



**Des services numériques pour une nouvelle bibliothèque numérique – des nouvelles façons de présenter les catalogues de bibliothèque**

**Asgeir Rekkavik**

Oslo Public Library, Deichmanske bibliotek  
Oslo, Norway  
E-mail: asgeir.rekkavik[at]kul.oslo.kommune.no

**Kim Tallerås**

Oslo and Akershus University College of Applied Sciences  
Oslo, Norway  
E-mail: kim.tallerås[at]hioa.no

**Anne-Lena Westrum**

Oslo Public Library, Deichmanske bibliotek  
Oslo, Norway  
E-mail: anne-lena.westrum[at]kul.oslo.kommune.no

*Translated by:  
Sophie Rigny  
Library of Congress  
Washington DC, USA*

**Meeting:**

**80 — *Inspired moments in cataloguing* — Cataloguing**

**Résumé :**

*Lorsqu'un usager de bibliothèque fait une recherche par nom d'auteur dans un catalogue en ligne, il obtient en général une longue liste qui contient plusieurs éditions et traductions de tous les livres de cet auteur, triés par titre ou par date de publication. Pour essayer de mettre un peu d'ordre dans cette pagaille, le projet PODE a appliqué une méthode de transformation automatisée en FRBR à partir de l'information contenue dans les notices MARC.*

*Le projet expérimente aussi une représentation RDF des notices MARC pour démontrer comment la production complète d'un auteur peut être présentée en une liste courte et lucide d'œuvres uniques. Ce résultat nous a inspiré à élaborer une conversion complète en représentation RDF de notre catalogue de bibliothèque.*

## **Introduction**

Les plans pour un nouveau bâtiment de bibliothèque à Oslo, Norvège, et les discussions sur les services de bibliothèques touchés par ces plans ont indirectement donné à la bibliothèque publique d'Oslo l'occasion d'examiner comment ses métadonnées peuvent être utilisées dans de nouveaux contextes et comment elles peuvent contribuer à l'amélioration des services.

Au cours des dernières années, le projet PODE a expérimenté des métadonnées descriptives liées à des composites, à des modèles de référence tels que FRBR (Groupe de travail IFLA sur les Fonctionnalités Requises des Notices Bibliographiques, 1998), à des nouvelles générations de catalogues d'accès public en ligne et aux données liées. Ce travail a abouti au moins à une conclusion significative: la qualité des services dépend de la qualité des métadonnées existantes.

Le projet PODE a pris fin il y a un an, mais ses résultats sont maintenant poursuivis dans un projet de création de services publics pour les usagers de bibliothèque. Le projet se base sur la représentation RDF des notices MARC et le but est de créer une station interactive qui présente le catalogue de bibliothèque de façon nouvelle. Le projet connectera le catalogue de la bibliothèque avec des sources de données externes et le présentera d'une manière intuitive et visuelle grâce à une interface à écran tactile.

## **Le projet PODE**

Le projet PODE a travaillé avec de nouvelles façons d'utiliser les métadonnées présentes dans les catalogues de bibliothèques, en particulier avec des façons de mélanger le contenu du catalogue avec d'autres ressources «composites». Le projet a également examiné les possibilités et les défis que représentent les technologies web en relation avec les systèmes et les pratiques d'aujourd'hui.

Le projet a utilisé des protocoles traditionnels fournis par les vendeurs de systèmes intégrés de gestion de bibliothèque, il a aussi converti et affiné les données pour les transformer en FRBR afin de pouvoir fournir des données liées de bibliothèque «significatives».

Le projet a également expérimenté une représentation RDF des notices MARC pour démontrer comment la production entière d'un auteur peut être présentée en une liste courte et lucide d'œuvres uniques qui puisse ensuite être cherchée par ses différentes expressions et manifestations. De plus, en liant des instances dans le fichier de données aux fichiers de données externes correspondants, la présentation a été enrichie par des informations supplémentaires sur les auteurs et les œuvres.

Notre deuxième constatation cruciale a été la suivante: une représentation RDF de notre catalogue de bibliothèque nous donne un fichier de données plus pratique qui nous permet de créer beaucoup plus de services pour les usagers que les fichiers de données MARC traditionnels.

