technique puisqu'elle nécessite l'utilisation d'éléments comme des interfaces de programmation et des formats de données standardisés. L'interopérabilité définie ici est l'interopérabilité technique, mais il en existe d'autres types dont l'interopérabilité sémantique qui est associée à un mode de description de l'information contenue dans une base de données (cette description forme les métadonnées)." L'article de Muriel Foulonneau [Foulonneau, 2005] différencie plusieurs types d'interopérabilité documentaire. « L'interopérabilité sémantique porte sur l'adéquation des concepts de description de la ressource. L'interopérabilité doit aussi être syntaxique afin de permettre une même présentation des données. Cette unité de présentation doit aussi être identique si le document est dans une autre langue afin de permettre une interopérabilité linguistique. »

J'ai retenu de ces définitions un certain nombre d'éléments sur les nuances que revêtent ces concepts. Le concept d'interopérabilité est utilisé depuis moins de temps que celui de compatibilité, je l'ai trouvé dans la littérature professionnelle plus ancienne. Cette notion d'interopérabilité recouvre une réalité plus technique que son quasi-synonyme compatibilité ou concordance qui ont un sens plus général. Les termes conciliation, harmonisation ou encore convertibilité, traductibilité concernent des réalités plus proches des langues et de la possibilité d'adapter les langages documentaires entre eux.

33(http://www.granddictionnaire.com/btml/fra/r_motclef/index800_1.asp)

4.2.1.2 Comment organiser cette interopérabilité? L'exemple du projet de correspondance et ses limites.

A partir du travail effectué sur les définitions, j'ai essayé de positionner les correspondances que j'ai effectuées. On peut retenir le terme d'interopérabilité dans son acception sémantique. La mise en équivalence de différents langages documentaires relève davantage d'un essai de concordance ou de compatibilité. Mais c'est justement cette compatibilité qui pose le plus de questions : Comment la mettre en place ? Est-elle vraiment toujours pertinente ? Est-il possible de parler de concordance ou de compatibilité ? Est-ce qu'il n'est pas plus prudent de parler de conciliation des deux langages ? Est-il possible d'arriver à une convertibilité parfaite ?

Tout d'abord revenons sur la conception de la méthodologie du travail effectué. Les correspondances ont été conçues par Gérard Vidal comme une injection et non comme une bijection. La bijection est une conception plus traditionnelle en documentation mais elle soulève des problèmes plus délicats qu'une simple injection. Celle-ci est une conception peu commune et assez originale.

La mise en correspondance ne pose aucun problème conceptuel dans le cadre d'une mise en équivalence des taxonomies disciplinaires de la Classification Décimale de Dewey, en effet ces deux langages documentaires ont en commun un même mode de hiérarchisation des termes par grandes classes. L'unité élémentaire de la classification ou de la taxonomie est le terme-sujet. Cette injection du vocabulaire "maison" et de la CDD relève d'un essai de compatibilité, mais il est peut-être plus prudent de parler de conciliation. J'ai commencé à voir les limites de ces équivalences dès que le travail avec les scientifiques a débuté. Ces limites tiennent aux limites même de l'indexation et partent de cette remarque simple: deux personnes n'indexent pas le même document de la même façon. Un travail en équipe permet de réduire ces difficultés et le point de vue des scientifiques de la discipline permettait de trouver le terme le plus proche ou à défaut la grande classe la plus appropriée, mais n'élimine pas les lacunes qu'il est possible de relever dans les résultats.

En revanche la mise en correspondance de la terminologie hiérarchique "maison" et les thésaurus pose davantage de problèmes. Une mise en équivalence entre un terme d'un langage hiérarchique et d'un thésaurus soulève des difficultés. Le terme du vocabulaire hiérarchique maison acquiert son sens grâce à sa "place" dans le classement, il peut être redondant et je le rappelle son unité de base est le terme-sujet. Le point fondamental de ces problèmes d'incompatibilité se manifeste dans la conception du terme dans le thésaurus qui est uni-conceptuel. Un terme équivaut à un seul et unique concept, ce n'est pas le cas dans une classification ni dans une taxonomie telle que nous l'avons définie. Par ailleurs, on peut souligner que l'organisation complexe d'un thésaurus qui n'est pas uniquement hiérarchique mais dont les sujets sont exprimés par une combinaison de termes empêche cette compatibilité. Nous touchons ici au problème de l'interopérabilité sémantique des langages documentaires. Selon Michèle Hudon [Hudon, 2005] *"l'interopérabilité sémantique implique la réconciliation de systèmes qui utilisent des termes différents pour décrire un même concept, ou à l'inverse des termes identiques pour représenter des concepts différents"*. Le problème majeur réside dans le fait que deux langages documentaires ne couvrant pas le même domaine de la connaissance auront une couverture sémantique différente.

La question qui se pose désormais se résume en ces termes : toutes ces correspondances pour rien ? Je pense que non. En effet un certain nombre de correspondances entre les taxonomies et les thésaurus sont valides puisqu'elles expriment exactement le même concept par un terme identique, dans ces cas là je ne vois pas en quoi la compatibilité ne serait pas valide. On peut affirmer qu'elle est partielle et que l'équivalence conceptuelle peut être une question de degré. Dans le cas d'une mise en équivalence de Mot Bis et des taxonomies en sciences de la vie et en physique, un terme sur deux correspond exactement. Il me semble que pour ce travail on peut parler d'essai de convergence de langages documentaires. Il me semble que les mises en équivalence ne peuvent être pérennes. En effet, chaque langage documentaire connaît une remise à jour régulière : la mise à jour de la Classification Décimale de Dewey est effectuée tous les sept ans, la prochaine édition est attendue pour 2012. Le thésaurus ETB n'a pas été remis à jour depuis 2002 et je n'ai pas réussi à savoir si une nouvelle version était prévue ou en cours.

