



# 68th IFLA Council and General Conference

## August 18-24, 2002

---

**Code Number:** 053-133-F  
**Division Number:** IV  
**Professional Group:** Cataloguing  
**Joint Meeting with:** -  
**Meeting Number:** 133  
**Simultaneous Interpretation:** Yes

### Orpailier MARC pour trouver... les FRBR ?

#### **Knut Hegna**

Bibliothèque universitaire d'Oslo  
Oslo, Norvège  
E-mail: [Knut.Hegna@ub.uio.no](mailto:Knut.Hegna@ub.uio.no)

#### **Eeva Murtomaa**

Bibliothèque universitaire d'Helsinki  
Helsinki, Finlande  
E-mails: [Eeva.Murtomaa@helsinki.fi](mailto:Eeva.Murtomaa@helsinki.fi)

---

#### **Abstract:**

*Cette intervention donne le résumé d'un projet consistant à analyser des données MARC provenant de deux bibliographies nationales à la lumière du modèle de données présenté dans l'étude FRBR de l'IFLA. Au cours du déroulement de ce projet nous avons découvert que même si les informations contenues dans les notices MARC comportent des attributs pertinents pour identifier les entités œuvre, expression et manifestation, la précision et la syntaxe formelle en sont trop simples pour que les programmes puissent les exploiter convenablement. Certains résultats peuvent servir à proposer de meilleures listes de réponses dans les OPACs. Le projet émet deux propositions d'interface utilisateur OPAC, sur la base des principes de l'étude FRBR et des résultats du projet. Le rapport intégral est disponible à l'adresse suivante :*

<http://folk.uio.no/knuthe/dok/frbr/datamining.pdf>

---

#### **FRBR et MARC**

En premier lieu nous nous sommes demandé si la structure des FRBR pouvait être décelée dans des notices MARC existantes. Si nous regardons une simple notice, nous constatons qu'elle nous donne des informations sur l'œuvre, sur l'expression, et sur la manifestation :

```
*008880325 no esp
*02000 $a 82-991075-2-0 $b h. $c Nkr 60.00
*04110 $a espnor
```

\*08200 \$a 839.822[S]  
 \*10010 \$a Ibsen, Henrik \$d 1828-1906  
 \*24510 \$a Puphejmo (1879) \$c Henrik Ibsen ; tradukis:  
 Odd Tangerud ; lingve kontrolita de Esperantista Verkista  
 Asocio (EVA)  
 \*26000 \$a Hokksund \$b Eldonejo Odd Tangerud \$c 1987  
 \*26900 \$a [Drammen] : Tangen-trykk  
 \*30000 \$a [1], 57 s. \$c 24 cm  
 \*50000 \$a Originaltittel: Et dukkehjem. -  
 Originalutgave: København : Gyldendal, 1879  
 \*99100 \ \$a Tangerud, Odd

Il s'agit d'une notice puisée dans la bibliographie nationale norvégienne et qui décrit un document contenant une version en espéranto de la célèbre pièce *Maison de poupée* d'Henrik Ibsen. Le titre de l'œuvre se trouve en zone 500, la relation entre l'œuvre et son créateur en zone 100. Ces informations peuvent permettre d'identifier l'œuvre dans cette notice.

En outre, nous trouvons des informations relatives à l'expression par le biais des codes de langue dans les zones 008 et 041, ainsi qu'une relation avec la personne responsable de la traduction en partie dans la mention de responsabilité et dans la zone 991.

Quant aux informations relatives à la manifestation, elles se trouvent dans les zones de description du document en main : 245, 260, 269, 300 etc.

Nous avons donc une première réponse à notre question :

**Oui** : POURQUOI ? Parce qu'une notice bibliographique peut

- décrire tant l'œuvre que la manifestation
- contenir des éléments relatifs à l'expression
- s'enrichir de relations grâce aux points d'accès et aux vedettes-matière.

Dans une certaine mesure on peut donc déceler des éléments du modèle FRBR dans une notice MARC !

Malheureusement, la réponse est également **Non**. POURQUOI ?

- Parce que les règles de catalogage sont adaptées aux catalogues sur fiches et aux bibliographies imprimées, pas au modèle FRBR
- les informations fondamentales sont souvent entrées d'une manière plus satisfaisante pour l'intelligence humaine et l'œil humain que pour la machine.