## **Transformation en FRBR – s’orienter à travers les listes de résultats des bibliothèques**

Knut Hamsun (1859-1952) est le romancier le plus important de Norvège et l’un des trois prix Nobel norvégiens de littérature. Sa production littéraire comprend environ 30 romans, quelques pièces de théâtre et plusieurs collections de nouvelles, une collection de poésie et quelques écrits documentaires et autobiographiques. La totalité de la production de Hamsun compte un total de 40 œuvres; c’est une bibliographie qui devrait être facilement navigable pour un usager de bibliothèque. Cependant, l’image que rencontre l’usager est bien différente. Dans le catalogue en ligne de la bibliothèque publique d’Oslo, une recherche par nom d’auteur sur «Hamsun, Knut» va produire une liste de 568 résultats (le 7 mai 2012). C’est bien sûr beaucoup trop pour un auteur qui a écrit 40 livres. Notez que ceci est le résultat d’une recherche avancée. Une recherche typique simple, comme la ferait la plupart des usagers de bibliothèque, fournit une liste encore plus longue.

Le problème est que le catalogue en ligne ne fait pas la distinction entre les différentes œuvres d’un auteur et les différentes versions d’une même œuvre. Dans notre liste de 568 résultats, il y en a 63 qui correspondent à des représentations différentes du même roman: *Hunger*. Il s’agit d’éditions variées, de formats différents et de traductions en d’autres langues. Les usagers doivent bien sûr pouvoir choisir entre le livre, le livre en bande sonore ou le film, et ils doivent pouvoir choisir la langue dans laquelle il veut lire le livre, mais pour la plupart des usagers c’est plus dérangeant qu’utile lorsque le catalogue en ligne les fait choisir entre plus de 20 différentes éditions de *Hunger* en norvégien.

### **Les normes de bibliothèque**

Le catalogue de bibliothèque s’est traditionnellement concentré sur la description d’objets physiques. Chaque manifestation d’un livre est représentée par une notice séparée, et il n’y a pas de connections fonctionnelles entre les notices qui décrivent les manifestations du même ouvrage. Un usager de bibliothèque qui fait une recherche dans le catalogue en ligne pour un titre spécifique pourra par exemple se mettre sur liste d’attente pour une édition particulière d’un roman classique sans réaliser que de nombreuses autres éditions du même ouvrage sont disponibles. Un autre usager pourra finir par ne pas obtenir le livre du tout s’il a commandé par erreur une édition dont les exemplaires ne sont plus disponibles. Pour ses expérimentations, le projet PODE a pris pour hypothèse que les usagers de bibliothèque recherchent en général un titre particulier et non pas une édition particulière de ce titre, et que cette attitude est typique pour les œuvres de fiction, où les différentes éditions ont en général un contenu identique. Dans la perspective du système de gestion de bibliothèque, l’usager devrait dans l’idéal pouvoir réserver un titre sans avoir à choisir entre les différentes éditions. Bien sûr, ceux pour qui les éditions sont importantes devraient encore avoir la possibilité de le spécifier, mais pourquoi forcer tous les autres à choisir?

Comme beaucoup de gens l’ont souligné, les normes actuelles de bibliothèque ont été développées avant le web et avant l’infrastructure actuelle de production, de distribution et d’utilisation des métadonnées. Il en est de même pour les normes qui ont introduit les données de bibliothèque à la sphère numérique qui ont été développées des années avant l’invention des modèles Entité-Relation (ER) et des bases de données relationnelles.

L'application de modèles de référence tels que FRBR (qui sépare les éditions des œuvres) présente des défis car ce modèle se base sur une analyse ER et sur des relations non mises en place dans et entre les notices MARC.

Le format MARC incarne les inscriptions techniques et la logique du catalogue sur fiches et il a été développé pour l'automatiser. Les métadonnées dans les catalogues sur fiches étaient lues et interprétées par des humains, une caractéristique qui se poursuit dans le format MARC qui prescrit des notices (séparées) consistant largement de différentes chaînes de caractères lisibles par les humains. Ceci est bien sûr une description simplifiée des métadonnées de bibliothèques. Le processus de création d'une notice MARC est caractérisé par une interaction complexe entre des règles de catalogage telles que AACR2 et ISBD, et la plupart des motifs derrière les chaînes de caractères sont à trouver dans ces règles. Les chaînes de caractères doivent être entièrement cohérentes pour que les ordinateurs puissent interpréter les données avec exactitude et créer une présentation utile de ces données.

