

Qualité de WorldCat



Livre blanc d'OCLC



Les bibliothèques du monde,
connectées.™



Qualité de WorldCat

Livre blanc d'OCLC

Contributeurs principaux

Karen Calhoun, vice-présidente, Services WorldCat et de métadonnées

Glenn Patton, directeur, Gestion de la qualité de WorldCat

Copyright © 2011, OCLC Online Computer Library Center, Inc.
6565 Kilgour Place
Dublin, Ohio 43017-3395

TOUS DROITS RÉSERVÉS. Aucune partie de cette publication ne peut être reproduite, stockée sur un système d'extraction ou transmise sous quelque forme que ce soit ni par quelque moyen que ce soit, y compris électroniquement, mécaniquement, par photocopie ou de toute autre façon, sans autorisation préalable signée par le détenteur du droit d'auteur.

Connexion, FirstSearch, OCLC, le logo OCLC, WorldCat, WorldCat.org, WorldCat Resource Sharing et « The World's Libraries. Connected » sont des marques de commerce ou des marques de services d'OCLC.

Les produits, les noms d'entreprise ou de service et autres dénominations spéciales de tierces parties sont des marques de commerce ou de service de leur propriétaire respectif.

Imprimé aux États-Unis d'Amérique

Catalogué dans WorldCat le 1^{er} septembre 2011
Numéro de contrôle OCLC : 748678720

ISBN : 1-55653-436-1
978-1-55653-436-2

Table des matières

Contexte	1
Facteurs affectant la qualité de WorldCat depuis 2008	3
Ré-implémentation du logiciel DDR (Duplicate Detection and Resolution)	5
« Notices parallèles »	7
Reproductions et réimpressions	8
« Dissémination » des fonds	9
A venir : GLIMIR (Global Library Manifestation Identifier)	11
A venir : autres projets à court terme	14
A venir : qualité des données au niveau local et mondial	16
Notes de fin	18

Contexte

L'étude récente d'OCLC portant sur les attentes des utilisateurs finaux et des bibliothécaires vis-à-vis des catalogues de bibliothèque en ligne,¹ dont WorldCat, suggère que ces deux groupes utilisent les catalogues et leurs données dans un but bien précis. Les utilisateurs cherchent généralement à trouver et obtenir les informations dont ils ont besoin; quant aux bibliothécaires et au personnel de bibliothèque, ils utilisent les données de catalogue dans le cadre de leurs fonctions professionnelles. La présence d'entrées en double réelles ou perçues dans le catalogue, bien que davantage tolérée par les utilisateurs finaux que par les bibliothécaires², empêche non seulement la recherche et la livraison des informations souhaitées côté utilisateurs, mais également l'utilisation de méthodes de travail efficaces côté bibliothécaires et personnel de bibliothèque. Pour cette raison, le programme de qualité de WorldCat d'OCLC est axé depuis de nombreuses années sur la gestion des doublons.

Les données non publiées issues de l'étude d'OCLC sur les catalogues en ligne menée en 2008 fournissent des indications sur la perception et la satisfaction des membres d'OCLC vis-à-vis de la qualité de WorldCat (dont la gestion des doublons) à cette date.³ Ce sondage regroupait près de 1 400 réponses obtenues auprès de bibliothèques universitaires, publiques et spécialisées en ou hors Amérique du Nord.

Les sondés ont évalué leur satisfaction globale vis-à-vis des données de WorldCat sur une échelle de 1 à 10 (1 = médiocre, 10 = excellent). Les niveaux de satisfaction présentaient des différences selon la région (figure 1). Trois-quarts des sondés nord-américains (76 %) ont évalué leur satisfaction à 10, 9 ou 8, contre deux-tiers des sondés situés hors Amérique du Nord (65 %).

Figure 1 : Satisfaction vis-à-vis des données de WorldCat à l'intérieur et l'extérieur de l'Amérique du Nord (2008).

	Amérique du Nord (n=869)	Extérieur de l'Amérique du Nord (n=304)
10 (excellent)	20%	17%
9	24%	19%
8	32%	29%
7	15%	17%
6	5%	9%
5	2%	5%
4 ou 3	2%	2%
2 ou 1 (médiocre)	0%	0%
Je ne sais pas	0%	2%

Contexte

Les sondés ont précisé les raisons de leur satisfaction ou insatisfaction vis-à-vis des données de WorldCat. Ceux ayant évalué leur satisfaction globale à 10 ou 9 ont fait des commentaires favorables sur la couverture exhaustive des informations bibliographiques et données sur les fonds de la base de données, les fonctions d'accès et de recherche faciles à utiliser, et l'utilité de WorldCat pour répondre aux besoins des utilisateurs finaux et des bibliothécaires. Alors que ceux ayant évalué leur satisfaction à 10 ont émis peu de critiques, 32 % de ceux ayant évalué leur satisfaction à 9 ont également indiqué des problèmes, notamment la présence de doublons et l'existence de notices incomplètes (figure 2). Ces problèmes ont été signalés de manière récurrente dans l'ensemble des évaluations.

Figure 2 : Raisons de la satisfaction vis-à-vis de WorldCat		
	Pourcentage de sondés ayant évalué leur satisfaction à 10 (excellent) (n=106)	Pourcentage de sondés ayant évalué leur satisfaction à 9 (n=148)
Base de données	72 %	68 %
<ul style="list-style-type: none"> -Bon taux de réussite; identification des ressources -Quantité de notices bibliographiques -Données sur les fonds -Précision des notices bibliographiques -Informations 		
Accès	47%	43%
<ul style="list-style-type: none"> -Facilité d'utilisation -Recherche -Rapidité -Fiabilité -Présentation claire 		
Utilisation dans les services	45%	32%
<ul style="list-style-type: none"> -Adéquation avec les besoins des bibliothécaires et des utilisateurs -Catalogage -PEB/partage de ressources 		
Problèmes	3%	32%
<ul style="list-style-type: none"> -Problèmes de recherche -Notices incomplètes -Notices en double -Indisponibilité 		

Les résultats du sondage mettent plus clairement en évidence que la perception de la qualité par les utilisateurs implique davantage que la qualité des données elles-mêmes. La manière dont ces données sont exploitées et présentées par le logiciel d'interface peut être un facteur déterminant dans la création d'une expérience positive pour l'utilisateur.