En ce qui concerne Mot bis, une nouvelle édition est parue cet été, autrement dit les correspondances effectuées mériteraient déjà d'être révisées. Aussi pour assurer une pérennité des correspondances, la mise à jour régulière est indispensable. Les limites de mon travail ne me sont apparues qu'à la fin de mon stage d'une part par la rencontre organisée au CNDP³⁴ et par la lecture d'articles professionnels qui m'ont ouvert les yeux sur les problèmes conceptuels de la mise en équivalence des langages documentaires. *"La disparité entre les familles est plus forte pour les langages documentaires que pour les langues naturelles. en fonction de leur type classificatoire ou combinatoire ils utilisent pour formuler un sujet des unités sémantiques de gabarit différent: la thématique ou le concept. Les énoncés produits de part et d'autre sont de ce fait, hétérogènes : synthétiques pour les premiers, analytiques pour les seconds. Bien entendu il est possible à un indexeur de traduire une formule d'indexation d'un langage A en une formule approchée d'un langage B. Mais comme toute traduction humaine, c'est un processus qui doit être écarté a priori pour des raison économiques."* Jacques Maniez est un partisan de la conversion ou de la mise en compatibilité des formules d'indexation davantage que des langages eux-mêmes.[Maniez, 1994]

4.2.1.3 Comment organiser l'interopérabilité? Quelles sont les solutions possibles?

L'organisation d'une interopérabilité entre système d'indexation n'est pas une question nouvelle mais reste une question difficile. Je cite Jacques Maniez :*"Tout projet d'harmonisation des langages documentaires est soumis à des conditions draconiennes de similarité sémantique et structurelle, ce qui explique que les réalisations soient en nombre limitées."* Les solutions ou propositions de solutions sont multiples. Michèle Hudon [Hudon, 2005] distingue quatre solutions possibles

³⁴Sylvie Dalbin m'a expliqué son point de vue sur ce sujet dans un mail que je cite: "J'ai tendance à dire qu'il faut les deux types d'entrées - classifications et indexation descripteurs, et que l'interopérabilité est possible entre classification ou entre langages conceptuels. La discussion est ouverte..."

aux problèmes posés par la multiplicité des langages documentaires et pour une tentative de compatibilité:

- La mise en équivalence ou correspondances que Michèle Hudon appelle "mapping"
- L'intégration ou la fusion des langages "merging"
- L'utilisation d'un langage faisant transition ou de commutation (switching language). Cette solution peut s'avérer intéressante lorsque les éléments pivots sont représentés par un code plutôt que par une expression du langage naturel.
- Le développement d'un langage de type méta, super, ou macro-thésaurus.

Les deux premières solutions proposent de faire des liens entre les langages documentaires alors que les deux dernières proposent de développer de nouveaux outils. Notre travail de correspondance se situe dans le premier type de correspondance appelé mise en équivalence ou "mapping". Les autres solutions proposés dans cet article relèvent d'une autre méthodologie mais sont importantes à garder en tête pour une analyse pertinente de notre travail.

Jacques Maniez rejette à priori le travail consistant à établir une compatibilité des langages en eux-mêmes, qui lui paraît trop difficile d'un point de vue conceptuel et qu'il rejette à cause du coût humain d'un tel travail. Il propose une conversion des formules d'indexation afin de permettre une interrogation commune à différents langages documentaires.

4.2.2 Quelle est la place des usagers ?

Cette question de la place des usagers dans cette volonté globale d'interopérabilité paraît être une question fondamentale. Mais en même temps il paraît légitime de s'interroger si cette vision n'est pas uniquement celle de gestionnaire de l'information et qui ne correspond pas à ce qui est l'idéal du point de vue de l'utilisateur. Quel est l'intérêt de vouloir imposer une même "feuille de style" à tous les usagers de toutes les disciplines? Sylvie Dalbin lors de notre rencontre au

CNDP a soulevé ces interrogations. En effet cette vision globale de gérer l'information, importe peu à l'utilisateur. Par exemple un étudiant ou un enseignant en physique se soucie rarement de chercher de l'information en sciences de la Vie. Ces problèmes se posent principalement pour des professionnels de la recherche d'information. Il est possible de mutualiser les connaissances en adoptant une certaine cohérence aux données sans pour autant tout unifier.

Ces interrogations légitimes ne doivent pas occulter que ces solutions présentent des avantages pour une meilleure gestion et un échange d'information facilité. L'utilisateur en retire une plus grande facilité, à condition de ne pas l'obliger à se servir d'outil aussi peu intuitif que la Classification Décimale de Dewey.

4.2.3 Enjeux/Collaborations

Le partage des documents est un enjeu majeur dans un contexte de diffusion des connaissances en réseau. La volonté d'utilisation des métadonnées et celle de rendre interopérable les langages documentaires proviennent d'un contexte plus général de partage de l'information. La langue naturelle n'est pas adaptée pour une recherche documentaire pertinente en effet elle ne permet pas une recherche d'information sur des concepts univoques comme le permet un langage documentaire. Mais l'adoption d'un langage documentaire unique et universel n'est pas non plus une solution viable puisque cet usage ne permet pas une description suffisamment spécifique de la ressource. A ce propos on peut citer Jacques Maniez *"La recherche de la compatibilité n'est pas un problème secondaire parce qu'elle se rattache à l'objectif essentiel de la documentation: mettre en communication ceux qui cherchent des sources d'information sur un sujet précis et les auteurs susceptibles de les satisfaire."*

Les enjeux de cette compatibilité ou d'interopérabilité sont majeurs. En effet dans le contexte de travail décrit dans les parties précédentes nous avons remarqué que le travail dans le secteur de la documentation pédagogique s'effectue

particulièrement par le réseau et la mise en place des universités numériques accélère ce mouvement. On peut évoquer à ce propos la collaboration de l'ENS à l'université thématique numérique en sciences UNISCIEL, dans laquelle nos taxonomies devraient être utilisées. Une utilisation plus étendue que les seuls sites des acteurs de ce projet est prévue. En effet, les taxonomies en sciences de la vie et de la terre ont déjà connu une diffusion plus large dans le cadre de la distribution de clé USB aux enseignants du secondaire. Ainsi ils ont été familiarisés avec ce type de classement de l'information. Une diffusion plus étendue serait souhaitable pour assurer une certaine pérennité à ces terminologies.

