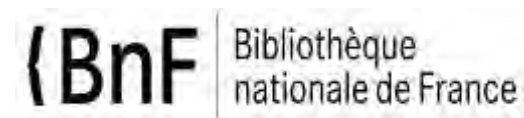


Les schémas de concepts et le Web sémantique : la norme sur les thésaurus ISO 25964 et le Web sémantique

Hélène Rabault - chef de projet chez Lexis-Nexis – helene.rabault@lexisnexis.fr
Hélène Zysman – consultante ECM chez Smile – helene.zysman@smile.fr



Sommaire

- **1 - Le contexte**
- **2 - Éléments constitutifs**
- **3 - Principes d'interopérabilité**
- **4 - SKOS et la norme 25964**
- **5 - Conclusion**

Le Contexte : Les acteurs

- **Instances normatives**

- l'ISO **TC46/SC9/WG8** : Information et documentation / "Identification and description" / "Structured vocabularies"
- Equivalent Afnor : **CG46/CN357/GE10** : INFORMATION ET COMMUNICATION / Modélisation, production et accès aux documents / Vocabulaires contrôlés

- **Composition du groupe international**

- Canada (2 personnes), Australie (1), France (2), Allemagne (3), Danemark (2), Royaume Uni (4), États-Unis (1), Espagne (1) - Auditeur hors WG : Chine (1)
- Chef de projet : Stella Dextre Clarke (UK)

- **Composition du groupe France**

- Représentant France : Sylvie Dalbin
- Dominique Chichereau ; Odile Contat ; Sylvie Dalbin ; Danièle Dégez ; Alina Deniau ; François Feyler ; Thierry Guillotin ; Dominique Ménillet ; Michel Mingam ; Hélène Rabault ; Bernard Vatant ; Virginie Triboulin ; Nathalie Yakovleff, Hélène Zysman

Le Contexte : positionnement de la norme

- La norme est centrée sur le thesaurus
 - Titre : Thesauri and interoperability with other vocabularies
 - Objet : Gestion électronique de thesaurus pour retrouver tout type de ressource
- La norme ISO 25964 est en cours d'élaboration
 - Partie 1 : thesaurus pour la recherche d'information :
 - Stade FDIS – publication en juillet 2011
 - Partie 2 : interopérabilité avec des vocabulaires contrôlés
 - Stade CD – publication en 2012
- Elle remplace les normes
 - ISO 2788:1986 (monolingue) ; ISO 5964:1985 (multilingue)
 - NF Z47-100:1981 (monolingue) ; NF Z47-101:1990 (multilingue)