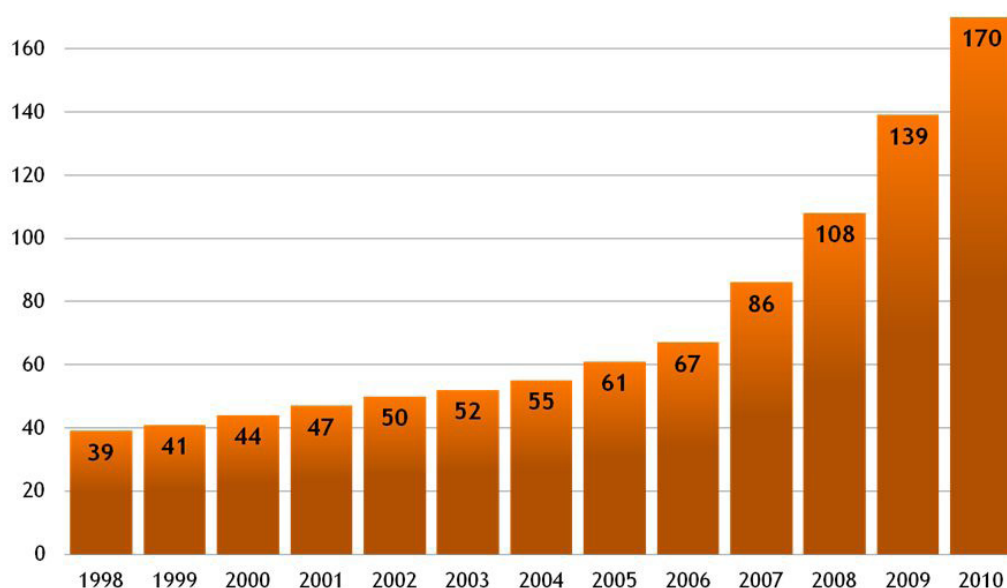
Facteurs affectant la qualité de WorldCat depuis 2008

Depuis 2008, la croissance sans précédent de WorldCat a fortement remis en question les technologies et méthodes de gestion des doublons d'OCLC (figure 3). Cette croissance a exacerbé à la fois la fréquence et la perception d'entrées en double dans WorldCat pour le même contenu d'information.⁴

Ces nouvelles métadonnées ont été introduites dans WorldCat en grande partie en raison des accords conclus avec les bibliothèques nationales et groupes hors Amérique du Nord. Lorsqu'elles ont été introduites dans WorldCat en grandes quantités, les nouvelles métadonnées :

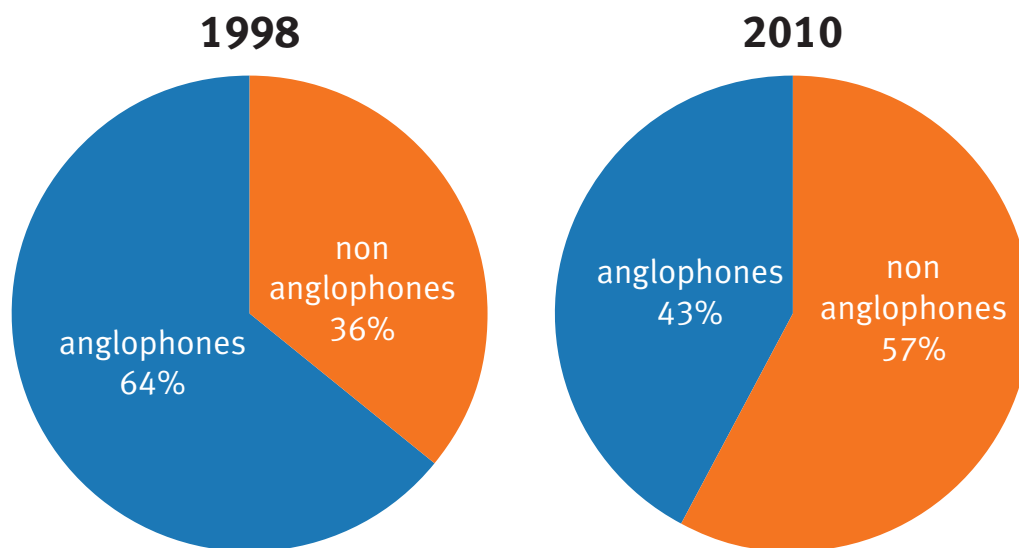
1. ont saturé les outils automatisés existants d'OCLC permettant de mettre en correspondance, fusionner et regrouper les notices d'un même ouvrage,⁵ et
2. ont créé de nouvelles opportunités de développement de services *multilingues* concernant les ouvrages, noms et terminologies en modifiant de façon significative la proportion des contenus non anglophones décrits dans WorldCat (figure 4).

Figure 3 : Croissance des données bibliographiques de WorldCat de 1998 à 2010 (en millions de notices)



Source : rapports annuels d'OCLC de l'exercice 1998 à l'exercice 2010

Figure 4 : Évolution des contenus non anglophones décrits dans WorldCat (1998 à 2010)



Source : rapports annuels d'OCLC, exercices 1998 et 2010

Un autre facteur affectant la qualité de WorldCat depuis 2008 est le chargement d'un plus grand nombre de notices de fournisseur pour les nouveaux titres et ceux à venir (on estime à l'heure actuelle que 1,59 % de la base de données bibliographique WorldCat représente des notices provenant de fournisseurs). Les bibliothèques redéfinissant les chaînes de travail de leurs services techniques et optimisant l'exploitation des métadonnées de fournisseur à des fins de sélection, d'acquisition et de traitement, OCLC a développé des partenariats visant à charger ces métadonnées dans WorldCat. Ces notices abrégées permettent d'accéder à des informations sur de nouveaux documents beaucoup plus tôt dans le processus de publication, mais du fait qu'elles contiennent bien souvent très peu de données, elles peuvent s'avérer peu utiles pour les utilisateurs finaux et les bibliothécaires. Elles posent également des problèmes concernant les processus de mise en correspondance et de fusion.