5 Conclusion générale

Pour conclure, je voudrais faire un point sur les limites et les manques de mon travail.

Le projet de terminologie scientifique élaborée au sein de l'ENS et de l'INRP va sans doute connaître une extension. Tout d'abord par une utilisation plus étendue de celles-ci. En sciences de la Vie, des contacts ont été pris durant mon dernier mois de stage avec Gilles Camus, responsable du site DGESCO en sciences de la vie, qui n'utilise pas encore la taxonomie. Je lui ai soumis pour avis la taxonomie de biologie, le JaHe pour lequel il a pu faire un certain nombre de remarques. Mais malheureusement, il n'a pas été possible de les intégrer. En effet une nouvelle réunion avec Françoise Jauzein et Sandrine Heusser, les responsables et créatrices du JaHe n'a pas pu être organisée. J'ai recensé toutes les modifications à faire pour une prochaine version de la taxonomie. Des modifications dans la taxonomie en science de la vie seront certainement nécessaires avant une utilisation plus étendue. A propos des remises à jour futures des taxonomies disciplinaires, il me semble que les deux taxonomies de physique et de sciences de la terre sont les plus viables à long terme après les modifications que nous avons effectuées après plusieurs années d'utilisation. Ce n'est absolument pas le cas de la taxonomie en chimie qui commence tout juste à servir et a été élaborée par des personnes n'indexant pas quotidiennement pour l'instant. Aussi, il me semble probable que les chimistes-utilisateurs éprouveront le besoin de faire des modifications assez rapidement.

Le projet peut aussi connaître une extension par l'élaboration de langage d'indexation et de classement dans d'autres disciplines. Nous avons eu des contacts pendant mon stage avec la personne responsable du site DGESCO en math : une première ébauche de classement de ces ressources a été commencé, mais pour que celle-ci ressemble aux autres taxonomies, il y a encore un travail à effectuer.

Au moment de faire un bilan sur mon stage, j'ai deux regrets de tâches que je

n'ai pas pu accomplir faute de temps.

J'ai eu le projet au début de mon stage de mettre en place un petit manuel d'aide à l'utilisation des taxonomies mais il me manquait des informations techniques et pratiques. Pour rédiger au mieux un manuel de ce type, il aurait fallu qu'en plus de la connaissance du langage d'indexation, que j'ai pu acquérir par un travail important lors de la phase d'amélioration, que j'indexe des documents ou que je travaille avec un scientifique qui indexe quotidiennement. Un autre élément qui aurait pu m'aider dans cette rédaction est une meilleure connaissance de l'outil utilisé pour renseigner les métadonnées des documents produits, à savoir Cinematic (voir partie 2-4 la présentation de cinematic).

D'autre part je regrette d'avoir manqué de temps pour le test de mon travail. D'une part l'amélioration des taxonomies aurait mérité un test de la part d'utilisateurs ainsi qu'un examen plus attentif de son utilisation pour la recherche sur les sites où elles sont utilisées. Ainsi dans Planet Terre, le ThoKaVi est entièrement utilisé pour une recherche directe dans la taxonomie. Il aurait été intéressant d'observer attentivement cet usage. Ce test n'a pas été possible avant mon départ, le changement de version nécessitant une réindexation des ressources, l'utilisation des taxonomies ne sera effective qu'à l'automne. D'autre part mon travail de correspondance n'a pas pu non plus être testé par manque de temps. La phase de test, si elle a lieu pourra aboutir à un certain nombre de conclusions sur la finalité des correspondances et la justesse des suppositions que j'ai pu faire dans la note de synthèse. Cette étape de test est capitale pour une viabilité à moyen terme de ce travail.

6 Bibliographie

Sur la documentation pédagogique en général :

* [Bourda, 2007] Yolaine Bourda, Pourquoi indexer des ressources pédagogiques?, [en ligne] Babel edit, L'indexation de ressources pédagogiques numériques, ENSSIB, janvier 2006. [<http://babel.enssib.fr/document.php?id=41>] [Consulté le 4 juillet 2007]

* [Kreczanik, 2004] Thomas Kreczanik, Vers une rationalisation de l'indexation des ressources pédagogiques électroniques , [en ligne] Sylvie Lainé-Cruzel (dir), juin 2004, mémoire de DEA, sciences de l'information et de la communication Université Jean Moulin Lyon 3, . [http://memic.ccsd.cnrs.fr/documents/archives0/00/00/01/63/mem_00000163_01/mem_00000163.pdf] [consulté le 10 juin 2007]

* [Noel, 2007] Elisabeth Noël, Quelle évaluation des ressources pédagogiques? , [en ligne] Journée d'étude Urfist, "Evaluation et validation de l'information sur Internet", 31 janvier 2007, Paris, [<http://urfistreseau.files.wordpress.com/2007/02/enobel31janvier2007.pdf>] [consulté le 22 avril 2007]

Sur le LOM et les métadonnées

* [Afnor, 2006] NF Z76-040, Norme homologuée en décembre 2006, Technologies de l'information pour l'éducation, la formation et l'apprentissage, Profil d'application du LOM (LOMFR) , Métadonnées pour l'enseignement, Afnor, 2006. ICS: 03.100.30.35.240.99

* [Arnaud, 2004] Michel Arnaud, La gestion des ressources avec les métadonnées,

[en ligne] Journées "Normes et standards éducatifs, GIP de l'Académie de Lyon, CAFOC, GIP de l'Académie de Grenoble, CAFOC et l'Espace Formateurs, 26 mars 2004, Lyon.
[<http://www.foademplois.org/La%20gestion%20des%20ressources%20avec%20les%20m%E9tadonn%E9es.doc>||url:] [consulté le 8 août 2007]

