Mais quelles opérations peut-on effectuer avec les données d'autorité ? Si à l'origine, le contrôle d'autorité est né d'un objectif – celui d'identifier et de contrôler la forme des points d'accès aux notices bibliographiques d'un catalogue – le souci permanent d'identifier personnes, collectivités, œuvres, etc. a conduit à un enrichissement progressif des fichiers d'autorité avec des informations supplémentaires. Ces fichiers sont devenus, ainsi, de véritables réservoirs d'informations aptes à être réutilisés pour de multiples applications, même en dehors du contexte des catalogues des bibliothèques. FRAD prend note de ce constat et définit les tâches que les différents utilisateurs, bibliothécaires ou autres chercheurs d'information, peuvent effectuer en utilisant les données d'autorité. Ces tâches sont : ***Trouver, Identifier, Contextualiser*** et ***Justifier***. Le modèle s'efforce également de répondre à la question : quelles relations et quels attributs sont nécessaires pour effectuer telle ou telle tâche ? Par exemple, la précision du type de relation entre deux personnes, tel que la relation de ***collaboration*** ou la relation ***parent/enfant,*** sert à effectuer la tâche contextualisation, à savoir à clarifier le contexte de vie de ces personnes ainsi que leurs liens avec les contenus et documents catalogués.

Enfin quel est l'impact de FRAD ? Les évolutions en cours dans le domaine du traitement bibliographique n'ont pas attendu la publication du modèle pour en intégrer ses concepts. Il est déjà devenu la référence en matière de données d'autorité dans l'univers bibliographique. Ainsi, les nouveaux Principes internationaux de catalogage[4] s'appuient sur FRAD pour tout ce qui concerne les données d'autorité. De même les RDA[5] en cours d'élaboration se veulent respecter FRAD.

Et ensuite ? Notons que pour compléter l'univers « FR.. » un autre modèle est en cours d'élaboration par un groupe de travail de l'IFLA, le groupe FRSAR (Functional Requirements for Subject Authority Records)[6]. Vous l'avez compris, il s'agit de la modélisation des données d'autorité relatives à l'indexation matière et de l'indexation matière elle-même.

Rendez-vous donc sur le site de l'IFLA, dans un premier temps, pour l'accès au texte final du modèle FRAD et, espérons le bientôt, pour des nouvelles sur le modèle FRSAR.

A. Angjeli
anila.angjeli@bnf.fr

BNF
Direction des services et des réseaux
Département Information bibliographique et numérique (IBN)
Anila Angjeli
☎ 01 53 79 53 95 85 86

1 FRANAR (Working group on Functional Requirements and Numbering of Authority Records) est le nom du groupe de travail de l'IFLA < http://www.ifla.org/VII/d4/wg-franar.htm > ; FRAD est le nom du modèle conceptuel élaboré par ce groupe (IFLA : International Federation of Library Associations and Institutions).

2 A review of the feasibility of an International Standard Authority Data Number (ISADN) / prepared for the IFLA Working Group on FRANAR By Barbara B. Tillett ; Edited by Glenn E. Patton, 1 July 2008. <http://www.ifla.org/VII/d4/franar-numbering-paper.pdf>

3 En attendant la publication de la version finale du modèle FRAD, c'est la version soumise à l'enquête internationale en 2007 qui est accessible sur le site de l'IFLA <http://www.ifla.org/VII/d4/franar-conceptual-model-2ndreview.pdf>.

4 La dernière version de la déclaration des *Principes internationaux de catalogage*, soumise à l'enquête internationale est accessible en ligne <http://www.ifla.org/VII/s13/icc/imeicc-statement_of_principles-2008_french.pdf> Le texte final a été validé par l'IFLA et est en cours de publication.

5 Resource Description and Access <http://www.collectionscanada.gc.ca/jsc/rda.html>

6 FRSAR : groupe de travail de l'IFLA <http://www.ifla.org/VII/s29/wgfrsar.h

TermSciences

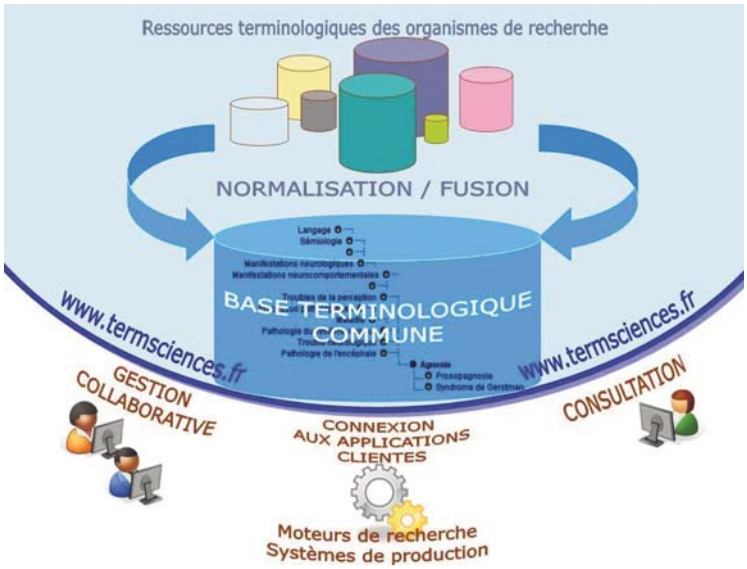
Contexte du projet

L'Institut de l'information scientifique et technique (INIST) a développé le portail TermSciences [www.termsciences.fr] en partant de deux constats :
1 la production terminologique des différents acteurs de la recherche est assez riche et diverse mais insuffisamment valorisée ;
2 la gestion des ressources terminologiques étant une tâche fastidieuse, une mise en commun de données et la fédération de compétences permettraient de mieux l'appréhender.
La réalisation du portail terminologique TermSciences vise donc à constituer un référentiel terminologique commun qui permet d'établir des passerelles entre les différents termes et variantes utilisés pour désigner une même notion tout en s'inscrivant dans une démarche de normalisation des données, d'interopérabilité des systèmes et de collaboration entre les spécialistes. Outre la valorisation et la mise en commun de la production terminologique, cet outil offre aux partenaires du projet la possibilité de gérer de manière collaborative le contenu ainsi produit et d'utiliser ce contenu pour l'indexation documentaire et/ou la recherche d'information dans les bases de données spécialisées ou sur le web.