Dans les bases de données relationnelles et dans les environnements de données liées, la meilleure doctrine en pratique est d'éviter la désambiguïsation en donnant des identifiants uniques – en utilisant respectivement des clés primaires/étrangères et des URIs. Les notices MARC manquent de tels identifiants explicites qui auraient aidé nos outils d'indexation et nos moteurs de recherche à séparer deux auteurs du même nom, ou peut-être à fusionner ces auteurs s'ils représentent vraisemblablement la même personne sur la base des relations concurrentes aux mêmes œuvres identifiées de façon unique. Des défis semblables vont bien sûr continuer lorsqu'on voudra que nos ordinateurs identifient des relations entre des documents, des auteurs et différentes entités FRBR. Quels livres représentent les manifestations d'une expression ou d'une œuvre particulières? Comment pouvons-nous identifier des œuvres telles que des nouvelles lorsqu'elles sont seulement décrites par du texte dans un champ de note MARC? Pour la plupart des livres publiés après 1970 nous avons des numéros ISBN qui identifient les manifestations, et nous avons des noms et des titres avec lesquels les ordinateurs peuvent raisonner, mais sans catalogues cohérents et sans identifiants lisibles par ordinateur, c'est tout de même une tâche ardue. Il faut au moins faire un bon nettoyage préalable.

### **Transformation des notices MARC en FRBR**

Le projet PODE a utilisé un outil développé par l'Université norvégienne de science et de technologie pour la transformation automatisée des notices MARC en FRBR (Aalberg, 2006). Cet outil utilise des transformations XSL sur des exports de catalogue dans le format MarcXchange pour trier les données à l'intérieur des notices bibliographiques sur lesquelles se basent les entités FRBR qui s'appliquent aux données.

Les notices des catalogues de bibliothèque sont principalement des descriptions de manifestations; la notice individuelle décrit une édition particulière d'une œuvre publiée avec des informations spécifiques telles que la date ou le lieu de publication, la description physique, l'ISBN, etc. Mais la notice contient aussi des informations qui ont trait à l'expression et à l'œuvre auxquelles cette manifestation est liée. Par exemple, les

données qui contiennent le nom de l'auteur et le titre original d'un livre sont des informations qui décrivent l'entité de l'œuvre, tandis que les informations sur la langue et le format relèvent de l'expression. L'outil de transformation en FRBR identifiera à quels champs s'appliquent les entités FRBR, et les utilise pour diviser chaque notice entre une partie œuvre, une partie expression et une partie manifestation, avec des relations FRBR entre elles. Si l'outil produit des descriptions d'œuvres identiques pour deux notices MARC différentes, on part du fait que les notices décrivent deux incarnations différentes de la même œuvre. L'outil produira aussi des entités des groupes 2 et 3, tels que des agents et des sujets.

Initialement, le projet a fait passer le corpus des notices sélectionnées (908 notices décrivant des manifestations de Hamsun et d'un autre auteur norvégien, Per Peterson) sans nettoyage préalable à travers le système automatisé de transformation en FRBR. La première tentative a produit des résultats qui étaient loin d'être parfaits. Ceci était dû principalement à des informations manquantes dans les notices MARC et à des pratiques de catalogage inconsistantes. Pour progresser dans cette expérimentation, le projet a dû nettoyer un nombre considérable de notices pour pouvoir obtenir des données suffisamment expressives et consistantes.

### **Le processus de nettoyage**

En bref, le processus de nettoyage a consisté essentiellement à identifier et à ajouter des titres originaux et des titres de forme aux notices où ils manquaient, à ajouter des informations sur les œuvres individuelles et les nouvelles rassemblées en un volume, et à ajouter des indicateurs pour distinguer les titres d'œuvres significatifs des titres d'œuvres non significatifs. Nous avons également passé du temps à corriger les erreurs et les fautes de frappe. En tout, nous avons passé environ 60 heures au nettoyage de la production de Knut Hamsun et de Per Petterson, y compris le temps passé à déterminer les règles de correction.