* [Gomez de Regil, 2004] Rosa Maria Gomez de Regil, Présentation des standards : (LOM) - Learning Object Metadata. [en ligne] Babel - edit, L'indexation des ressources pédagogiques numériques (journée d'étude du 16/11/ 2004). ENSSIB - janvier 2006 [en ligne] [consulté le 30 mars 2007]

* [Fressard, 2006] Olivier Fressard, Métadonnées et valorisation de l'information, BBF 2006, Paris, t.51, n°4. [en ligne]
[<http://bbf.enssib.fr/sdx/BBF/frontoffice/2006/04/document.xsp?id=bbf-2006-04-0094-012/2006/04/fam-tourhorizon/tourhorizon&nDoc=33&statutMaitre=non&statutFils=non&tri=||url:>][consulté le 28 mars 2007]

Sur l'interopérabilité

* [Foulonneau, 2005] Muriel Foulonneau, Assurer l'interopérabilité des systèmes documentaires, in Les archives ouvertes: enjeux et pratiques, guide à l'usage des professionnels de l'information, coord. Christine Aubry et Joanna Janik, chapitre 5, 2005, ADBS, Paris, p.163-200.

* [Mai Chan, 2002] Lois Mai Chan, Marcia Lei Zeng, La réalisation de l'interopérabilité entre vocabulaires d'accès matière et systèmes d'organisation de la connaissance: une analyse méthodologique, 68th IFLA council and general conference, August 18-24,2002.[<http://www.ifla.org/IV/ifla68/papers/008-122f.pdf>||url:] [consulté le 15 juillet 2007]

* [Maniez, 1997] Jacques Maniez, Fusion des banques de données documentaires et

compatibilité des des langages d'indexation, Documentaliste-sciences de l'information, 1997, vol.34, n°4-5, p.212-224.

* [Degez, 1998] Danièle Dégez, avec des contributions de Claudine Masse, Florence Rénier et Anne-Marie Sardon, Compatibilité des langages d'indexation: mariage, cohabitation ou fusion? quelques exemples concrets, Documentaliste- sciences de l'information, 1998, vol.35, n°1, p.3-14.

* [Kreczanik, 2006] Thomas Kreczanik, Modélisation de parcours dans des hypertextes pédagogiques: typage des ressources et des liens, [en ligne] Colloque International: Discours et Document, Schedae, 2006, prépublication n°15, fascicule n°1, p.123-128.
[<http://www.unicaen.fr/services/puc/ecrire/preprints/preprint0152006.pdf>][url:]
[consulté le 2 juillet 2007]

* [Hudon, 2005] Michèle Hudon, Accès thématique en bibliothèque numérique : le rôle du langage documentaire de type « thésaurus » , in Les bibliothèques numériques, Fabrice Papy dir., Paris, Lavoisier, 2005, p. 151-177.

Les langages documentaires

* [Degez, Menillet, 2001] Danièle Dègez, Dominique Ménillet, Thésauroglossaire des langages documentaires, Un outil de contrôle sémantique, ADBS edition, Paris, 2001.

* [Francis, Quesnel, 2007] Elie Francis et Odile Quesnel, Indexation collaborative et folksonomies, Langages documentaires et outils linguistiques, numéro spécial Documentalistes-Sciences de l'information, février 2007, vol. 44 , n°1, p.58-63 .

* [Lefèvre, 2001] Philippe Lefèvre, La recherche d'information, du texte intégral au thésaurus, Hermès Science Publication, Paris, 2001.

* [Maniez, 1994] Jacques Maniez, Actualité des langages documentaires, les fondements théoriques de la recherche d'information, ADBS Edition, Paris, 1994.

* [Menon, 2007] Bruno Menon, Les langages documentaires : un panorama, quelques remarques critiques et un essai de bilan, l'oeil de l'ADBS mars 2007.

* [Menon, 2007] Bruno Menon, Les langages documentaires : un panorama, quelques remarques critiques et un essai de bilan, Langages documentaires et outils linguistiques, numéro spécial Documentalistes-Sciences de l'information, février 2007, vol. 44 , n°1, p.18-28.

* [Aitchison, Gilchrist, 1992] Jean Aitchison, Alan Gilchrist, Construire un thésaurus, Manuel pratique, (traduction Dominique Hervieu, révision scientifique Jacques Maniez,) ADBS, Paris, 1992.

* [Hudon, 1994] Michèle Hudon, Le thésaurus: conception, élaboration, gestion, Coll.Clé en main, Asted, Montreal, 1994.

Les ontologies

* [Chaumier, 2007] Jacques Chaumier, Les ontologies, antécédents, aspects techniques et limites, Langages documentaires et outils linguistiques, numéro spécial Documentaliste Sciences de l'information, février 2007, vol. 44 , n°1, p.81-83.

* [Gandon, 2006] Fabien Gandon , Ontologies informatiques, [en ligne] Interstices, mai 2006. url: http://interstices.info/display.jsp?id=c_17672 [consulté le 11 avril 2007]

* [Ghouas Dziri, 2006] Anissa Ghouas Dziri, Ontologies et thésaurus : des systèmes d'organisation des connaissances différents et complémentaires, in Indice, index, indexation, colloque 3 -4 novembre 2005, Université Lille-3 par les laboratoires CERSATES et GERICO, ADBS, Paris, 2006.