Un troisième facteur relatif à la résolution des problèmes de qualité de WorldCat est la croissance de l'activité de la communauté d'experts, bibliothèques membres d'OCLC qui partagent la maintenance de WorldCat via leur participation au programme d'amélioration d'OCLC (OCLC Enhance Program), aux programmes BIBCO et CONSER de l'initiative Program for Cooperative Cataloging, et via leur utilisation des fonctionnalités de mise à niveau et d'enrichissement des notices bibliographiques accessibles à l'ensemble des utilisateurs de catalogage. Ces fonctionnalités ont été étendues de manière significative en 2009 avec Expert Community Experiment (expérimentation menée auprès de la communauté d'experts), permettant ainsi aux utilisateurs de catalogage de modifier la quasi-totalité des notices de WorldCat. L'activité de la communauté d'experts au cours de l'exercice 2011 a atteint un peu plus d'un million de transactions, soit près du double par rapport à 2008.⁶

Ré-implémentation du logiciel DDR (Duplicate Detection and Resolution)

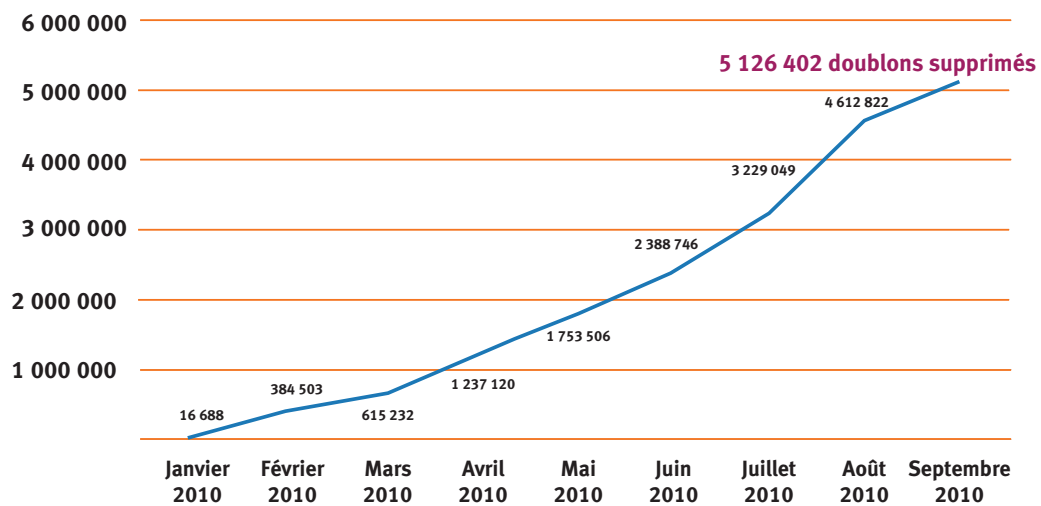
En 1991, OCLC lance son logiciel de détection et d'élimination des doublons (Duplicate Detection and Resolution, DDR), qui permet de supprimer les doublons par une mise en correspondance des notices bibliographiques de WorldCat pour le format Poche avec ses propres notices. À l'été 2005, lors de la migration de WorldCat vers sa nouvelle plate-forme, 16 analyses avaient déjà été réalisées, permettant ainsi l'élimination d'1,6 million de doublons. L'année 2005 vit également le lancement d'un projet visant à réinventer le logiciel DDR afin de l'adapter au nouvel environnement et d'étendre ses fonctionnalités non seulement aux notices de livres, mais également à celles des ressources, partitions musicales, enregistrements sonores, documents visuels, cartes et autres ressources électroniques.

Le logiciel DDR prend une notice et regroupe un ensemble d'autres notices WorldCat pouvant représenter la même ressource. Les algorithmes DDR comparent ensuite cet ensemble de notices (en utilisant des parties significatives des données dans chaque notice) afin d'identifier les doublons éventuels. Si le logiciel identifie des doublons, il sélectionne une notice à conserver (en utilisant une hiérarchie qui prend en compte la source de la notice, son exhaustivité relative, et le nombre de fonds), ainsi que les données appropriées (telles que les cotes et vedettes-matières de types qui ne sont pas encore représentés dans la notice conservée) à fusionner dans celle-ci. Tous les fonds (y compris les données détaillées sur les fonds locaux) sont déplacés vers la notice conservée, et les numéros OCLC des doublons sont enregistrés dans le champ 019 de celle-ci.

Ce projet de grande envergure échelonné sur plusieurs années a atteint un jalon en septembre 2010, lorsque le logiciel DDR a effectué l'analyse complète de WorldCat et supprimé ainsi plus de 5,1 millions de doublons (figure 5). En outre, le logiciel DDR a été intégré dans le traitement des fichiers journaux quotidiens, ainsi que dans le chargement par lots. Ces processus continuent à fusionner quotidiennement les doublons à mesure que de nouvelles notices sont ajoutées dans WorldCat. Au 30 avril 2011, plus de 7,5 millions de doublons avaient été supprimés par le logiciel DDR.

Le nouveau logiciel DDR fusionne un grand nombre de notices bibliographiques, et les bibliothèques pourront noter la diminution du nombre de doublons dans WorldCat. Néanmoins, le logiciel DDR *ne suffit pas à lui-seul* pour relever le défi et la complexité de la gestion des doublons dans la base WorldCat actuelle, une base de données multilingue et véritablement globale qui représente de nombreuses langues de catalogage, traditions de métadonnées internes ou externes aux bibliothèques, ainsi qu'une gamme croissante de catalogues individuels et collectifs du monde entier.

Figure 5 : Logiciel DDR : traitement de la base de données WorldCat (janvier à septembre 2010)

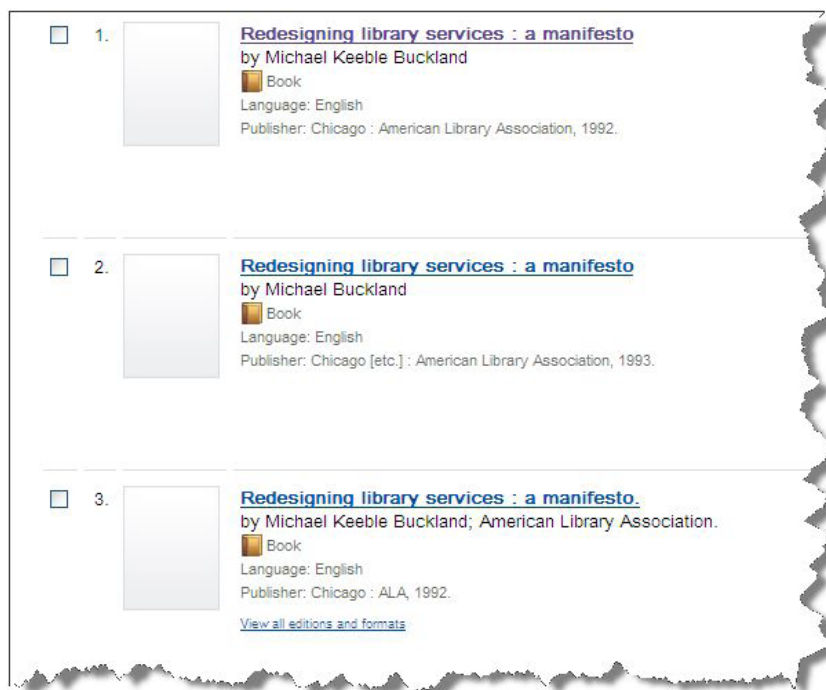


« Notices parallèles »