L'une des tâches principales du processus de nettoyage a été l'identification et l'amélioration des notices pour les ouvrages de traduction qui soit n'avaient pas du tout le titre original en norvégien, ou bien qui n'indiquaient le titre original que dans une note lisible par un humain. Une autre tâche a consisté à ajouter les titres de forme dans des cas où les titres des titres non traduits différaient du titre original. À cause des réformes orthographiques norvégiennes, plusieurs des œuvres de Hamsun sont sorties avec des titres différents à des époques différentes. Par exemple, la nouvelle *Paa tourné* (titre original) a aussi été publiée avec les orthographes *På tourné* et *På turné*.

Une autre tâche a été de déterminer des indicateurs pour séparer les titres significatifs des titres non significatifs. Les titres non significatifs se trouvent en général dans les collections des œuvres d'un auteur lorsque le contenu a été mis en place et la publication autorisée par une autre personne que l'auteur. De telles publications ne devraient pas être répertoriées comme œuvres dans une bibliographie FRBR, bien que les romans et nouvelles contenus dans les collections doivent l'être<sup>1</sup>.

Exemples :

a) 245 10 †aGrowth of the soil †cKnut Hamsun ; translated from the Norwegian by W. Worster

(traduction anglaise d'une œuvre originale de Hamsun. Le premier indicateur mis sur 1 veut dire que c'est un titre d'œuvre significatif)

b) 245 00 †aTales of love and loss †cKnut Hamsun ; translated by Robert Ferguson

(collection anglaise de nouvelles qui n'étaient pas publiées ensemble à l'origine. Le premier indicateur mis sur 0 veut dire que c'est un titre non significatif)

Voir l'annexe pour plus d'informations sur le processus de nettoyage.

### **Résultats**

Après que le premier essai de transformation en FRBR ait identifié 149 œuvres par Hamsun, la liste a été encore réduite à 84 après le nettoyage des notices MARC. Dans le cas de Per Petterson, nous avons constaté une augmentation du nombre des œuvres qui sont passées de 14 à 41 à cause de l'ajout des titres des collections individuelles de nouvelles et d'essais à certaines notices du catalogue. La liste finale des œuvres correspondait presque exactement aux bibliographies réelles des deux auteurs. La seule exception étant un ouvrage de Hamsun qui était répertorié dans la bibliographie de Hamsun de la Bibliothèque nationale norvégienne, mais qui manquait à la collection de la bibliothèque publique d'Oslo. Le processus de nettoyage nous a donc involontairement donné une méthode d'identification des œuvres manquantes dans les collections de la bibliothèque. À part cela, c'est une procédure fastidieuse de se retrouver avec de longues listes de centaines de manifestations.

### **Transformation en RDF**

Les fichiers de données transformés en FRBR ont été convertis en RFD en utilisant XSLT<sup>2</sup> et un mélange entre les champs MARC et les prédicats RDF développés par le projet. Ce mélange a utilisé principalement des vocabulaires et des ontologies connus tels que les termes de métadonnées du Dublin Core, l'ontologie bibliographique, le Core FRBR, FOAF et SKOS. Mais ce projet a aussi construit plusieurs sous-propriétés plus spécifiques pour exprimer nos données plus exactement que ces vocabulaires ne le permettaient. Par la suite, le projet a découvert que les vocabulaires RDA<sup>3</sup> contiennent des prédicats qui recouvrent beaucoup d'informations spécifiquement documentaires qu'il nous fallait exprimer. Dans des révisions ultérieures du mélange, certains de ces termes ont remplacé nos propres prédicats.

Voir le mélange complet à <http://www.bibpode.no/blogg/?p=1573>.

