



Identifiants: franchir les barrières linguistiques

Jan Pisanski
Maja Žumer
University of Ljubljana
Ljubljana, Slovenia

&

Trond Aalberg
Norwegian University of Science and Technology
Trondheim, Norway

Traduit de l'anglais par :
Anila Angjeli,
Bibliothèque nationale de France

Meeting: **93. Cataloguing**

WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY

10-15 August 2010, Gothenburg, Sweden
<http://www.ifla.org/en/ifla76>

Résumé :

L'identification des entités bibliographiques est un élément important permettant de franchir les barrières culturelles et linguistiques dans le monde des bibliothèques car elle autorise l'utilisation et la réutilisation des données bibliographiques dans de multiples applications. Le présent article se penche sur les identifiants existants susceptibles d'être utilisés pour l'identification des entités formelles et étudie la possibilité de leur utilisation à grande échelle au sein des bibliothèques et au-delà.

Introduction

La masse des informations traitées par les bibliothèques étant en permanent accroissement, il est important que les utilisateurs comprennent clairement l'environnement avec lequel ils entrent en interaction. En termes de Fonctionnalités requises pour les notices bibliographiques (FRBR) – le nouveau paradigme de l'univers bibliographique – les utilisateurs ont besoin de distinguer entre les instances des entités afin d'effectuer les tâches des utilisateurs : trouver, identifier, sélectionner, obtenir, ainsi que d'autres tâches non identifiées par les FRBR telle que la tâche « explorer ». Il existe un besoin de plus en plus grand d'identifier de façon univoque les instances de toutes les entités importantes dans l'univers bibliographique. Le modèle FRBR étant l'unique modèle de l'univers bibliographique officiellement reconnu, il paraît naturel d'avoir des identifiants pour les

entités qui constituent l'ossature du modèle FRBR. Certes, FRBR est un modèle qui se prête à différentes interprétations dont quelques définitions sont vagues. Cela ne veut pas dire pour autant que des efforts ne doivent pas être faits pour identifier les différentes entités. Bien qu'une parfaite identification n'est pas toujours possible, étant donné que l'univers bibliographique lui-même n'est ni tout noir ni tout blanc, un système d'identifiants distincts pour les entités devrait aider les utilisateurs à trouver leur chemin, et cela indépendamment de la façon d'interpréter l'univers bibliographique.

Le besoin d'identification dans les bibliothèques

Traditionnellement les bibliothèques se sont fiées à la combinaison d'attributs comme unique moyen d'identifier leurs documents et les autres entités (vedettes pour les noms de personnes, vedettes uniformes). Ce type d'identification fonctionne relativement bien dans l'enceinte d'une bibliothèque traditionnelle et dans un environnement bien délimité. Or, aujourd'hui les bibliothèques ne peuvent plus se permettre la redondance de l'information produite par la duplication du catalogage et du travail sur les données d'autorité lorsque les bibliothèques travaillent de façon isolée. De plus, l'application d'éléments d'identification dans les catalogues particuliers a souvent laissé à désirer, ce qui a entraîné des incohérences dans les données. Certes, les coûts du travail d'identification sont élevés ; cependant le développement des technologies de l'information et de la communication cré les conditions d'un partage efficace et rentable du travail, même à une échelle mondiale. Même si les identifiants locaux sont satisfaisants pour les applications locales, ce sont les identifiants reconnus internationalement qui possèdent une véritable force d'identification. Les utilisateurs ne veulent pas limiter leurs recherches à une seule base bibliographique et de plus en plus exigent un accès intégré à une large gamme de ressources accessibles en ligne, ce qui accentue le besoin d'harmonisation de l'information bibliographique concernée. Et enfin, les bibliothèques ont besoin de pouvoir intégrer le web social et en faire totalement partie afin de pouvoir moissonner et diffuser les contenus produits via la contribution des utilisateurs en tant que ressources complémentaires.