Mais il ne suffit pas de regarder une simple notice. Nous avons décidé d'analyser des lots de notices générés par une recherche dans les bibliographies nationales finlandaise et norvégienne. Nous nous sommes limités à la recherche auteur.

Le titre de notre projet se termine par un point d'interrogation. Nous ignorions quels résultats nous obtiendrions. Nous avons le sentiment que notre recherche ne pouvait pas aboutir sur des résultats définitifs. À tout le moins nous nous attendions à ce qu'émergent certains problèmes à ce sujet avec les règles de catalogage et le format MARC.

Nous avons commencé par regarder les tableaux de l'étude FRBR qui visent à dégager les attributs des entités qui revêtent la plus grande importance dans l'identification de l'œuvre, de

l'expression et de la manifestation. Puis nous avons tenté de mettre en correspondance les informations puisées aux notices MARC avec ces attributs.

Attribut FRBR	Valeur FRBR	NORMARC	FINMARC	Attribut retenu ?
titre de l'œuvre	haute		241 \$a	oui
		500 \$a	500 \$a	oui
		505 \$a	505 \$a	oui
			248 \$h	oui
		240 \$a	240 \$a	oui
	245 \$a	245 \$a	oui	
relation avec la personne responsable	haute	100 \$a	100 \$a \$h	oui
		70010 \$a	70010 \$a \$h	
complétude visée	haute	?	?	non
forme de l'œuvre	moyenne	interprétation de la Dewey	008 pos. 24-27, 29-30, 33-34	non

Tableau 1 : Attributs permettant d'identifier l'entité œuvre dans les formats MARC

Pour obtenir le titre de l'œuvre nous cherchons les titres uniformes ou les titres originaux. Nous vérifions tout d'abord s'il y a une zone 241 (titre original). La bibliographie nationale norvégienne n'utilise pas cette zone, contrairement à la bibliographie finlandaise. S'il n'y a pas de zone 241, nous recherchons la présence dans les zones 500 et 505 de la mention *Titre original* dans les deux langues : *Originaltittel* ou *Originaltitler* (pluriel), abrégé *Orig. tit.*, en norvégien, *Alkuteos* ou *Alkuteokset* (pluriel), en finnois. Cette mention est toujours suivie des titres originaux. Si à l'issue de ce test on n'a trouvé aucun titre original dans les zones 50X, nous passons à d'autres zones de titre dans l'ordre suivant : 248 \$h (titre propre de partie dans les notices finlandaises) et 240 \$a. En dernier recours, le ou les titres originaux sont pris de la zone 245.

Habituellement les titres figurant en zone 505 sont répétés dans les zones d'accès titres (745 \$a en Finlande, 740 \$t en Norvège), mais il est impossible de déterminer si les données contenues dans ces zones représentent des titres originaux ou non. Tantôt il s'agit bien de titres originaux, tantôt d'autres formes du titre. Ni la 745 ni la 740 n'accompagnent le titre d'un qualificatif. Dans bien des cas ces zones 74X sont des points d'accès relatifs à une autre œuvre qui est en relation d'une manière ou une autre à l'œuvre principale reflétée dans la notice.

Nous avons eu le sentiment qu'en mettant le titre original en relation avec le créateur de l'œuvre il nous était possible d'identifier l'œuvre ou les œuvres reflétées dans chaque notice. Sur cette base, à l'aide de ces données, nous avons pu trouver les cooccurrences d'œuvres identiques dans plusieurs notices et les différencier d'autres œuvres. Cela fait, nous avons utilisé d'autres données pour différencier et trouver les cooccurrences d'expressions, identifiées au moyen du code de langue et du nom du traducteur. D'autres données ont permis d'identifier les manifestations.

Voici quelques résultats :

<b>Auteur</b>	<b>Nombre de notices</b>	<b>Nombre d'identifiants d'œuvres</b>	<b>Nombre d'identifiants d'œuvres uniques</b>
Ibsen, Henrik (norv.)	744	914	220
Wassmo, Herbjørg (norv.)	149	159	19
Gaarder, Jostein (norv.)	237	237	14
Solstad, Dag (norv.)	92	93	35
Kunnas, Mauri (finl.)	130	133	41
Jansson, Tove (finl.)	576	595	92
Linna, Väinö (finl.)	168	168	20

Tableau 2 : Résultats de la procédure de réduction du nombre d'œuvres reflétées dans des notices provenant des bibliographies nationales norvégienne (norv.) et finlandaise (finl.)