Le thesaurus : définition générale

Thesaurus

controlled and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms



Recentrage sur le concept

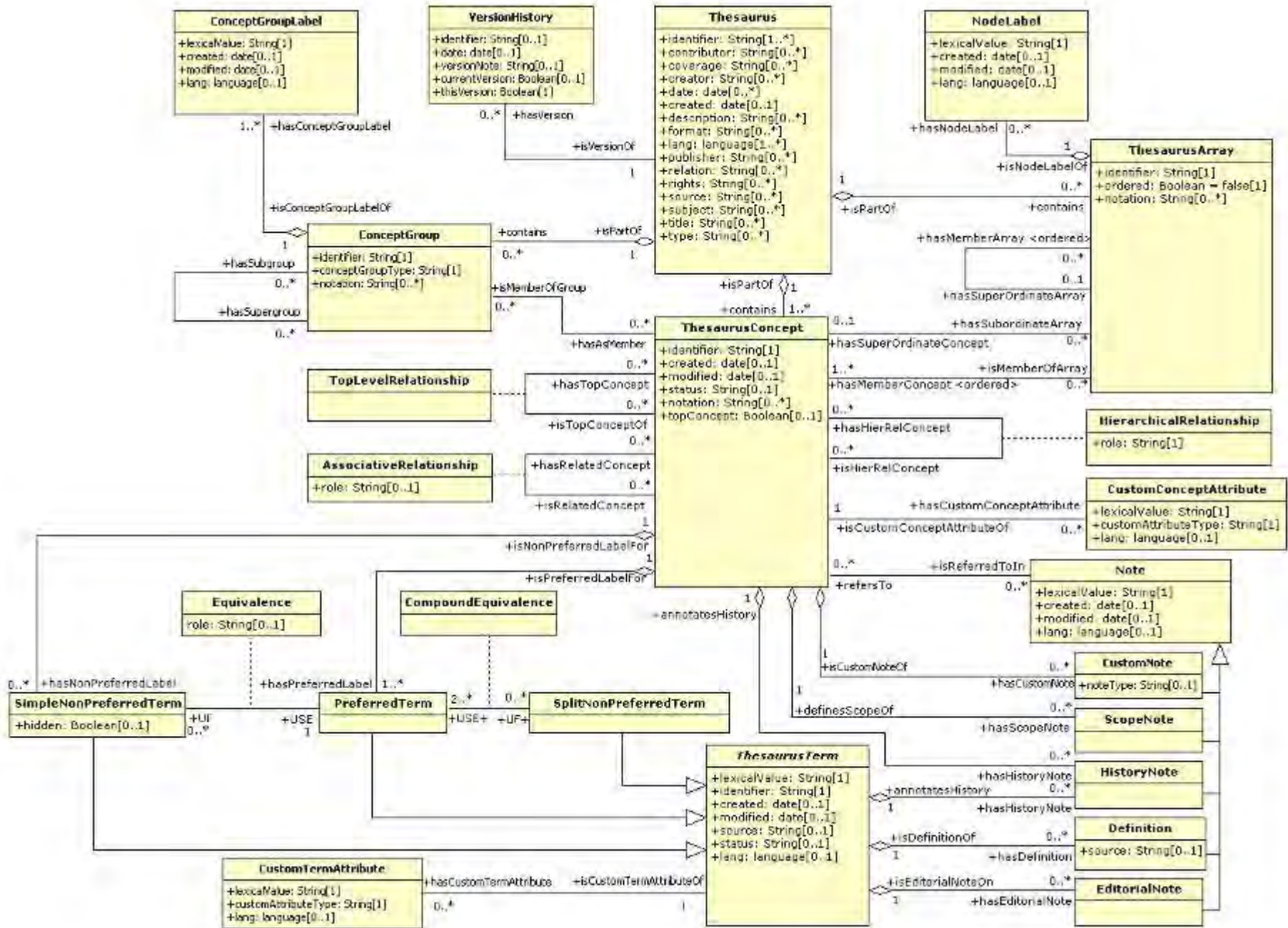
Périmètre des concepts

- **Objets** et leurs caractéristiques physiques, matériau, processus et activités, propriétés, disciplines et domaines de la connaissance, unités de mesure, types de personnes et d'organisation, organismes vivants ...
- **Entités nommées**
 - Lieux géographiques
 - Personnes physiques ou morales
 - Marques, objets spécifiques
 - ...

Un effort de formalisation et sémantisation

- **Modèle de données métier sous forme d'un schéma de classes UML**
 - Un ensemble de classes d'objets (concepts, termes, notes, ...)
 - Un ensemble de propriétés pour chaque classe
 - Identifiant, Libellé, Langue, Dates, ...
 - Ajouts possibles (customAttribute)
 - Un ensemble de classes de relations
 - Equivalence, hiérarchique, associative, top level
 - Attribut "Rôle" pour préciser la nature de la relation (cause / effet, relation hiérarchique partitive, ...)

2. Le modèle métier



Le multilinguisme

- Chaque terme est associé à une langue
- Les termes exprimant un même concept sont en relation d'équivalence cross-lingue, pouvant être explicite
 - aide économique
 - en** economic aid
 - es** ayuda económica
- On peut envisager des niveaux de langues par ex. comme autant de langues différentes

Les degrés d'équivalence crosslingue

- **Equivalence exacte**
 - en: sun / es: sol / fr: soleil / de: Sonne
 - en: rail network / fr: réseau ferroviaire / de: Schienennetz / nl: spoorwegnet
- **Equivalence inexacte ou proche**
 - en: information retrieval / fr: recherche d'information
 - fr: jeu didactique / en: learning games
- **Equivalence partielle**
 - équivalence générique/spécifique
 - de: Wissenschaft / en: science
 - SN: Comprend les sciences naturelles et sociales, et la connaissance (les savoirs ?) en général.
 - équivalence "1-à-n", "une-à-plusieurs"
 - fr: sécurité / de: Sicherheit / en: safety, security
- **Non-équivalence**