7 Annexes

Exemple de Fiche LOM de Culture Science physique remplie par Catherine Simand

```
<lom xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://ltsc.ieee.org/xsd/LOM"
xsi:schemaLocation="http://ltsc.ieee.org/xsd/LOM
http://ltsc.ieee.org/xsd/lomv1.0/lom.xsd">
  <metaMetadata>
    <contribute>
      <role>
        <value>creator</value>
      </role>
      <!--creator/validator -->
      <entity>Catherine Simand</entity>
      <date>
        <dateTime>2007-06-26</dateTime>
      </date>
    </contribute>
    <language>FR</language>
  </metaMetadata>
  <general>
    <title>
      <string>Diffraction à l'infini</string>
    </title>
    <identifiant>
      <catalog>ENS-Lyon CSPhysique</catalog>
      <entry>??</entry>
    </identifiant>
    <language>fr</language>
    <description>
      <string language="fr">Saurez-vous associer à chaque figure de
diffraction l'ouverture diffractante correspondante ?</string>
    </description>
    <keyword>
      <string language="fr">diffraction</string>
    </keyword>
    <keyword>
      <string language="fr">écran diffractant</string>
    </keyword>
    <keyword>
      <string language="fr">lumière</string>
    </keyword>
    <keyword>
      <string language="fr">onde</string>
    </keyword>
  </general>
</lom>
```

```

</keyword>
<keyword>
  <string language="fr">diffraction de Fraunhofer</string>
</keyword>
<keyword>
  <string language="fr">diffraction à l'infini</string>
</keyword>
<keyword>
  <string language="fr">diffraction à grande distance</string>
</keyword>
<keyword>
  <string language="fr">figure de diffraction</string>
</keyword>
<keyword>
  <string language="fr">transformée de Fourier</string>
</keyword>
<keyword>
  <string language="fr">transformée de Fourier à deux
dimensions</string>
</keyword>
<keyword>
  <string language="fr">transformée de Fourier 2D</string>
</keyword>
<keyword>
  <string language="fr">transformée de Fourier spatiale</string>
</keyword>
<keyword>
  <string language="fr">ouverture diffractante</string>
</keyword>
<keyword>
  <string language="fr">diffraction par une fente</string>
</keyword>
<keyword>
  <string language="fr">diffraction par un trou</string>
</keyword>
<keyword>
  <string language="fr">interférence</string>
</keyword>
<keyword>
  <string language="fr">interférences</string>
</keyword>
<aggregationLevel>
  <value>2</value>
</aggregationLevel>
</general>
<lifeCycle>
  <status>
    <value>final</value>
  </status>
  <!-- draft/final/revised/unavailable-->
<contribute>
  <role>
    <value>author</value>
  </role>
  <entity>Catherine Simand</entity>
  <date>

```

```

        <dateTime>2007-06-20</dateTime>
    </date>
</contribute>
<contribute>
    <role><value>publisher</value></role>
    <entity>ENS Lyon</entity>
</contribute>
</lifeCycle>
<technical>
    <size>10000</size>
    <!--en bytes -->
    <format>text/xml</format>
    <!--video/mpeg ou ...-->
    <location>xmldb:exist://ilafegh.ens-
lyon.fr/db/csphysique/data/QSfigurediffraction.xml</location>
</technical>
<educational>
    <learningResourceType>
        <value>problem statement</value>
    </learningResourceType>
    <!--lecture/simulation/figure/graph/index/slide/ta
ble/narrative
text/exam/experiment/problem statement/exercise -->
    <intendedEndUserRole>
        <value>teacher</value>
    </intendedEndUserRole>
    <intendedEndUserRole>
        <value>learner</value>
    </intendedEndUserRole>
    <!--teacher/learner/author/manager -->
    <!--ne mettre que teacher pour zone réservée
prof -->
    <typicalAgeRange><string>tout public</string></typicalAgeRange>
    <!--collège/seconde/première/terminale/université/au
tre/tout
public -->
</educational>
<rights>
    <cost>
        <value>no</value>
    </cost>
    <copyrightAndOtherRestrictions>
        <value>yes</value>
    </copyrightAndOtherRestrictions>
    <description>
        <string>Copyright ENS Lyon</string>
    </description>
</rights>
<relation>
    <kind>
    </kind>
    <!--ispartof/haspart/isversionof/isformatof/hasfo
rmat/isreferencedby/isbasedon/isbasisfor/requires/isrequiredby-->
    <resource>
        <identifier>
        <entry/>

```

```

        </identifier>
        <description>
            <string language="fr">description du fichier dont la
ressource est une
                partie</string>
            </description>
        </resource>
    </relation>
    <classification>
        <taxonPath>
            <source><string>Chebocar</string></source>
            <taxon><entry><string>Ondes</string></entry></taxon>
            <taxon><entry><string>Ondes
électromagnétiques</string></entry></taxon>
            <taxon><entry><string>Optique</string></entry></taxon>
            <taxon><entry><string>Optique
ondulatoire</string></entry></taxon>
            <taxon><entry><string>Diffraction</string></entry></taxon>
        </taxonPath>
        <taxonPath>
            <source><string>Chebocar</string></source>
            <taxon><entry><string>Ondes</string></entry></taxon>
            <taxon><entry><string>Ondes
électromagnétiques</string></entry></taxon>
            <taxon><entry><string>Optique</string></entry></taxon>
            <taxon><entry><string>Optique
ondulatoire</string></entry></taxon>
            <taxon><entry><string>Interférences</string></entry></taxon>
        </taxonPath>
        <taxonPath>
            <source><string>Chebocar</string></source>
            <taxon><entry><string>Physique à notre
échelle</string></entry></taxon>
            <taxon><entry><string>Ondes</string></entry></taxon>
            <taxon><entry><string>Ondes
électromagnétiques</string></entry></taxon>
            <taxon><entry><string>Optique</string></entry></taxon>
            <taxon><entry><string>Optique
ondulatoire</string></entry></taxon>
        </taxonPath>
    </classification>
</lom>

```

Langages documentaires et ontologies

Dans le cadre de mon stage de master 2 à l'INRP et à l'ENS sciences de Lyon je travaille à l'amélioration des taxonomies, que les concepteurs, des scientifiques de différentes disciplines, appelaient thésaurus. J'ai à l'origine cherché des informations sur les langages documentaires pour définir et donner un nom exact aux outils qui ont été élaborés. Au fil de mes recherches, je me suis intéressée aux ontologies, je propose ici une synthèse de mes lectures sur ces sujets. Il existe des ressemblances structurelles entre ontologies et langages documentaires. Mais malgré des analogies entre ces modes de représentation des connaissances, il est inexact de les assimiler. J'ai choisi de façon partielle de considérer sous l'appellation de langage documentaire, les classifications, les taxonomies et les thésaurus. Ce choix est lié à ma problématique de travail et omet volontairement un certain nombre de langages documentaires.