À la demande expresse de ses membres, OCLC a modifié sa politique en 2003 afin de permettre l'introduction de « notices parallèles » dans WorldCat par langue de catalogage (remarque : **non pas** la langue du contenu décrit, mais la langue de catalogage utilisée pour décrire ce contenu). Auparavant, les notices d'un même titre, mais cataloguées en différentes langues, telles qu'en anglais, espagnol et français, étaient considérées comme des doublons. OCLC ne les considère plus aujourd'hui comme tel, mais comme des notices parallèles (figure 6). Il a depuis longtemps été prévu que WorldCat aurait besoin d'une structure de notice parallèle pour afficher les notices par langue de catalogage, et de par son évolution, cela est maintenant devenu réalité.

La figure 6 présente les résultats de recherche pour la publication de Michael Buckland *Redesigning Library Services* dans WorldCat.org, l'interface d'OCLC pour les utilisateurs finaux. Intentionnellement, seuls quelques éléments de données sont affichés pour identifier le titre; de ce fait, les métadonnées qui révéleraient les différentes langues de catalogage de ces notices n'apparaissent pas. Cependant, les données sous-jacentes (accessibles par exemple via Connexion, l'interface de catalogage d'OCLC) permettent d'identifier les différences (voir numéros de notice OCLC 25628636, 611064627, 464816186).

Figure 6 : Exemple de notices parallèles pour le même titre cataloguées en différentes langues



Reproductions et réimpressions

Gérer l'affichage des entrées en double des ouvrages originaux et de leurs reproductions constitue un autre défi dans le cadre de la gestion des doublons dans WorldCat, une base de données qui doit répondre aux besoins de recherche et de consultation de multiples membres parmi les bibliothécaires et utilisateurs finaux. Ces deux groupes peuvent avoir des approches assez différentes concernant ce qui est ou n'est pas une entrée en double pour le même titre. À titre d'exemple, le texte original de l'édition 1993 en anglais de Philipp Lenard de *Great Men of Science*, publiée par Macmillan, a été numérisé et fait maintenant partie de la HathiTrust Digital Library. Les *règles de catalogage anglo-américaines, 2^{ème} édition*, prévoient des descriptions distinctes pour la publication originale et sa reproduction numérique, et de ce fait, OCLC ne fusionne pas les notices de ces descriptions avec le logiciel DDR (figure 7). C'est également le cas pour les réimpressions.

Même si les règles de catalogage et leurs interprétations concernant les reproductions et réimpressions sont importantes dans le cadre de la gestion appropriée des collections et des descriptions de catalogage, on peut avancer que les entrées en double affichées entraînent au mieux une certaine confusion pour l'utilisateur de WorldCat.org, et au pire l'échec du système à le guider correctement jusqu'à l'emplacement ou le lien pertinent vers le contenu.

Figure 7 : Exemple de notices de catalogage distinctes pour une publication originale et sa reproduction



Voir numéros de notice OCLC 1394475, 296526894

« Dissémination » des fonds

Même si l'agrégation des métadonnées WorldCat décrivant les collections des bibliothèques est unique, les *fonds* (emplacements des bibliothèques) constituent sans doute la valeur unique et inimitable de WorldCat. Il est important de pouvoir identifier tous les fonds pertinents pour une recherche. Pour un catalogueur, le résultat de recherche idéal peut se limiter à une édition ou un format spécifique, tandis qu'un utilisateur est susceptible de s'intéresser à n'importe quel format ou édition (quelle que soit la langue de catalogage) à sa disposition (et peut-être d'avoir une préférence pour le contenu en ligne s'il est disponible).

Les numéros OCLC, outre le fait d'être les pierres angulaires de la base de données, permettent d'établir la liaison entre les expériences de *recherche* et de *livraison* de l'utilisateur; en d'autres termes, ils permettent dans WorldCat d'établir la liaison entre les métadonnées *décrivant* un contenu spécifique et les informations sur *l'emplacement* de ce contenu dans une bibliothèque en particulier (ou sur Internet). Les numéros OCLC permettent également d'accéder directement aux notices dans WorldCat, ainsi que d'établir une liaison directe vers les métadonnées ou contenus des bases de données de catalogues externes (lorsque ces catalogues ou bases de données comportent des identificateurs OCLC intégrés et indexés).

La croissance rapide de WorldCat depuis 2008 a atténué l'efficacité des numéros OCLC permettant d'assurer la cohérence des informations sur les fonds, et ainsi la prise en charge de la visibilité des, et des liaisons vers les, ressources de bibliothèque référencées dans WorldCat. De multiples bibliographies nationales ayant été chargées, plus de notices parallèles ayant été introduites, et davantage de notices pour les reproductions (en particulier les reproductions numériques) ayant été référencées dans WorldCat, les informations sur les fonds ont été disséminées sur plusieurs notices portant des numéros OCLC différents.