Toutefois, dans ces circonstances, les identifiants dépendant des traditions culturelles – y compris ceux dépendant d'une langue ou d'une écriture – vont inévitablement présenter des inconvénients. Certes, les utilisateurs veulent, méritent et doivent recevoir des solutions culturellement appropriées dans l'affichage de l'information. Le besoin d'afficher l'information de façon bien identifiable et le besoin d'identifier de façon unique les instances des entités sont, quoi qu'il en soit, deux problématiques bien distinctes. La véritable identification univoque est celle effectuée au moyen d'identifiants indépendants des langues. Utiliser les noms pour une identification universelle de tout type d'entité n'est pas une bonne solution. Non seulement les noms peuvent changer, ce qui n'en fait pas de bons candidats pour une identification univoque, mais ils ont également l'inconvénient de dépendre des pratiques culturelles et des langues. Si certains pensent que nous pouvons tous utiliser la même forme d'un nom à fin d'identification, une telle solution impose la mainmise d'une culture sur toutes les autres. Il serait difficile de trouver une solution acceptable pour tous. En fait, il n'y a aucunement besoin que l'identification se fasse de cette manière. Comme l'a formulé le Groupe de travail sur l'avenir du contrôle bibliographique de la Bibliothèque du Congrès (2008, p. 24) : « L'utilisation des chaînes de caractères de nature linguistique [...] comme identifiants, à la fois pour l'affichage et pour la manipulation des données, empêche l'échange des données entre langues et entre différentes communautés de données.

Les identifiants peuvent constituer la base des fichiers d'autorités et ainsi permettre d'éliminer la redondance dans les catalogues et même rendre les catalogues potentiellement

plus faciles à utiliser. L'effective utilité des fichiers d'autorité pour les utilisateurs dépend beaucoup des applications. Il y a des cas où les fichiers d'autorité ne sont pas liés aux notices bibliographiques. Une solution où tout changement dans le nom d'un auteur (par exemple résultant d'un mariage) oblige à modifier manuellement toutes les données dans les notices bibliographiques, ne devrait pas être une solution acceptable.

Qui plus est, les identifiants sont susceptibles d'être encore plus importants dans le contexte de nouvelles perspectives d'utilisation et de réutilisation des données produites et fournies par les bibliothèques. Par exemple, les identifiants sont la clé d'une intégration réussie des données bibliographiques dans le web sémantique et les bibliothèques s'intéressent de plus en plus à l'utilisation des Linked Data¹.

Bien que le besoin d'identification dans les bibliothèques ne soit pas nouveau (Tillett, 2007), l'âge numérique a fait émerger plusieurs initiatives variées, dont certaines ont été décrites par Vitiello (2004), Hakala (2006), Tillett (2007) et Babeu (2008).

Les identifiants

L'identification des entités est l'un des champs les plus intrigants de l'application des identifiants dans les bibliothèques et au delà. Le fait que le Groupe de travail de l'IFLA sur les Fonctionnalités requises pour les notices d'autorité (Functional Requirements and Numbering of Authority Records - FRANAR) a été chargé d'évaluer la faisabilité d'un Numéro international normalisé pour les données d'autorité (International Standard Authority Data Number – ISADN) est une preuve évidente de la reconnaissance de l'importance d'une identification correcte dans le milieu des bibliothèques bien au delà du groupe 2 des entités [FRBR]. Toutefois, le groupe en a conclu que l'établissement d'un tel numéro n'est pas faisable, bien que le rapport du groupe (Tillett, 2008) n'ajoute pas à cette conclusion de raisonnement économique détaillé.

Trois possibilités sont discutées dans le rapport. Le numéro normalisé ISADN a été considéré comme une très bonne solution en principe, mais coûteuse et difficile à maintenir ; l'identification fondée sur des chaînes de caractères textuels (vedettes autorisées uniques) a été complètement exclue ; l'identification fondée sur l'appariement des fichiers d'autorité provenant de sources différentes (tel que le projet VIAF) a été considéré comme une option appropriée pour de futurs développements. Le rapport reconnaît également l'établissement en cours par l'ISO de l'ISNI (International Standard Name Identifier = Code international d'identification des noms) en tant qu'identifiant normalisé pour les parties concernées et qui pourrait profiter à plusieurs communautés.