Une fois les données converties en RDF, elles ont été enrichies par des liens à d'autres fichiers de données. Les œuvres ont été liées à des instances de DBpedia et du projet Gutenberg, tandis que les personnes ont été liées à DBpedia et VIAF. Pour permettre de

mettre une liste d'œuvres en ordre chronologique, le projet a ajouté des informations sur la date de la première édition aux instances de l'œuvre. Il n'est pas facile d'extraire cette information d'une notice MARC, pour peu qu'elle soit présente. Avec ce nouveau fichier enrichi de données, le projet a pu développer une application web qui a permis à un usager de parcourir la collection complète que possède la bibliothèque de ces auteurs en choisissant à partir d'une courte liste de titres d'œuvres au lieu de parcourir une longue liste de centaines de manifestations. De plus, cette application pourrait donner aux usagers des informations pertinentes sur les auteurs provenant de DBpedia ainsi que des liens aux versions numériques en texte intégral du Projet Gutenberg.

Une application web simple a été développée pour permettre aux usagers de parcourir cette partie de la collection de la bibliothèque, concentrée en entités FRBR avec des informations complémentaires fournies par des sources externes grâce à la liaison de données (4).

La représentation RDF des deux auteurs nous a tout de suite donné un fichier de données plus exploitable. L'expérience a été si positive que nous avons décidé de convertir tout le catalogue de la bibliothèque pour voir si cela pourrait exaucer un souhait fondamental: améliorer les services aux usagers.

### **Les données liées et RDF comme base de nouveaux services**

En liaison avec les projets «recommandations de livres» et «rayonnages actifs» de la bibliothèque publique d'Oslo, nous avons beaucoup travaillé à la conversion du catalogue de bibliothèque en format RDF. L'effort que nous avons fourni à ce stade a été beaucoup plus détaillé et minutieux et nous a ouvert beaucoup plus de perspectives qu'à l'époque du projet Pode. Le catalogue de la bibliothèque publique d'Oslo comprend environ 420 000 notices et remonte aux années soixante-dix.

RDF est un format qui rend les données d'un catalogue de bibliothèque vraiment lisibles par ordinateur. La lecture par ordinateur dans le format MARC veut dire que des ordinateurs peuvent lire, conserver et traiter les *caractères* d'une notice de catalogue, tandis que la lecture par ordinateur en RDF signifie que l'ordinateur peut lire la *signification* réelle, ou la sémantique, de nos données. En bref, dans la conversion des données de bibliothèque en RDF il s'agit de «traduire» des notices de catalogue en fichiers d'affirmations simples, où nous utilisons des identifiants uniques à la fois pour les choses dont nous voulons parler et pour l'information que nous voulons donner au sujet de ces choses.

Pouvons-nous avoir de nouvelles idées sans oublier notre ancienne expertise? Nous ne sommes pas la première bibliothèque du monde à convertir nos données de bibliothèque en RDF. Mais tandis que beaucoup de travaux se sont concentré sur l'extraction et la conversion d'un petit sous-ensemble de données d'informations essentielles des notices de catalogue, nous avons essayé de «tout» emmener avec nous. Le travail des catalogueurs en bibliothèque est minutieux et inestimable, et quiconque a travaillé en bibliothèque sait qu'il est crucial de pouvoir accéder à plus d'informations dans nos documents que le titre, le nom du créateur, le sujet et l'ISBN. Nous avons essayé de

conserver autant d'informations que possible des notices MARC et de les exprimer de la façon la plus précise et la plus correcte possible. Nous avons systématiquement passé en revue la norme NORMARC, champ par champ, et nous avons décidé comment interpréter les données et comment exprimer le contenu en RDF.

Voir notre script complet de conversion qui convertit le MARC binaire en RDF en utilisant le mappage YAML à <https://github.com/digibib/marc2rdf>

En même temps, il est important qu'une version RDF du catalogue ne devienne pas tout simplement un fichier de données qui dise exactement la même chose dans une autre langue. Les règles de catalogage et le format MARC ont été conçus il y a plusieurs décennies pour obtenir les meilleurs résultats possibles selon certaines limitations, et certaines de ces limitations ne sont plus aussi pertinentes aujourd'hui qu'à l'époque où ces normes ont été inventées. De plus, beaucoup de ces règles ont la double fonction de description et de localisation des documents. Si on reprend les anciennes règles et les anciennes idées dans le nouveau format, elles viendront avec les mêmes limitations. Le format RDF nous donne des possibilités qui n'existent pas dans MARC et il est important d'en tirer parti.