Classification

Une classification peut être considérée comme le langage documentaire le plus simple. Il est composé de termes ou d'objets répartis par classes ou catégories selon un certain ordre et une certaine méthode. Taxonomie. Une taxonomie (de termes, de classes, de concepts) désigne la hiérarchie ou l'arborescence autour de laquelle est construite différents types d'instruments, comme les thésaurus ou les ontologies. D'un point de vue fonctionnel, une taxonomie est un cadre d'organisation pour des ressources numériques de toute nature (et pas seulement documentaires), destiné à en permettre une présentation ordonnée. L'accès se fait par navigation hypertextuelle.

Thésaurus

Un thésaurus est un index de termes organisés selon une liste hiérarchique. Le terme - est défini comme une représentation linguistique d'un concept unique. Les

termes sont organisés en thèmes et possèdent des liens sémantiques entre eux : synonymie, terme spécifique (lien vers un concept de sens plus précis), terme général (lien vers un concept de sens plus large) et terme rejeté ou équivalent. Ces derniers sont rejetés du thésaurus lors de son élaboration mais un système de renvoi vers un « synonyme » ou un terme proche est mis en place. Ces relations sémantiques entre les termes caractérisent le thésaurus et le différencie des thésaurus. Un thésaurus a trois grandes fonctions : l'indexation des ressources, la recherche documentaire en réduisant par un vocabulaire contrôlé l'impact de la synonymie dans le langage naturel et permettre un repérage exhaustif des ressources par des liens entre les concepts et mots proches.

Ontologie

Selon le Petit Robert, l'ontologie est la partie de la métaphysique qui étudie l'être en tant qu'être, indépendamment de ses déterminations particulières. Depuis une dizaine d'années, c'est dans le domaine de l'ingénierie des connaissances que l'ontologie est apparue sous une acception informationnelle. Une ontologie est un ensemble de concepts, de propriétés, d'axiomes, de fonctions et de contraintes explicitement définies. C'est une conceptualisation qui est le résultat de la formulation exhaustive et rigoureuse d'un domaine. Le vocabulaire qui compose une ontologie est choisi de manière consensuelle et la conceptualisation de celui-ci doit pouvoir être comprise et interprétée par un ordinateur. L'ontologie organise les concepts de manière hiérarchique, élabore des relations entre concepts limités à un nombre déterminé à l'avance ainsi que des règles et axiomes qui les contraignent. Les objets sont définis d'une part et leurs relations d'une autre.

Les ontologies : concepts et représentations

Cette partie repose sur l'article de Fabien Gandon, qui a l'atout d'une certaine clarté malgré la complexité de la question. Pour exprimer un concept nous choisissons une représentation symbolique souvent linguistique, verbale ou iconique. Pour un même objet, nous pouvons donner diverses représentations linguistiques.

On pourrait parler de voiture, d'auto, de véhicule automobile, de tacot, de bagnole etc. Nous dissocions les concepts et leurs manifestations linguistiques. Un terme n'est pas un concept et vice-versa. Un terme peut être ambigu quand un concept n'a qu'un seul sens, une seule définition. Il faut alors gérer les problèmes de synonymie (un concept dénoté par plusieurs termes) et d'homonymie (un terme dénotant plusieurs concepts). Dans une ontologie, il faut distinguer les intentions et les extensions d'un concept. L'intention est un ensemble de propriétés communes aux individus auxquels le concept s'applique et permettant de définir ce concept. Par exemple : le concept que nous allons appeler pour le formaliser « C » est une sous-catégorie de véhicules de transports automobiles, aménagés pour le transport d'un petit nombre de personnes et comprenant au minimum 3 roues. L'extension est un ensemble d'entités entrant dans cette catégorie. Par exemple : la twingo de Rose, la kangoo d'Olivier, la deux-chevaux d'Emilie. Un roman est un livre, un livre est un document. Un homme et une femme sont des sous-types d'humains qui sont eux-mêmes sous-types d'être vivants.

Fabien Gandon explique que dans une ontologie les intentions sont organisées, structurées et contraintes de façon à représenter une conception du monde et de ses contraintes. Par exemple, une voiture est forcément un véhicule. L'intention permet de définir de manière non ambiguë comment concevoir de façon précise le sens. Ce qui permet sa manipulation par une machine. Les intentions sont organisées par taxonomie ou hiérarchie de types. Cette organisation a une grande importance car nous effectuons souvent et sans nous en rendre compte, ces relations hiérarchiques entre les concepts Exemple : « - Tu connais un restaurant proche ? - Il y a une pizzeria au coin de la rue. - Merci. »

La première personne a généralisé sa demande au concept de restaurant, qui est un concept suffisamment abstrait pour recouvrir un grand nombre de réponse. La deuxième personne a utilisé sa taxonomie de concept connue de tous implicitement pour en déduire qu'une pizzeria est un restaurant et que sa réponse est pertinente. Cette taxonomie partagée suppose que la réponse sera comprise sans préciser qu'une pizzeria est un restaurant et que c'est effectivement le cas. Le recours à des conceptualisations partagées est au centre de nombreux échanges d'informations. Le fait de rendre explicite ces connaissances et de s'assurer de leur nature

consensuelle est un des problèmes majeurs de l'ingénierie ontologique. Ainsi, dans un système d'information, le simple ajout de cette connaissance peut permettre d'améliorer considérablement les capacités des machines. Il s'agit d'expliquer au système une partie de notre réalité, l'ontologie permet de factoriser des connaissances et de pouvoir les réutiliser ou les échanger. Mais les connaissances ontologiques dépassent ces connaissances taxinomiques.