Il est important de savoir si les identifiants déjà existants pourraient être utilisés à plus grande échelle dans les bibliothèques et s'ils ne sont pas injustement oubliés et de voir si de nouveaux identifiants normalisés sont nécessaires dans le cadre des FRBR et de leur application par les bibliothèques. Une utilisation cohérente des identifiants, indépendamment de leur forme mais de préférence aussi normalisés que possible, devrait constituer à l'avenir, même si cet avenir est quelque peu flou, la pierre angulaire d'une utilisation en parallèle de données natives FRBR et de données héritées du passé. Les identifiants devront également ouvrir de nouvelles façons de réutiliser l'information bibliographique et de lier cette information à d'autres ressources, facilitant ainsi l'interopérabilité avec d'autres domaines.

¹ NdT : une traduction possible de l'expression « Linked Data » est « Web de données »

Toutefois il ne faut pas oublier que les différences culturelles dans l'interprétation de l'univers bibliographique peuvent limiter l'intérêt d'identifiants largement utilisés.

Il existe plusieurs identifiants internationaux visant à identifier les documents bibliographiques. Toutefois, ils ont été conçus non pas pour identifier les entités FRBR mais pour satisfaire les besoins de différentes communautés. Il est, par conséquent, difficile de définir de manière univoque l'entité FRBR que chacun d'eux identifie, comme cela peut être constaté dans le Tableau 1 qui présente des estimations effectuées par certains auteurs éminents. Notons que le tableau présente quelques inexactitudes dues au fait que les estimations ont été effectuées bien avant la reconnaissance internationale de certaines de ces normes. Comme ce tableau le montre, les identifiants de manifestations sont les moins problématiques.

Tableau 1: Comparaison des estimations effectuées sur l'identification des entités FRBR du groupe 1

	VITIELLO (2004)	GATENBY (2008)	LEBOEUF (2005)	HAKALA (2006)
ISBN	M	M	M	M
ISSN	M	M	M	M
ISRC	E	M	E	
ISAN	Œ, E*	Œ	Œ	
ISWC	Œ	Œ	Œ, E	
ISTC		Œ	E	Œ, E
ISMN	M	M	M	
V-ISAN		M	E	

*Même si le texte n'est pas suffisamment clair, il se peut que l'auteur puisse faire référence à ISAN qui identifie des œuvres et à V-ISAN qui identifie des expressions.

Les identifiants de manifestations sont de loin les identifiants d'entités les plus largement utilisées dans les bibliothèques. C'est pourquoi le constat que seulement 30% des documents référencés dans WorldCat possèdent un identifiant international, fait réfléchir (Gatenby, 2008). Si l'on considère que les données de WorldCat sont représentatives de l'état actuel des notices bibliographiques dans le monde, cette estimation devrait susciter une forte préoccupation, car il révèle un manque de préparation même lorsqu'il s'agit des manifestations.

Et pourtant, le succès relatif de l'ISBN et, à une moindre échelle, celui d'autres identifiants de manifestations tel que l'ISMN, montre que l'identification peut être réussie à une échelle mondiale si elle est proprement implémentée. Son adoption a paisiblement aidé tous ceux qui ont à faire d'une manière ou d'une autre aux livres et rares sont les discussions portant sur le coût du système qui fournit cette identification transparente. Toutefois, pour des raisons de couverture (types de ressources, âge et géographie), nombre de manifestations restent sans ISBN, de sorte que la solution a définitivement ses inconvénients. Holdsworth (2008) pointe un autre problème : en termes de FRBR, le même ISBN est parfois attribué à différentes manifestations de la même expression.