Base de données terminologique

La base terminologique du portail TermSciences résulte de la fusion de plusieurs ressources terminologiques produites par différents organismes de recherche :
- les vocabulaires d'indexation de l'INIST-CNRS ;
- le thésaurus MeSH bilingue de la National Library of Medicine, traduit par l'Inserm avec la participation de l'INIST ;
- le thésaurus de la banque de données en santé publique (BDSP) ;
- des dictionnaires et lexiques produits par l'INRA ;
- le thésaurus de linguistique (thésaulangue) de l'ATILF-CNRS (Analyse et traitement informatique de la langue française) ;

Un portail pour valoriser la production terminologique des organismes publics de recherche



- le thésaurus GéoEthno de l'UMR 7186 (CNRS) ;
- le lexique hydrologique pour l'ingénieur du Cemagref ;
- le thésaurus myobase de l'Association française contre les myopathies (AFM) ;
- d'autres ressources (thésaurus Popin de l'INED, thésaurus de météorologie de Météo France, etc.) qui seront intégrées prochainement.

Le contenu terminologique est formalisé selon la norme ISO 16642 (Terminological Mark-up Framework) définie dans le cadre du comité technique 37 de l'ISO et publiée en 2003. Cette norme correspond à un modèle abstrait de représentation de terminologies multilingues informatisées en XML. Elle repose sur une méthodologie qui distingue la structure générale d'une base terminologique des informations (catégories de donnée, ISO 12620) qui servent à décrire les différents niveaux de cette structure.

Concrètement, la base terminologique est une collection d'entrées terminologiques correspondant chacune à un concept ; chaque concept étant décrit par un ou plusieurs termes regroupés par langue. Actuellement, la base regroupe près de 180 000 concepts décrits par près de 600 000 termes essentiellement en français et en anglais (autres langues présentes : espagnol, allemand, arabe, roumain, polonais, italien).

L'approche est onomasiologique c'est-à-dire qu'elle part du sens (le concept) pour aller vers les signes (les termes) qui le désignent. Dans la mesure où les ressources versées dans la base terminologique étaient majoritairement des langages documentaires (vocabulaires d'indexation, thésaurus) centrés sur le terme (le descripteur), une étape de conceptualisation a été nécessaire pour passer à une organisation centrée sur le concept. Par exemple, le traitement des renvois de type « employer » (EM) très fréquents dans les langages documentaires où ils sont utilisés, aussi bien pour renvoyer des vrais synonymes sur une forme préférentielle (par exemple, « nom de montagne » EM « oronyme »), que pour faire des regroupements de termes considérés comme des synonymes documentaires (par exemple, « beurre », « yaourt », « fromage » EM « produit laitier ») s'est révélé très fastidieux.

Le versement d'une nouvelle ressource terminologique dans la base commune se fait par comparaison (mapping) avec le contenu existant.

La fusion des données est faite en conservant les origines et les attributs de chaque donnée – nom de l'organisme, nom de la ressource, terme préférentiel, etc.

Gestion du contenu terminologique

Un système de gestion des contenus est accessible sur le web pour les membres. Cette interface permet la mise à jour collégiale et à distance de la base terminologique.

La base terminologique évolue donc de manière continue au rythme des enrichissements et des modifications apportées au contenu terminologique actuel mais également au rythme de l'intégration de nouvelles ressources terminologiques (vocabulaires, thésaurus, dictionnaires, etc.) dans la base.

Utilisation pour la recherche d'information

Au sein du portail, un couplage avec des sources d'information via un système de formulation de requêtes est proposé à l'utilisateur pour interroger des bases de données bibliographiques ou d'autres sources d'information du web.

Ce système permet à l'utilisateur de composer ses requêtes de manière automatique ou assistée en utilisant tous les termes (synonymes, variantes, traductions) présents dans l'entrée terminologique pour composer une requête dans le langage d'interrogation de la base cible.

Des web services de récupération de contenus sont également proposés.

Utilisation pour l'indexation

Une application utilisant la base terminologique pour l'indexation des contenus est en préparation. Elle consiste en une plateforme d'indexation sociale dédiée à la collecte, à l'annotation et au partage de références bibliographiques ou de documents en général. L'annotation des documents est mixte c'est-à-dire qu'elle utilise soit des « tags » libres soit des concepts contenus dans la base terminologique.

L'utilisation de la base terminologique de TermSciences pour l'indexation sociale permet d'apporter :

- la normalisation des termes utilisés et la gestion des renvois entres formes synonymes ;
- la traduction des termes utilisés par l'indexeur grâce au multilinguisme de la base terminologique ;
- la structuration de ces termes grâce à leurs réseaux sémantiques (génériques et associés) contenus dans la base terminologique.

Majid Khayari
majid.khayari@inist.fr

INIST - TermSciences
www.termsciences.fr
Majid Khayari ☎ 03 83 50 46 00 47 33
Stéphane Schneider
Stephane.SCHNEIDER@inist.fr
Nicolas Thouvenin
Nicolas.THOUVENIN@inist.fr
2 allée du Parc-de-Brabois CS 10310
54519 VANDŒUVRE-LÈS-NANCY