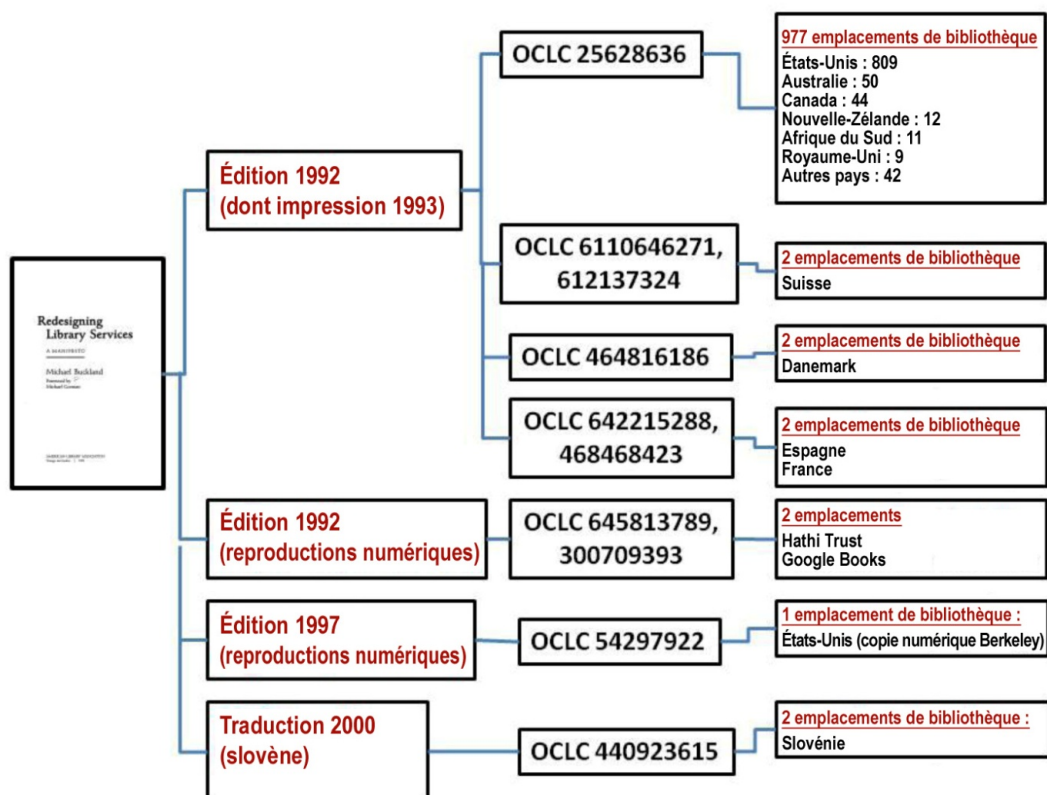
La figure 8 présente un exemple de « dissémination de fonds » affectant maintenant la visibilité et les liens vers les catalogues de bibliothèque pour la publication de Michael Buckland *Redesigning Library Services*. Telles qu'elles sont actuellement référencées dans WorldCat, les informations sur les fonds pour l'édition imprimée 1992 en anglais sont disséminées sur six numéros OCLC qui identifient plusieurs notices parallèles (différentes langues de catalogage), ainsi que sur une notice d'une impression de 1993. Deux numéros OCLC identifient les reproductions numériques de l'édition 1992 provenant de deux sources (HathiTrust et Google Books), tandis qu'un autre identifie une reproduction numérique d'une édition (censée être de) 1997. Dans presque tous les cas, les notices portant ces numéros OCLC s'affichent sous forme de résultats de recherche distincts dans les interfaces WorldCat. OCLC a mis en place le regroupement FRBR⁷ dans WorldCat.org afin de limiter la confusion. Néanmoins, le succès de l'utilisateur à trouver les fonds de bibliothèque les plus pertinents ou le lien vers le contenu souhaité repose sur la sélection de l'entrée « appropriée » dans les résultats de recherche. Par exemple, si un utilisateur nord-américain a sélectionné l'entrée représentant le numéro de notice OCLC 25628636, il accèdera à un très grand nombre de fonds de bibliothèque. La sélection de l'une des

« Dissémination » des fonds

autres entrées décrivant l'édition 1992 pourrait amener l'utilisateur à croire qu'il n'y avait pas de bibliothèque à proximité possédant l'article.

Notez que la situation actuelle affecte également les « données bibliographiques sociales ». Ces données (analyses, listes et systèmes de recommandation) sont maintenant disséminées dans plusieurs résultats de recherche. Par ailleurs, cela érode l'efficacité du partage des ressources, de l'échange de données entre bibliothèques, ainsi que de la gestion et de l'analyse collaboratives des collections.

Figure 8 : Répartition des fonds dans WorldCat pour le titre *Redesigning Library Services*



À venir : GLIMIR

(Global Library Manifestation Identifier)

Comme nous l'avons décrit précédemment, le programme DDR d'OCLC a permis de résoudre les doublons dans WorldCat. Aujourd'hui, cependant, la qualité de WorldCat doit être renforcée via la fourniture d'un identificateur unique pour plusieurs notices de métadonnées dans les cas où la redondance est non seulement **intentionnelle**, mais également en conformité avec les normes de catalogage et la politique d'OCLC concernant les notices parallèles. L'époque où l'on s'appuyait sur une seule notice originale dans WorldCat pour une publication donnée est maintenant révolue, et le temps est venu d'adopter une nouvelle approche du regroupement des métadonnées et des informations sur les fonds.

GLIMIR est un projet d'OCLC entamé en 2009 qui vise à résoudre les problèmes décrits concernant les numéros OCLC et la dissémination consécutive sur plusieurs notices des métadonnées et des informations sur les fonds. Dans ce projet, les nouveaux identificateurs GLIMIR permettent de regrouper diverses notices (décrivant des variantes linguistiques, spécifiques au format ou autres) qui décrivent le contenu d'information donné dans un groupe consolidé ayant son propre identificateur GLIMIR et ses propres numéros OCLC croisés. L'impact de GLIMIR, une fois déployé dans les diverses interfaces de WorldCat pour les utilisateurs finaux et les bibliothécaires, sera le suivant:

- Affichage des résultats de recherche plus facile à utiliser et plus intuitif pour les utilisateurs finaux et les bibliothécaires
- Réduction des coûts de traitement des acquisitions, du catalogage, de la sélection et du partage des ressources en facilitant l'identification et la sélection de la notice « correcte »
- Amélioration de la liaison entre WorldCat.org et les catalogues de bibliothèque, et flux de données uniforme dans WorldCat et les systèmes de bibliothèque, tant au niveau local que mondial
- Amélioration de la précision du regroupement FRBR dans les « ensembles d'ouvrages » WorldCat
- Amélioration de la visibilité des différences significatives parmi les occurrences des ensembles d'ouvrages (par exemple, publication originale et révisions, traductions, versions audio)
- Partage du contenu enrichi par sa mise disposition pour toutes les notices d'un groupe
- Facilitation de l'ancrage et de la divulgation des ressources sur Internet
- Fourniture d'un nouvel identificateur permettant de regrouper les autres notices dans d'autres systèmes

- Prise en charge de statistiques WorldCat de meilleure qualité permettant de déterminer le nombre de ressources uniques à la disposition des bibliothèques dans le monde

La figure 9 est une réorganisation *pro forma* des notices de l'ensemble *Redesigning Library Services* de la figure 8 (organisées sous forme d'ensemble d'ouvrages et affichées en rouge). La figure 9 présente l'étape suivante en indiquant les notices associées à l'édition 1992 et ses reproductions numériques (organisées sous forme de groupe GLIMIR et affichées en bleu). La meilleure organisation des données dans le cadre du regroupement GLIMIR assurera un affichage WorldCat plus intuitif, le regroupement des fonds associés, l'échange de données, l'agrégation et la syndication, et l'amélioration de la liaison entre WorldCat, les catalogues de bibliothèque, le contenu et autres services de données disponibles sur Internet.