Interopérabilité : mapping des vocabulaires

- Mapping direct

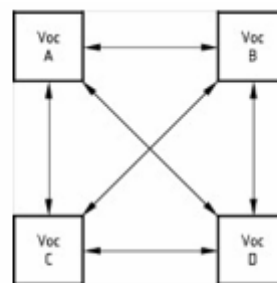


Figure 1 — Model 2 (direct-linked) as applied to four vocabularies

- Mapping en hub

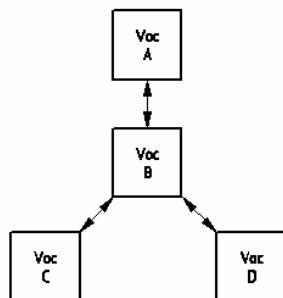


Figure 2 — Model 3 (hub model) as applied to four vocabularies

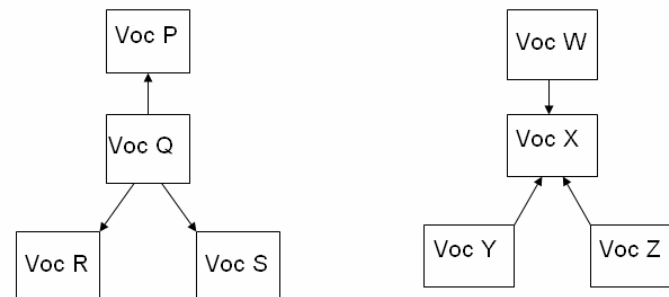


Figure 3 — Two variations on Model 3, with one-way mappings

Interopérabilité : mapping des concepts

EQ	Equivalence (sans préciser la nature)	Eau EQ Water Commande EQ Control
=	Equivalence exacte	Eau =EQ Water Téléphone mobile =EQ téléphone cellulaire Eau =EQ Eau
~	Equivalence inexacte (sans préciser la nature)	Commande ~ EQ Control Horticulture ~ EQ Jardinage Education secondaire ~EQ Education + Ecoles secondaires
BM	Equivalence vers un concept plus vaste	Roi BM monarque
NM	Equivalence vers un concept moins vaste	monarque NM rois
RM	Autre équivalence	Education RM Puniton

Interopérabilité

- Entre 2 ou plusieurs thesaurus
- Entre un thesaurus et d'autres vocabulaires
 - Classification,
 - Référentiel de classements Records Management
 - Taxonomie
 - Vedettes matière
 - Ontologie (correspondance avec OWL)
 - Terminologie (correspondance TBX / norme thesaurus)
 - Liste d'autorité
 - Anneau de synonymes

Implémentation de la norme

- **Eléments fournis**

- Schéma UML
- Schéma XML simple, pour l'échange de thesaurus fourni à titre indicatif <http://www.niso.org/schemas/iso25964/>

- **Formats d'échange évoqués**

- Sont brièvement cités quelques formats (MARC, SKOS, Zthes,...)

- **Protocoles de recherche**

- Objectif : présenter et utiliser le thesaurus
- Utilisation de protocoles existants ou développement spécifique
- Citation de protocoles spécifiques aux thésaurus (SKOS API, ADL) ou orientés Web (Search Web Services, SPARQL, Z3950)

SKOS et la norme ISO 25964

- Format et spécifications fonctionnelles permettant l'exploitation, sur le Web, de systèmes d'organisation de concepts (thésaurus, classification,...)
- Publié le 18 août 2009, s'est appuyé sur la norme BSI 8723, qui est également le point de départ de la norme ISO 25964
- SKOS est en grande partie compatible avec la norme
- Quelques différences
 - Les données de gestion ne sont pas intégrées à SKOS
 - Pas de gestion des équivalents composés, ni de relations entre termes
 - SKOS:ConceptScheme
 - Propriétés de mapping assimilées à des relations sémantiques

<http://www.w3.org/TR/2009/REC-skos-reference-20090818/>

Conclusion

- **Une norme qui positionne clairement le thesaurus par rapport à d'autres types de vocabulaires contrôlés.**
 - **Une norme adaptée aux évolutions des systèmes d'information**
 - **Une norme en phase avec le web sémantique et ses standards**
 - Typologie plus fine des relations
 - Gestion d'attributs, qui peuvent être personnalisés
 - Plus grand formalisme dans la description des objets métiers
 - Les concepts deviennent des ressources utilisables dans une optique de base de connaissance et de web sémantique
- **Un renouveau du thesaurus ?**

Références

Une nouvelle norme pour le thésaurus, Hélène Zysman, Hélène Rabault, 5-à-7 ADBS, février 2011

<http://www.slideshare.net/2helene/adbs-5-7-une-nouvelle-norme-pour-le-thesaurus>

Thésaurus pour la recherche documentaire. Présentation aux éditeurs, Hélène Zysman, Sylvie Dalbin, mars 2011

<http://www.slideshare.net/Dalb/iso-25964-thsaurus-pour-la-recherche-documentaire>

Sur le site de l'ISO

ISO FDIS 25964-1 (<http://tinyurl.com/iso25964-1>) et ISO CD 25964-2 (<http://tinyurl.com/iso25964-2>)