Selon Fabien Gandon, on peut introduire dans une ontologie : des définitions complètes « une personne est directeur seulement s'il dirige une organisation », des contraintes d'intégrité « un parent ne peut pas être plus jeune que ses enfants », des fonctions de calcul, des relations réciproques « si Thomas est marié à Stéphanie alors Stéphanie est mariée à Thomas », des connaissances par défaut « une voiture a 4 roues », Le calcul d'une fermeture transitive « si un coupé est une voiture et une voiture est un véhicule alors un coupé est un véhicule. »

Web sémantique

Le développement des ontologies va de pair avec le concept du web sémantique. Le web sémantique est une extension du web dit syntaxique. Cette nouvelle conception du web, appelé aussi web 2.0, propose d'améliorer la recherche d'information par une représentation sémantique des ressources, qui devrait permettre des contenus exploitables et interprétables par des machines. Le développement du web sémantique repose sur un certain nombre de composants comme des métadonnées, une méthode de raisonnement, un format et des systèmes d'organisation des connaissances. Ces trois derniers points étant la composante même des ontologies. Selon Jacques Chaumiez « Pour être susceptible d'être exploitées automatiquement, les métadonnées doivent être entièrement explicites et exprimées selon un vocabulaire formellement défini. Les ontologies sont le réceptacle de ces définitions. » Contrairement aux autres langages de formalisation, les ontologies n'ont pas comme seule application l'aide à la gestion de ressources documentaires. Les ontologies se retrouvent dans de nombreux

domaines d'application comme : l'intégration d'informations géographiques, la gestion de ressources humaines, l'aide à l'analyse en biologie, le commerce électronique, l'enseignement assisté par ordinateur, les bibliothèques numériques, les échanges commerciaux entre partenaires industriels, le suivi médical informatisé...

Ontologies et thésaurus

Les ontologies, de la même façon que les thésaurus, les taxonomies sont des systèmes de représentation des connaissances d'un domaine et utilisées dans des manipulations sémantiques afin de rechercher de l'information. Le tableau ci-dessous recense les ressemblances et différences des thésaurus et ontologies.

| | Thésaurus | Ontologies |
|--------------------------|--|---|
| Origine | Sciences documentaires | Philosophie puis le concept a été repris en intelligence artificielle |
| Fonctions | | |
| Indexation de ressources | manuelle | automatique |
| Recherche documentaire | <ul style="list-style-type: none"> • Permet une recherche pertinente mais se pose un problème d'incompatibilité entre les thésaurus • Possibilité de trouver des concepts signalés comme proches dans le thésaurus | <ul style="list-style-type: none"> • Permet une recherche intelligente et raisonnée par les liens sémantiques entre concepts contrôlés par des règles. • Permet le rappel de notion liées par une généralisation de la requête s'il y a trop peu de résultats ou à l'inverse un regroupement thématique des résultats s'il sont en grand nombre • Recherche sémantique sur le web par des recherches dans des sources hétérogènes, • Annotations de ressources web, |
| Utilisation | Production manuelle de description documentaire | Description des ressources automatique |
| Autres utilisations | | <ul style="list-style-type: none"> • Gestion de mémoire d'entreprise, • Permettre l'interopérabilité de ressources en réseau • Piloter des traitement de traitement automatique des langues, • Vérifier la cohérence de modèle • Permettre des raisonnement temporels et spatial • Faire des approximations logiques |
| Conception | | |
| Elaboration | Normes | Aucune norme. Fréquemment son élaboration est basée sur un thésaurus. Seulement un article de référence A.Gomez-Perez, Methontology, 1998 |
| Liens sémantiques | Liens simples <ul style="list-style-type: none"> • hiérarchiques, • associatif • d'équivalence par la gestion des synonymies et antonymie (descripteurs et non-descripteurs) | Liens complexes <ul style="list-style-type: none"> • hiérarchiques, • règles ou axiomes définissant et contrôlant les liens sémantiques complexes en fonction de l'objectif de l'ontologie |
| Langage de formalisation | Langage naturel | Langage opérationnel compréhensible par une machine |

Pour commenter ce tableau et conclure cette partie, on peut affirmer qu'un thésaurus peut être transformé en ontologie par un travail de conceptualisation et de formalisation des connaissances. On peut à ce propos citer Bruno Bonchimont « Les ontologies n'étant pas sans rapport avec les terminologies, on peut trouver dans les thésaurus des ressources pour amorcer une ontologie. Mais il faut prendre garde au fait que ce ne sont que des "ressources pour" et pas des "embryons de" . » Le même auteur précise que « les ontologies sont des artefacts construits en fonction d'une tâche précise et ne peuvent être réutilisées, en tant qu'objets formels, pour une autre tâche. » (citations tirées de l'article de Jacques Chaumier). Ces citations appuient et résument ce qui a été montré dans le tableau, c'est à dire que si thésaurus et ontologies sont souvent rapprochées, on ne peut pas les confondre.

Pour conclure, s'il y a des points de ressemblances entre les langages documentaires et les ontologies, il est inexact de les assimiler et important de souligner qu'une ontologie est un outil beaucoup plus complexe. Une ontologie a une vocation plus large. « Une ontologie n'est pas un catalogue du monde, une taxonomie, une terminologie ou une liste d'objets, de choses ou de quoi que ce soit d'autre. Une ontologie serait plutôt le cadre général, la structure au sein duquel catalogues, taxonomies, terminologies peuvent recevoir une organisation appropriée » (Roberto Politic et traduit par Bruno Menon).