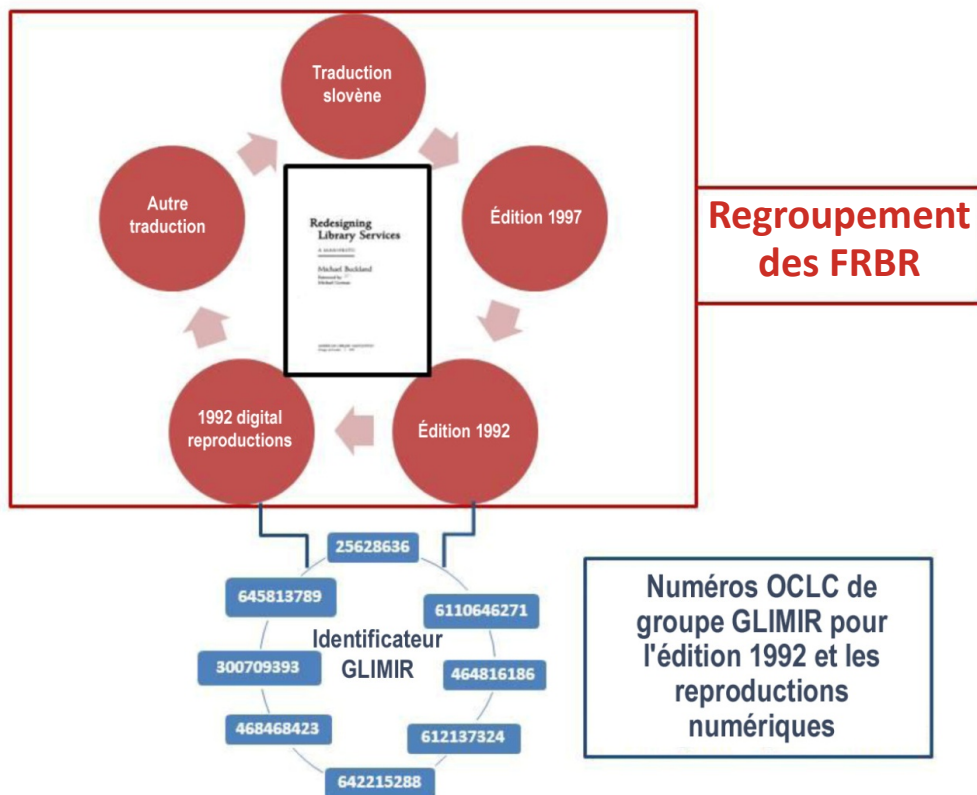
Principaux avantages de GLIMIR :

- Les groupes GLIMIR amélioreront la précision des ensembles d'ouvrages FRBR et réduiront l'affichage de notices en double strictement identiques.
- Les groupes GLIMIR amélioreront le regroupement des informations sur les fonds, en assurant une liaison plus fiable entre WorldCat et les catalogues de bibliothèque.

Un regroupement similaire améliorera également l'affichage d'autres cas dans lesquels le même contenu est présenté sous différentes formes physiques. Citons par exemple l'enregistrement sonore du même spectacle musical présenté sur disques vinyles, cassettes, CD, fichiers MP3 et streaming audio, et le même film présenté dans sa version originale ainsi que sur bande vidéo, DVD et streaming vidéo.

À la date où nous rédigeons ce document, l'infrastructure GLIMIR est presque terminée. La mise en conformité de WorldCat à cette infrastructure prendra plusieurs mois et s'achèvera début 2012 du fait de la taille importante de la base de données. Lorsqu'un nombre suffisant de groupes GLIMIR seront prêts, les utilisateurs de WorldCat.org et de WorldCat Local (interfaces utilisateurs d'OCLC) commenceront à constater les améliorations en termes de recherche, d'affichage et de liaison aux fonds, catalogues de bibliothèque et contenus en ligne. Plus tard dans l'année, les groupes GLIMIR seront intégrés dans Connexion (interface de catalogage en ligne d'OCLC) afin de réduire le temps nécessaires aux catalogueurs pour choisir la notice correcte et fournir une meilleure vue des fonds pour les occurrences d'un ouvrage. Cette tâche dans Connexion devrait également améliorer le regroupement (si ce n'est l'élimination des doublons) des notices de fournisseur incomplètes, permettant ainsi aux catalogueurs d'économiser du temps et des efforts.

Figure 9 : Présentation pro forma de l'impact de GLIMIR sur les notices WorldCat et les fonds pour la publication Redesigning Library Services



À venir : autres projets à court terme

Le tableau suivant décrit d'autres améliorations relatives à la qualité de WorldCat prévues pour 2012. Ces activités s'efforcent de trouver un équilibre entre d'une part l'amélioration des métadonnées existantes et de leur organisation, et d'autre part l'enrichissement des métadonnées et de leurs liaisons externes. En outre, le programme 2012 prévoit de commencer à résoudre les problèmes rencontrés avec WorldCat Registry et d'étendre encore davantage la communauté d'experts.

<i>Activités qualité de WorldCat prévues pour 2012</i>		
<i>Action</i>	<i>Impact</i>	<i>Commentaires</i>
Reconduire une enquête de satisfaction sur la qualité de WorldCat	Valider la pertinence et la priorité du programme de qualité des données 2012; identifier des nouvelles activités et améliorations.	Comparer les résultats de référence du sondage 2008.
Améliorer le logiciel DDR	Améliorer la flexibilité du logiciel DDR et de la mise en correspondance du chargement par lots, ainsi que leur capacité à traiter les variantes présentes dans les données entrantes dues aux différences entre les méthodes de catalogage, à un codage incorrect, etc. Introduire un nombre moindre de notices en double strictement identiques dans WorldCat.	Actuellement, le logiciel DDR est prudent et prend en compte les petites différences entre les notices. Tester et améliorer le logiciel afin de permettre une mise en correspondance plus importante lorsque cela est justifié.
Réduire les liens manquants ou rompus vers les catalogues de bibliothèque et/ou les catalogues collectifs	La recherche par le numéro OCLC associé à un ensemble de notices fusionnées établit automatiquement la liaison vers les notices correspondantes dans le catalogue externe.	Lorsque des notices sont fusionnées par le logiciel DDR, les numéros OCLC des notices supprimées sont enregistrés dans le champ 019 de la notice WorldCat conservée. Les fonds des notices fusionnées sont également déplacés vers la notice conservée. Dans la plupart des cas, les liens directs de WorldCat.org vers le catalogue d'une institution pour cette notice fusionnée échouent car l'institution dispose encore du numéro OCLC original dans son catalogue et le catalogue n'est pas « informé » du numéro OCLC de la notice conservée. Ce projet permettra d'établir une meilleure liaison entre WorldCat et le catalogue de bibliothèque à l'aide des numéros provenant du champ 019.