Voici un exemple précis: Qu'est-ce que cela signifie lorsqu'une notice de catalogue dit que Karl Marx est l'auteur du «Manifeste communiste» tandis que Friedrich Engels en est le co-auteur ? Cette distinction entre les deux fonctions n'a rien à voir avec les fonctions réelles des deux personnes. En fait nous n'avons pas vraiment de champ pour les auteurs dans le format MARC: nous avons le champ d'«accès principal». La fonction de ce champ est de dire où placer un document dans une collection. Un document physique ne peut se trouver qu'à un seul endroit, et c'est pourquoi nous ne pouvons avoir qu'un seul accès principal. Le nom de Marx est cité avant celui de Engels sur la page de titre, et c'est pourquoi il est le seul à recevoir l'honneur d'être cité comme auteur du livre, tandis que le rôle de Engels est réduit à celui de «co-auteur».

Lorsque nous exprimons cette information en RDF nous n'avons pas besoin de faire attention à ce genre de distinctions artificielles et non-fonctionnelles. Karl Marx est auteur du livre ainsi que Friedrich Engels. Il n'y a pas de conflit entre ces affirmations.

Autre exemple: lorsque l'information fournie par les notices du catalogue est ambiguë et dépend du contexte. Prenez par exemple ce point d'accès secondaire:

\*700 0 \$a King, Stephen

\$d 1947-

\$e forf.

\$j am.

\$t Rita Hayworth and Shawshank Redemption

Quel est le lien entre la nouvelle «Rita Hayworth and Shawshank Redemption» et le document catalogué dans ce cas? Tout dépend du contexte. Si le document catalogué est un livre, cette affirmation veut probablement dire que cette nouvelle fait partie du document. Si le document catalogué est un DVD, cela veut probablement dire que le film est tiré de l'histoire de cette nouvelle. Une interprétation correcte des données demande que l'on connaisse les textes et les films et leurs connexions. En d'autres termes, il faut un lecteur humain.

Ces deux exemples nous montrent les limitations de ce que l'on peut exprimer en format MARC. Le premier exemple nous montre comment les concepts d'accès principal et secondaire n'ont pas été conçus originellement pour décrire des livres, des films ou des documents sonores. Ils ont été créés pour organiser des collections physiques et pour décider de quelles façons l'utilisateur doit pouvoir consulter un catalogue sur fiches. Le deuxième exemple concerne le degré de lecture par ordinateur. Lorsque l'interprétation correcte des notices de données dépend du savoir et du bon sens humains, les données ne sont pas vraiment lisibles par ordinateur.

SPARQL est le langage de requête que nous utilisons pour accéder aux données RDF. On peut effectuer n'importe quelle recherche ordinaire sur le catalogue avec SPARQL, mais on obtient beaucoup plus.

La recherche documentaire CCL nous offre beaucoup de façons de combiner des requêtes pour faire des recherches avancées et spécifiques. Mais il y a une limitation à laquelle on n'échappe jamais: quelle que soit la façon dont on construit sa requête, on obtiendra à l'arrivée des notices de catalogue. Vous pouvez poser la question que vous voulez à condition que cela commence par les mots : «Quels livres ...»

On ne peut pas demander une liste de sujets dans une langue particulière, ou obtenir une liste d'auteurs qui ont écrit sur un sujet particulier. Pour répondre à de telles questions, il faudrait trouver les *livres* que la bibliothèque a dans cette langue ou sur ce sujet, et ensuite exécuter la tâche fastidieuse de passer cette liste en revue pour trouver les sujets ou les auteurs différents.

Une autre limitation des recherches documentaires traditionnelles est qu'elles ne peuvent pas contenir d'entités inconnues. Par exemple, on ne peut pas construire une recherche CCL qui donne tous les auteurs des livres ayant un certain titre. Il faut d'abord chercher le titre connu, trouver le nom de l'auteur et ensuite faire une autre recherche pour trouver les livres écrits par cet auteur. Cela n'est pas très difficile, mais que se passe-t-il si l'information dont vous avez besoin dépend de plus d'une seule entité inconnue? Une question telle que «Quels romans puis-je trouver à la bibliothèque qui soient aussi disponible en DVD ?» ne trouve pas de réponse avec les outils traditionnels de recherche documentaire, à moins de disposer d'énormément de temps. À l'aide de SPARQL vous pouvez poser la question:

Donne moi toutes les paires de documents, doc-1 et doc-2, de sorte que :

Doc-1 est un film en DVD

Doc-2 est un livre contenant un roman

Doc-1 est basé sur l'œuvre X

Doc-2 est une manifestation de l'œuvre X

### **Remarques pour conclure**

Nous allons continuer notre travail de représentation RDF de notre catalogue. La prochaine étape sera de mettre en place une recherche pour voir si les résultats seront meilleurs que ceux de notre catalogue d'accès public en ligne à l'heure actuelle. Vous pouvez suivre notre progrès à <http://digital.deichman.no/>

### **Annexe – nettoyage et corrections**

La tâche principale a été de s'assurer que toutes les notices traduites (soit à cause d'une langue étrangère, soit à cause des réformes de la langue norvégienne) contiennent le titre original de l'ouvrage dans le champ 240. Lorsque le champ 240 manquait, il a été ajouté automatiquement sur la base de l'information présente dans le champ de note 574, un champ spécifique à NORMARC contenant des informations sur le titre original. Lorsque ce champ de note manquait ou lorsque la conversion automatique avait échoué, le titre était ajouté manuellement.

Une autre partie du travail qui a pris beaucoup de temps a été d'identifier et de déterminer les titres réels (originaux) des ouvrages significatifs dans les champs 245 et 740/700 |t et de mettre l'indicateur adéquat selon les règles de correction.

En ce qui concerne le second indicateur dans le champ 740, le projet a choisi d'utiliser la valeur 2 seulement lorsque le champ contenait un titre d'œuvre significatif. Pour tous les autres titres l'indicateur a été mis à 0, même si ce sont des points d'accès analytiques. Même si cette pratique ne se conforme pas à la convention d'utiliser 700‡a + ‡t pour les points d'accès titres analytiques, il a été décidé que ce serait l'approche la plus efficace pour détecter des œuvres analytiques publiées de façon non ambiguë.

Se basant sur les résultats du premier essai de transformation FRBR, les corrections dans les 980 notices pour Hamsun et Petterson comprenaient :

- correction du code de langage dans le 008 pour 5 notices
- ajout du titre de forme (ou « titre original » pour correspondre précisément à la terminologie NORMARC) dans les champs 240 de 85 notices et correction de fautes de frappe des champs 240 existants pour 24 notices
- correction de fautes de frappe dans 245‡a (ou mauvaise syntaxe ISBD) pour 6 notices
- correction du premier indicateur dans 245: avant la correction 137 notices avaient pour indicateur 1=0 ou nul, tandis que l'indicateur 1=1 était utilisé dans 774 notices. Après correction la distribution était de 263-651
- correction des champs 700. Dans les notices originales, nous avons trouvé 948 champs 700‡a et 545 700‡t. Dans les notices corrigées, ces chiffres sont réduits respectivement à 917 pour 700‡a et 481 pour 700‡t. Ce changement est dû à une utilisation plus

systématique des champs 740 dans toutes les notices qui avaient le même auteur (qui est marqué dans 100).

- Changement du deuxième indicateur dans le champ 700 pour pouvoir clarifier si un point d'accès est une œuvre unique ou non

1. Bien que les relations entre les champs MARC soient exprimées implicitement dans une notice bibliographique, il n'est pas facile d'exprimer des relations entre des champs sans ambiguïté et de façon explicite dans les cas où le document comprend plusieurs œuvres.

2. Dans des travaux ultérieurs de ce type, nous avons utilisé des scripts Ruby avec un fichier de mappage YAML au lieu de XSLT. Pour plus de détails voir <https://github.com/bensinober>

3. <http://rdvocab.info/>

4. <http://bibpode.no/linkedauthors/>

## **Références**

Aalberg, T. (2006). A Tool for Converting from MARC to FRBR. *ERCIM News*, (66). Retrieved from [http://www.ercim.eu/publication/Ercim\\_News/enw66/aalberg.html](http://www.ercim.eu/publication/Ercim_News/enw66/aalberg.html)