Bibliographie :

* Jean Aitchison, Alan Gilchrist, Construire un thésaurus, Manuel pratique, (traduction Dominique Hervieu, révision scientifique Jacques Maniez,) ADBS, Paris, 1992. ISBN 2 901 046 43 6

* Jacques Chaumier, Les ontologies, antécédents, aspects techniques et limites, Langages documentaires et outils linguistiques, p.81-83, numéro spécial Documentalistes-Sciences de l'information, février 2007, vol. 44 , n°1.

* Renaud Eppstein, Ontologies http://artist.inist.fr/article.php3?id_article=243
[consulté le 11 avril 2007]

* Fabien Gandon , Ontologies informatiques
http://interstices.info/display.jsp?id=c_17672[consulté le 11 avril 2007] (une partie de
cet article est largement citée, la plupart des exemples de la partie portant sur les
ontologies en sont tirés)

* Anissa Ghouas Dziri, Ontologies et thésaurus : des systèmes d'organisation des
connaissances différents et complémentaires, Indice, index, indexation, actes du
colloque organisé les 3 et 4 novembre 2005 à l'université Lille-3 par les laboratoires
CERSATES et GERICO, ADBS, Paris, 2006. ISBN 2 84365 088 7

* Michèle Hudon, Le thésaurus: conception, élaboration, gestion, Coll.Clé en main,
Asted, Montreal, 1994. ISBN 2 921 548 14 3

* Bruno Menon, Les langages documentaires : un panorama, quelques remarques
critiques et un essai de bilan, l'il de l'ADBS mars 2007
[http://www.adbs.fr/site/publications/oeil_adbs/17/oeil_adbs_no17.html#metiers_mot_cle||brunomenon]
[consulté le 11 avril 2007]

* Bruno Menon, Les langages documentaires : un panorama, quelques remarques
critiques et un essai de bilan, p. 18-28, Langages documentaires et outils
linguistiques, numéro spécial Documentalistes-Sciences de l'information, février
2007, vol. 44 , n°1.

* Chantal Reynaud, Brigitte Safar, Ontologies, fondements du Web sémantique
[[http://master.lri.fr/recherche/intranet/current/cours/affiche_cours.php?wsl_dir_project=
/users/master/master/WWW/recherche/intranet/current&cours=53||Reynayd, Safar](http://master.lri.fr/recherche/intranet/current/cours/affiche_cours.php?wsl_dir_project=/users/master/master/WWW/recherche/intranet/current&cours=53||Reynayd, Safar)]
[consulté le 11 avril 2007]

* Robert Laurini, Ontologies et Interopérabilité, janvier 2001, <http://lisi.insa-lyon.fr/~laurini/disic/dea3/sld001.htm> [consulté le 11 avril 2007]

* En ce qui concerne l'indexation collaborative :

* Elie Francis et Odile Quesnel, Indexation collaborative et folksonomies, p.58-63 ,
Langages documentaires et outils linguistiques, numéro spécial Documentalistes-
Sciences de l'information, février 2007, vol. 44 , n°1.

Tableau de correspondance

J'ai décidé de mettre en annexe seulement une partie d'un des tableaux de correspondances, leurs versions intégrales sont accessibles au format pdf sur le site de Pr@tic. [\[http://pratic.ens-lyon.fr/projets/meta-LOMCDM/thesaurus/thesaurus/projection-dewey\]](http://pratic.ens-lyon.fr/projets/meta-LOMCDM/thesaurus/thesaurus/projection-dewey) en effet les tableaux sont assez volumineux l'extrait du JaHe ci dessous ne comprend que la première page sur soixante et une.

Extrait de la version 1.0 du tableau de correspondance entre le JaHe (taxonomie en sciences de la vie) et la classification en CDD, et les thésaurus ETB et Mot bis.

Tableau de Correspondances

JaHe =Sciences de la vie=4

CDD = Classification décimale de Dewey

| | |
|--|----------------------------|
| | 1er niveau de la taxonomie |
| | indice CDD |
| | ETB/TEE |
| | Mot Bis |

| | code classification | Cotes Dewey | intitulé CDD | CDD anglais | ETB | Mot Bis |
|-----------------------|---------------------|-------------|---|--|-------------------------|----------------------|
| Sciences de la vie | | 570 | Sciences de la vie. Biologie | Life science. Biology | Sciences de la vie | Sciences de la vie |
| Environnement | 4:1 | 577 | Écologie | Ecology | Environnement | Environnement |
| Environnement | 4:1 | 333.709 | Géographie environnementale | Environmental geography | | |
| Environnement naturel | 4:1:1 | 577 | Écologie | Ecology | Environnement | Environnement |
| Hydrosphère | 4:1:1:1 | 551.4 | Géomorphologie et hydrosphère | Geomorphology and hydrosphere | Océan | Écosystème aquatique |
| Hydrosphère | 4:1:1:1 | 551.462 | Circulation océanique | Ocean circulation | Mer | |
| Hydrosphère | 4:1:1:1 | 551.48 | Hydrologie | Hydrology | Cours d'eau | |
| Hydrosphère | 4:1:1:1 | 577.14 | Chimie environnementale | Environmental chemistry | | |
| Atmosphère | 4:1:1:2 | 551.51 | Composition de l'atmosphère | Composition | Phénomène atmosphérique | Atmosphère terrestre |
| Atmosphère | 4:1:1:2 | 551.6 | Climatologie | Climatology and weather | Météorologie | |
| Atmosphère | 4:1:1:2 | 577.14 | Chimie environnementale | Environmental chemistry | | |
| Lumière | 4:1:1:3 | 535 | Lumière visible (optique) et phénomènes de l'infrarouge et de l'ultraviolet | Light and infrared and ultraviolet phenomena | Lumière | Lumière |
| Sol | 4:1:1:4 | 577.57 | Écologie des sols | Soil ecology | Sciences de la terre | Pédologie |
| Sol | 4:1:1:4 | 631.4 | Pédologie, sciences de préservations des sols | Soil science | | |