Tandis que l'identification des manifestations est relativement simple, l'identification des œuvres et des expressions est moins claire, en particulier parce que l'univers bibliographique couvre tous les types de ressources. De plus, les identifiants de ces niveaux plus abstraits

figurent rarement sur les manifestations elles mêmes. Malheureusement, très souvent pour les bibliothèques l'information qui ne figure pas dans ou sur la manifestation n'existe pas, même si elle est au mieux des intérêts de l'utilisateur. Il ne faut pas non plus oublier que jusqu'à présent tout le catalogage a été effectué au niveau de la manifestation et que le besoin d'identifier des œuvres et des expressions a été moins saillant. Par conséquent, l'identification des œuvres et des expressions est pour le moment presque inexistante dans les bibliothèques.

L'ISRC (International Standard Recording Code = Code international normalisé des enregistrements) est un identifiant d'expression et l'ISAN (International Standard Audiovisual Number = Numéro international normalisé d'œuvre audiovisuelle) est un identifiant d'œuvre ; ils n'appartiennent toutefois qu'à une petite partie de l'univers bibliographique et ne sont pas particulièrement bien utilisés dans les bibliothèques. D'un autre côté, la nature de l'ISWC (International Standard Music Work Code = Code international normalisé pour les œuvres musicales) et celle de l'ISTC (International Standard Text Code = Code international normalisé des œuvres textuelles) ne sautent pas aux yeux. Dans l'ISWC les arrangements, les adaptations et les traductions reçoivent des ISWC distincts (Antelman, 2004). La base de données ISWC Net (www.iswc.org) montre, en effet, des attributions incohérentes d'identifiants ISWC en termes de niveaux FRBR mais également en termes de couverture géographique.

La norme ISTC présente, elle aussi, des incohérences internes. Étant donné que l'ISTC est appliqué aux « œuvres textuelles » et que les textes sont associés aux expressions FRBR, l'ISTC devrait être attribué à ce niveau. Mais en fait, les exemples en annexe E de la norme ISTC (ISO 21047:2009) montrent que des identifiants ISTC distincts sont attribués à différentes versions de la même œuvre textuelle (par exemple des révisions et des traductions dans l'exemple E. 4). On peut donc déduire que les identifiants ISTC sont effectivement attribués au niveau de l'expression et non au niveau de l'œuvre au sens FRBR. Or cela est en contradiction avec la mention de l'annexe B (B. 13) de la version publiée de la norme ISTC qui stipule que la même œuvre textuelle ne doit pas recevoir plus d'un ISTC. Tandis que le texte de l'exemple E. 4 décrit les relations entre une œuvre textuelle et d'autres œuvres textuelles dérivées de celle-ci, les instances fournies indiquent clairement que les traductions et les révisions sont toujours considérées comme étant la même œuvre textuelle, alors qu'une adaptation pour enfants de l'œuvre textuelle « originale » ne l'est pas.

Parce que nous vivons dans un monde où l'identification des objets numériques est aussi importante que celle des objets non numériques, n'oublions pas des identifiants tel que le DOI. Le DOI (Digital Object Identifier = Identifiant d'objet numérique) est un identifiant numérique relativement bien utilisé pour les objets du domaine de la propriété intellectuelle. Cela ne présume pas qu'il identifie une entité particulière. En fait il peut identifier des manifestations physiques ou numériques, des représentations ou des œuvres abstraites (International DOI Foundation, 2006). Du point de vue FRBR cela signifie qu'une utilisation du DOI, même avec les meilleures intentions, peut conduire en réalité à la plus grande confusion.

Conclusion

Il existe différentes normes internationales pour identifier les différentes parties du présent univers bibliographique. Tandis que certains sont bien utilisés et applicables à une entité FRBR définie (par exemple ISBN, ISMN), la plupart des outils d'identification des entités FRBR à une échelle mondiale ne sont pas encore en place. Encore plus important, les

identifiants actuels visant à identifier seulement des fragments de l'univers bibliographique (par exemple texte, musique, ressource audiovisuelle) sont sous-utilisés, voire mal utilisés – parfois dû à des différences culturelles – et parfois utilisés pour identifier les instances des différentes entités FRBR sans mécanismes pour identifier proprement les entités elles-mêmes. Cette situation rend très peu vraisemblable la possibilité, dans un futur proche, d'identifier efficacement les entités FRBR au moyen des identifiants existants. D'un autre côté, des problèmes similaires sont susceptibles de surgir même avec l'adoption de nouveaux identifiants.