Activités qualité de WorldCat prévues pour 2012		
Action	Impact	Commentaires
Résoudre les problèmes relatifs aux notices de fournisseur incomplètes	Améliorer les métadonnées utilisées dans WorldCat et les systèmes de bibliothèque; permettre aux catalogueurs d'économiser du temps et des efforts.	S'engager avec des fournisseurs spécifiques à améliorer et aligner les chaînes de travail de métadonnées OCLC/fournisseur afin d'afficher plus rapidement des notices de meilleure qualité, et remplacer ainsi les notices très abrégées utilisées dans les processus de sélection et de commande, mais qui s'avèrent inadéquates dans le cadre du catalogage.
Lier et mettre à jour les vedettes spécifiques des notices WorldCat par rapport aux vedettes des fichiers d'autorité de sujet et de nom LC	Améliorer la fiabilité et la précision de la recherche et de l'affichage concernant les auteurs et sujets dans les interfaces WorldCat; améliorer la qualité des vedettes des notices exportées vers des systèmes desservant des communautés anglophones.	Les vedettes des notices seront automatiquement mises à jour, réduisant ainsi de façon significative la nécessité pour les catalogueurs de contrôler manuellement chaque vedette.
Enrichir les notices WorldCat par l'ajout de résumés, tables des matières, remarques, cotes, etc.	Améliorer la capacité des utilisateurs finaux et des bibliothécaires à évaluer l'utilité des documents identifiés dans le cadre de la recherche et à déterminer ceux qui leur semblent les plus intéressants.	Exécuter le logiciel d'enrichissement de données sur l'ensemble de la base de données WorldCat et les notices WorldCat récemment ajoutées/modifiées.
Étendre l'implication de la communauté dans la maintenance de WorldCat	Augmenter le pool de bibliothèques autorisées à modifier et corriger les notices ajoutées via l'initiative Program for Cooperative Cataloging (c'est-à-dire les notices PCC BIBCO).	Étendre le programme Expert Community en permettant aux institutions OCLC NACO de modifier les notices PCC BIBCO.
Fournir un jeu d'outils dans le cadre des activités qualité de WorldCat Registry	Améliorer l'utilité des données d'institution dans WorldCat Registry.	Assurer le reporting afin de faciliter la maintenance des données d'institution.

À venir : qualité des données au niveau local et mondial

Les projets naissants d'OCLC permettant d'améliorer en continu la qualité de WorldCat reposent sur un engagement des bibliothèques à organiser le savoir local, régional, national et mondial sur plusieurs types de supports (physiques, électroniques, numériques). Les approches d'OCLC en matière de création de métadonnées et de gestion de la qualité appliquées de longue date et performantes, mais centrées sur l'anglais, devront être adaptées aux réalités de la coopérative OCLC multinationale caractérisée par un multilinguisme croissant. Une approche exhaustive répondant aux exigences de qualité à la fois des *utilisateurs finaux et des bibliothécaires* sera incontournable pour la viabilité et l'attrait futurs de WorldCat.

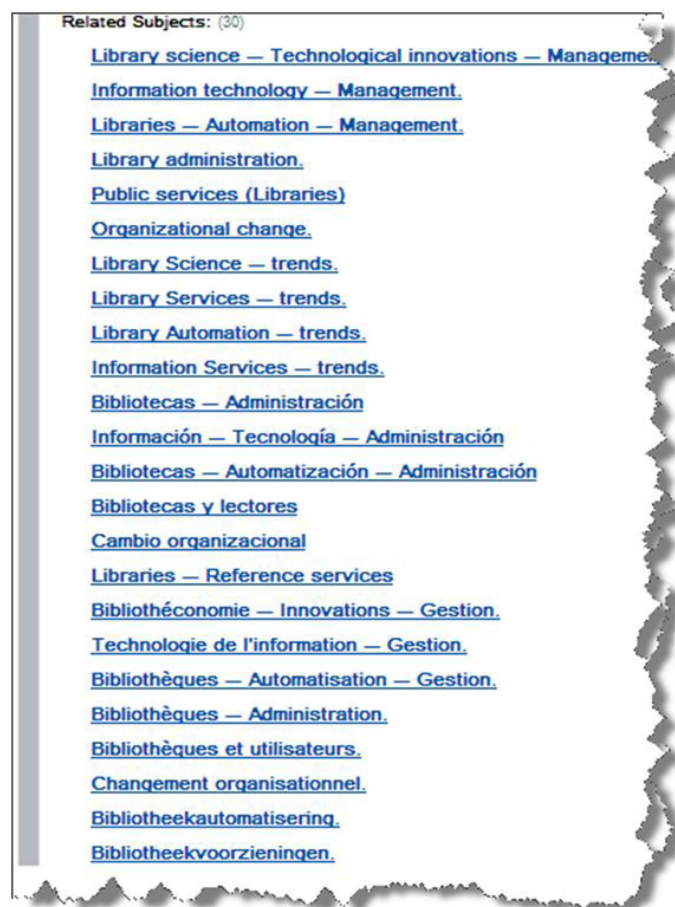
La feuille de route qualité de WorldCat au-delà des 12 prochains mois est à juste titre moins bien définie, mais inclut les principaux éléments suivants :

- Avancées supplémentaires dans la gestion des doublons au sein et hors des systèmes OCLC
- Efforts supplémentaires visant à améliorer la gestion et l'affichage des informations sur les fonds
- Extension de GLIMIR afin d'activer de nouveaux services de données à des fins d'utilisation interne et externe (cette tâche impliquera une meilleure optimisation de FRBR et de GLIMIR, ainsi que la mise à disposition des identificateurs GLIMIR afin d'accroître l'utilisation)
- Projets relatifs à la qualité de WorldCat et nouveaux services de données concernant les fichiers d'autorité et terminologiques utilisés hors Amérique du Nord
- Extension accrue de la communauté d'experts (dont davantage de participation de la part des bibliothèques hors Amérique du Nord)
- Déploiement de données multilingues à partir de la 23^{ème} édition du système de classification décimale Dewey (DDC) et des ensembles terminologiques associés des nouveaux services Web accessibles de l'extérieur
- Gestion dynamique des descriptions de notices en différentes langues à l'aide d'algorithmes, du système de classification décimale Dewey et des fichiers d'autorité