Nous constatons qu'il y avait 744 notices où Henrik Ibsen figurait en tant qu'auteur ou auteur secondaire. Nous avons pu dégager de ces notices 914 identifiants d'œuvres et sur ces 914 identifiants nous avons pu identifier 220 œuvres uniques. Soit ; mais Ibsen a écrit 26 pièces et quelques recueils de poèmes : alors, d'où viennent ces 220 œuvres ?

Certaines notices comportaient des informations déficientes, voire carrément erronées : recueils de pièces ou théâtre choisi où les titres originaux ou uniformes étaient absents, ou présents mais non identifiables (12) ; 33 notices où le complément de titre était inclus dans le titre propre, générant 8 nouvelles œuvres uniques ; 3 notices avec des coquilles dans le titre original ; 14 notices où l'orthographe plus ou moins modernisée différait du titre original ; 10 notices totalement dépourvues de titres originaux générant des vedettes d'œuvres dont le titre était libellé dans une langue étrangère.

Tous ces problèmes sont détaillés dans notre rapport.

## **L'interface utilisateur**

Nous avons donc trouvé une structure. La seconde question que nous nous sommes posée portait sur le bénéfice que nous pourrions tirer de cette structure dans l'affichage des réponses à la requête d'un utilisateur.

En premier lieu nous avons regardé les tableaux des FRBR visant à déterminer quels étaient les attributs les plus importants pour choisir entre des entités analogues et nous avons mis en correspondance ces attributs et les étiquettes MARC.

Pour faciliter le choix il importe en outre de déterminer le meilleur ordre d'affichage. Le principe de classement devrait être facilement identifiable par l'utilisateur et devrait être fonction du niveau de l'entité présentée.

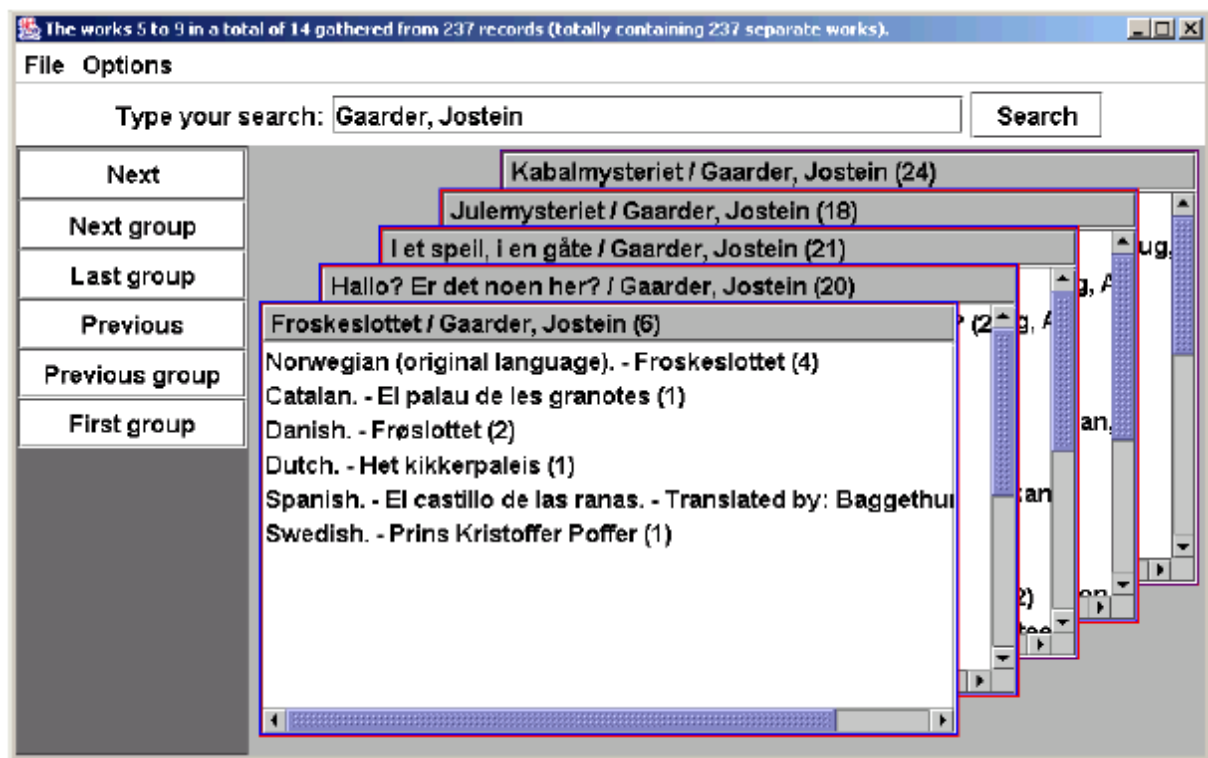
Nous sommes d'avis que la liste des réponses à une requête devrait toujours s'afficher en fonction de la requête lancée et des réponses elles-mêmes. Quand on recherche une personne en particulier, la liste des réponses devrait comporter soit les œuvres de cette personne classées selon un ordre ou un autre, alphabétique peut-être ou bien chronologique, soit une liste des différents rôles que cette personne a joués par rapport aux diverses entités (auteur, illustrateur, traducteur...).