La communauté des bibliothèques est suffisamment mûre pour admettre que les identifiants internationalement utilisés, idéalement partagés par différentes communautés, sont essentiels à une intégration des bibliothèques au niveau mondial. Quels identifiants et que vont-ils identifier reste à voir. Si les coûts associés au maintien d'un système d'identification efficace sont élevés et que l'identification parfaite peut ne pas être toujours possible, maintenir le *statu quo* peut avoir des coûts encore plus élevés pour les bibliothèques.

Bibliography

Antelman, K. (2004). Identifying the Serial Work as a Bibliographic Entity. *Library Resources & Technical Services*. 48 (4), 238-255.

Babeu, A. (2008). *Building a »FRBR-Inspired« Catalog: The Perseus Digital Library Experience*.

<http://www.perseus.tufts.edu/~ababeu/PerseusFRBRExperiment.pdf>

Gatenby, J. (2008). *The activities of OCLC on FRBR*. Workshop on FRBR in the European Library, 9. October 2008, Lisbon, Portugal.

http://frbr.bnportugal.pt/documentos/The_activities_of_OCLC_on_FRBR.ppt.

Hakala, J. (2006). The seven levels of identification. *Program*. 40 (4), 361-371.

Holdsworth, M. (2008). *The Identification of Digital Book Content*. Report prepared for the Book Industry Study Group, January 2008.

http://www.bisg.org/docs/DigitalIdentifiers_07Jan08.pdf

International DOI Foundation (2006). *DOI Handbook*.

<http://www.doi.org/hb.html>

ISO 21047 (2009). *Information and documentation – International Standard Text Code (ISTC)*, 22 p.

Le Bœuf, P. (2005). *Identifying 'textual works'*. FRBR in 21st Century Catalogues, Dublin, Ohio, May 2-4 2005.

<http://www.oclc.org/research/events/frbr-workshop/presentations/leboeuf/ISTC.ppt>

Library of Congress (2008). *On the record: Report of The Library of Congress Working Group on the Future of Bibliographic Control*.

<http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>

Tillett, B. (2007). Numbers to Identify Entities (ISADNs – International Standard Authority Data Numbers). *Cataloging and Classification Quarterly*. 44 (3/4), 343-361.

Tillett, B. (2008). *A Review of the Feasibility of an International Standard Data Authority Number (ISADN)*. Prepared for the IFLA Working Group on Functional Requirements and Numbering of Authority Records, edited by G. Patton.

<http://archive.ifla.org/VII/d4/franar-numbering-paper.pdf>

Vitiello, G. (2004). Identifiers and identification systems. *D-Lib Magazine*. 10 (1).

<http://www.dlib.org/dlib/january04/vitiello/01vitiello.html>

International Standards Discussed in the Paper

ISO 2108 (2005). *Information and documentation – International Standard Book Number (ISBN)*, 21 p.

ISO 3297 (2007). *Information and documentation – International Standard Serial Number (ISSN)*, 20 p.

ISO 3901 (2001). *Information and documentation – International Standard Recording Code (ISRC)*, 9 p.

ISO 10957 (2009). *Information and documentation – International Standard Music Number (ISMN)*, 13 p.

ISO 15706-1 (2002). *Information and documentation – International Standard Audiovisual Number (ISAN) – Part 1: Audiovisual work identifier*, 12 p.

ISO 15706-2 (2007). *Information and documentation – International Standard Audiovisual Number (ISAN) – Part 2: Version Identifier*, 20 p.

ISO 15707 (2001). *Information and documentation – International Standard Music Work Code (ISWC)*, 10 p.

ISO 21047 (2009). *Information and documentation – International Standard Text Code (ISTC)*, 22 p.