Les difficultés croissantes associées à l'augmentation de la pertinence mondiale de WorldCat depuis 2008 s'accompagnent de nouvelles opportunités. Comme nous l'avons mentionné précédemment dans ce document, la croissance rapide de WorldCat a modifié de façon significative la proportion des métadonnées et des

contenus non anglophones représentés dans la base de données, créant ainsi de nouvelles opportunités de développement de services *multilingues* concernant les ouvrages, noms et terminologies. La consultation d'une liste partielle des vedettes-matières linguistiquement différentes associées au titre pris en exemple dans ce document (*Redesigning Library Services*) révèle la présence de données multilingues nécessaires (figure 10). Par ailleurs, le développement de WorldCat Identities et de VIAF⁸ a permis de capturer des formes linguistiques différentes des noms d'auteur et de créateur. Il s'agit maintenant de déployer ces données linguistiquement différentes dans un programme qualité WorldCat coordonné afin de prendre en charge les services Web localisés des utilisateurs finaux et des bibliothécaires dans lesquels non seulement l'interface, mais les données elles-mêmes sont proposées dans la langue souhaitée.

Figure 10 : Vedettes-matières linguistiquement différentes décrivant les sujets de la publication *Redesigning Library Services*



Les approches d'OCLC en matière de création de métadonnées et de gestion de la qualité appliquées de longue date et performantes, mais centrées sur l'anglais, devront être adaptées aux réalités de la coopérative OCLC multinationale caractérisée par un multilinguisme croissant.

Notes de fin

1. Calhoun, Karen et Diane Cellentani. 2009. *Catalogues en ligne : les souhaits des utilisateurs et des bibliothécaires : rapport d'OCLC*. Dublin, Ohio : OCLC.

2. Calhoun, Karen et Diane Cellentani. 2009. *Catalogues en ligne : les souhaits des utilisateurs et des bibliothécaires : rapport d'OCLC* : document synoptique. Dublin, Ohio : OCLC, page 14. Disponible à l'adresse : <http://www.oclc.org/reports/onlinecatalogs/>.

3. Données non publiées préparées pour *Catalogues en ligne : les souhaits des utilisateurs et des bibliothécaires*. 2008.

4. Ce livre blanc porte sur la base de données de notices bibliographiques et de fonds bibliothécaires de WorldCat qui sont en majeure partie saisis par les membres d'OCLC, bibliothèques nationales et consortiums, ainsi que certains éditeurs et fournisseurs. La croissance rapide du nombre de notices de métadonnées des articles, livres électroniques, contenus numérisés du domaine public ou d'autres contenus numériques qui sont indexés pour la recherche via WorldCat.org n'est pas abordée dans ce document. Il faut admettre que l'expansion de WorldCat.org afin d'inclure ces métadonnées a contribué à la perception que celui-ci est « parasité ». L'équipe chargée de la qualité de WorldCat d'OCLC est informée des problèmes de recherche, liaison et livraison associés à ces nouvelles métadonnées dans WorldCat.org, mais ceux-ci ne sont pas traités dans ce document.

5. FRBR ([Functional Requirements for Bibliographic Records](#)) est une recommandation de 1998 de l'IFLA (International Federation of Library Associations and Institutions) visant à restructurer les bases de données de catalogues afin de prendre en compte la structure conceptuelle des ressources d'information. Le modèle FRBR regroupe les notices bibliographiques intellectuellement liées sous forme d'« ouvrages ». Le regroupement des ressources sous l'égide « ouvrages » permet aux utilisateurs de passer en revue la multitude de ressources disponibles sous forme de textes divers, copies numérisées, accords, traductions, etc. Ce regroupement leur permet d'obtenir l'ouvrage, ou le contenu, qu'ils recherchent, indépendamment du « conteneur » dans lequel il se trouve. Dans les bases de données de grande taille telles que WorldCat, ce type de collocation est indispensable pour la recherche, la navigation et le traitement rentable des bibliothèques. La conformité de WorldCat.org à FRBR a été assurée par l'algorithme FRBR Work-Set (ensemble d'ouvrages FRBR) développé par OCLC.

6. Pour plus d'informations sur la communauté d'experts, le programme d'amélioration d'OCLC (OCLC Enhance Program) et autres fonctionnalités de mise à jour et d'enrichissement, consultez le site

<http://www.oclc.org/us/en/worldcat/catalog/quality/>. Pour plus d'informations sur l'initiative Program for Cooperative Cataloging et les programmes BIBCO et CONSER, consultez le site <http://www.loc.gov/catdir/pcc/>.

7. L'algorithme FRBR Work-Set développé par OCLC regroupe les notices WorldCat en fonction d'informations sur l'auteur et le titre provenant des métadonnées décrivant des auteurs et titres spécifiques. Les noms d'auteur et les titres sont normalisés de façon à créer une clé. Toutes les notices ayant la même clé sont regroupées dans un « ensemble d'ouvrages FRBR ».

8. WorldCat Identities compile les données explorées à partir de WorldCat et d'autres ressources afin de présenter les personnes physiques et morales référencées dans WorldCat. Sont incluses des informations telles que les variantes d'appellation sous lesquelles ces personnes physiques ou morales sont connues, un historique de publication concernant les ouvrages rédigés par ou sur elles, leurs rôles par rapport aux ouvrages, etc. Chaque personne physique ou morale, le nombre total étant actuellement de près de 30 millions, est représentée par une page d'identité distincte. Ces pages d'identité sont intégrées dans l'interface WorldCat.org et sont également disponibles à l'adresse <http://worldcat.org/identities>. VIAF (Virtual International Authority File) est un projet commun regroupant la Library of Congress (bibliothèque du Congrès des États-Unis), la Deutsche Nationalbibliothek (bibliothèque nationale allemande) et la Bibliothèque nationale de France, et mené en collaboration avec un nombre croissant d'autres bibliothèques nationales et organismes. Il combine les fichiers d'autorité de nom de ces organisations participantes en un service d'autorité de nom unique (<http://viaf.org/>).



6565 Kilgour Place
Dublin, Ohio 43017-3395
1-800-848-5878 +1-614-764-6000
Fax: +1-614-764-6096
www.oclc.org

ISO 9001 Certified

ISBN: 1-55653-436-1
978-1-55653-436-2