La liste des réponses est représentée sous la forme d'un empilement de fiches. Le dessus de l'ensemble des fiches, intégralement visible, définit un axe horizontal qui s'enfonce dans la profondeur de l'écran ; la surface des fiches constitue le second axe, vertical, de ce système à deux axes et donne sous la vedette de chaque fiche des informations différenciées.

Les deux axes peuvent représenter les binômes *auteurs-œuvres*, *œuvres-expressions*, *expressions-manifestations* ou *sujets-œuvres*, selon la requête initialement lancée. Si l'utilisateur recherche un auteur en particulier et qu'il y a plusieurs réponses, l'axe horizontal donnera la liste des noms de personnes, et l'axe vertical la liste des œuvres de chacune de ces personnes. Si au contraire la requête n'aboutit qu'à une seule réponse, l'axe horizontal donnera la liste des œuvres de l'auteur, et l'axe vertical la liste des expressions de chaque œuvre.

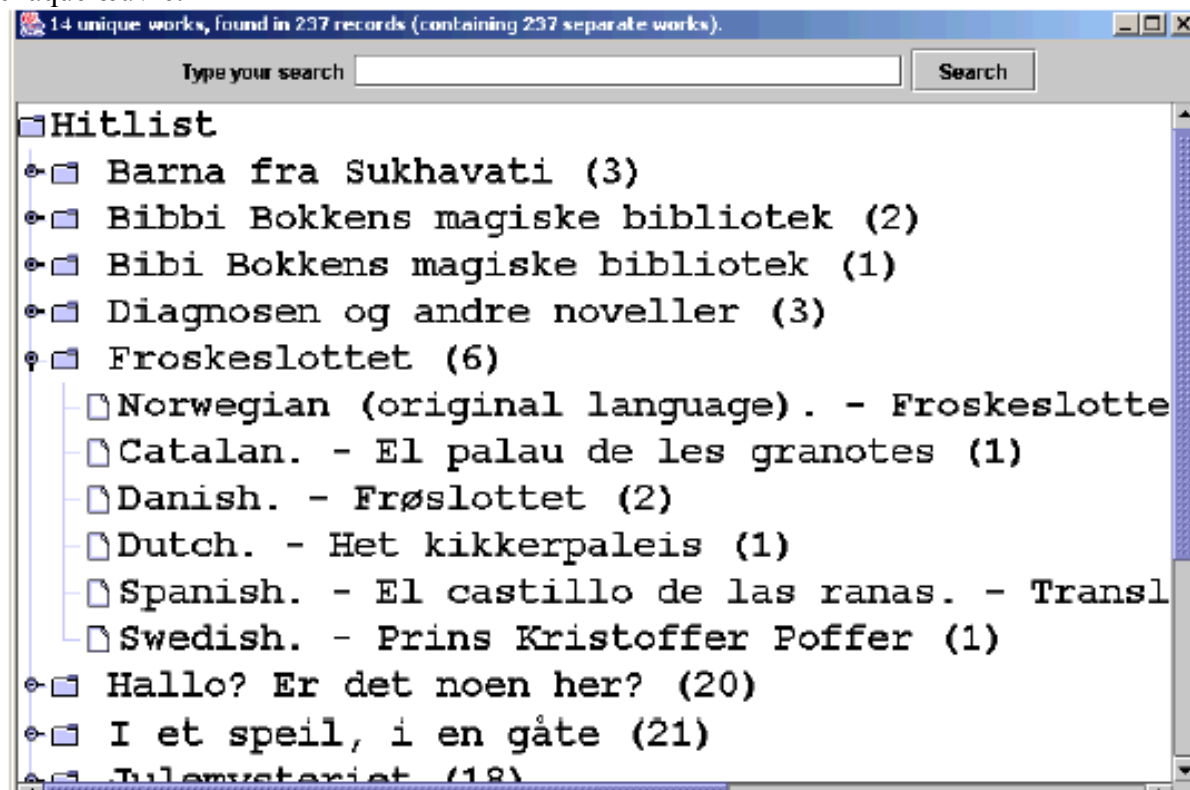
On peut en trouver un exemple ci-dessous dans la première interface, avec les œuvres de *Jostein Gaarder* le long de l'axe horizontal et les expressions le long de l'axe vertical sous le titre de chacune des œuvres. Le nombre d'expressions identifiées pour chaque œuvre est

indiqué entre parenthèses à la suite de la vedette titre de l'œuvre. Le nombre de manifestations pour chaque expression figure à la fin des informations relatives à l'expression.



L'utilisateur pourrait alors choisir une expression en particulier, provoquant ainsi l'ouverture d'une nouvelle fenêtre avec toutes les manifestations listées sous cette expression. Les manifestations de l'expression choisie sont classées par ordre chronologique, et l'année d'édition apparaît en tête de chaque notice. La fenêtre comporte des boutons associés à diverses fonctions telles que *impression*, *enregistrement* et *tri* et elle ne disparaît que si l'utilisateur la referme expressément. Cela implique que l'utilisateur peut garder autant de lots de manifestations qu'il le souhaite, chacun dans une fenêtre séparée.

La seconde interface utilisateur présente les œuvres sous forme d'arborescences à partir d'une liste de réponses. Les œuvres sont classées par ordre alphabétique du titre original et le nombre qui figure à la suite du titre correspond au nombre d'expressions identifiées pour chaque œuvre.



Sur chaque œuvre il est possible d'ouvrir une grappe d'expressions. Les expressions sont classées de la même manière que dans l'interface précédente, la langue originale venant en premier puis les autres langues, dans l'ordre alphabétique.

En cliquant sur le symbole correspondant à une expression on obtient l'ouverture d'une nouvelle fenêtre avec les manifestations, comme précédemment.

## En bref

L'une des principales raisons qui expliquent le bruit que nous avons obtenu au cours de nos expériences tient à l'utilisation de la zone 700 sans qualificatif et à la manière dont les systèmes indexent cette information en association avec la zone 100 sans prendre en considération la fonction qu'exerce réellement la personne dont le nom est en zone 100.

Nous avons le sentiment que les règles de catalogage devraient rendre obligatoire l'indication de fonction en zone 700 et que la liste des codes de fonction devrait en outre être enrichie. Les systèmes doivent utiliser ces codes de fonction pour proposer à l'utilisateur final des listes de réponses mieux structurées.

Cela permettrait aux systèmes de présenter toutes les fonctions qu'une personne est susceptible d'avoir dans une base de données, de sorte que l'utilisateur final pourrait choisir soit l'ensemble soit un sous-ensemble de ces fonctions au cours de sa navigation dans le catalogue.

Nos recherches ont également montré que l'analyse serait facilitée si les titres originaux pouvaient être saisis de manière plus cohérente, dans des zones distinctes et répétables.

Les codes de langue constituent l'un des attributs les plus importants pour identifier les diverses expressions d'une œuvre donnée. Ils jouent également un rôle important pour déterminer si une manifestation contient une traduction ou non.

La normalisation des données est en soi un procédé d'établissement de liens. L'utilisation de fichiers d'autorité dans l'environnement bibliographique aide à établir la structure de navigation, à la fois en montrant les cooccurrences des entités et en les différenciant. Les fichiers d'autorité les plus fréquents concernent les noms de personnes et de collectivités. Notre travail démontre qu'il y aurait également besoin d'autorités pour les œuvres, afin de montrer les cooccurrences d'une même œuvre sous une même vedette.

Merci.

Translations by Patrick Le-Boeuf.