



World Library and Information Congress: 69th IFLA General Conference and Council

1-9 August 2003, Berlin

Code Number: 120-F
Meeting: 111. Document Delivery and Interlending & Cataloguing
Simultaneous Interpretation: Yes

Avantages bibliographiques d'un catalogue collectif centralisé pour le PEB et le partage des ressources

Philip Hider

Bibliographic Manager
Singapore Integrated Library Automation Services (SILAS)
National Library Board
Singapore

Abstract :

Cet article examine quelques-uns des avantages bibliographiques d'un catalogue collectif à base de données centralisée par rapport à un catalogue collectif distribué ou « virtuel ». Il peut être bon de prendre ces avantages en considération lorsqu'un réseau de bibliothèques doit choisir entre l'un et l'autre modèle pour établir un service de fourniture de documents. La nature de ces avantages bibliographiques est discutée dans le contexte de la fourniture de documents entre bibliothèques, puisque ce sont là les circonstances qui les produisent et les rendent plus significatifs. On donne un compte rendu d'une étude succincte du fossé qui s'est creusé entre les catalogues de deux grandes bibliothèques de Singapour à la suite de l'adoption d'un modèle distribué. Cela a montré que le contenu bibliographique d'un catalogue collectif distribué peut être sensiblement plus pauvre que celui d'une base de données centralisée, notamment en termes d'une augmentation : (1) des doublons ; (2) de l'incohérence ; (3) des erreurs ; (4) des oublis. Il y a au moins quatre raisons principales qui expliquent ce fait, puisque dans un système centralisé : (a) la duplication des notices ne peut être imputable qu'à une erreur d'un catalogueur ; (b) il est plus facile de mettre en œuvre des mécanismes de contrôle de qualité ; (c) il y a des chances pour que les catalogueurs fassent plus d'efforts pour se plier à des normes et des principes consensuels ; (d) les notices sont plus susceptibles d'être corrigées et améliorées par le biais des autres bibliothèques.

Introduction

Les catalogues collectifs ont toujours été un outil très profitable au PEB et à la fourniture de documents. On y trouve « tout sous la main », pour parler familièrement, avec un accès simultané à de multiples collections, que ces collections soient matérielles, virtuelles, ou les

deux. Il importe de garder à l'esprit que les catalogues collectifs ne doivent pas seulement servir comme « outils » pour répondre à des demandes de PEB et de fourniture de documents : ils peuvent aussi servir à formuler des demandes de fourniture de documents, et c'est souvent le cas.

Si une bibliothèque vise à ce que son service de fourniture de documents fonctionne à plein, un moyen important d'y parvenir, c'est de faire en sorte que le public ait accès au catalogue collectif, afin que non seulement le personnel de la bibliothèque, mais aussi l'utilisateur final puisse identifier tous les PEB potentiels et autres transactions entre bibliothèques plus commodément accessibles. Certes, si la demande dépasse les possibilités de l'offre, un moyen de la juguler serait de limiter délibérément l'accès à l'information bibliographique, mais beaucoup considèrent une telle politique comme erronée, car dans le contexte de la plupart des services de fourniture de documents il n'est pas nécessaire que l'offre coïncide avec la demande : les PEB n'ont pas besoin d'être encouragés ; la fourniture en ligne de documents peut être refusée en suivant un algorithme automatisé. S'il faut établir des priorités entre les requêtes, dans la plupart des cas il est plus satisfaisant de faire appel à d'autres procédés que la limitation de l'accès aux catalogues de bibliothèques et autres bases de données.

Il est à noter toutefois qu'il y a encore des catalogues collectifs qui pourraient être rendus accessibles au public mais qui ne le sont pas, alors même que les bibliothèques qui y sont représentées proposent des services de fourniture de documents entre bibliothèques. De nombreuses raisons expliquent ce fait, notamment l'angoisse de finir par crouler sous les demandes de PEB. D'un autre côté, il semble que la proportion de catalogues collectifs accessibles au public ait tendance à augmenter, avec pour beaucoup d'entre eux un accès gratuit via Internet.

L'auteur du présent article part de l'hypothèse que les services de PEB et autres types de fourniture de documents souhaitent globalement encourager les utilisateurs à formuler des requêtes potentiellement valides, et considèrent le catalogue collectif comme un outil très précieux pour y parvenir. La question des avantages bibliographiques qu'un type de catalogue collectif présente par rapport à un autre demeure même si le catalogue collectif est vu sous un autre angle, en tant que simple outil permettant de retrouver les ressources demandées. Mais cette question revêt toute son importance dès lors que le catalogue collectif sert à l'utilisateur final à identifier les ressources qu'il cherche à obtenir, *avant même* qu'il n'accède à la banque de PEB.

Catalogues collectifs centralisés et catalogues collectifs distribués

Il y a plusieurs manières de concevoir un catalogue collectif. La manière habituelle consiste à établir une base de données centralisée qu'il est possible d'alimenter directement (en créant des notices dans la base centrale) ou indirectement (en créant des notices dans un catalogue local d'abord, puis en les reversant dans le catalogue collectif), ou les deux. Des protocoles tels que Z 39.50 ouvrent une autre possibilité, consistant à établir un catalogue collectif *distribué*. Dans ce modèle, les catalogues locaux sont reliés par leurs serveurs respectifs, probablement par le biais de Z 39.50, pendant que le client Z 39.50 recherche dans chacun des catalogues tour à tour, créant « à la volée » un catalogue collectif virtuel. Le concept de catalogue collectif distribué est devenu très populaire dans certains milieux bibliothéconomiques ces dernières années. Dans certaines situations, les bibliothèques ont tendance à privilégier ce type de catalogue, dans la mesure où le schéma traditionnel mettant en œuvre une base de données centrale peut présenter un inconvénient majeur : le flux de

catalogage peut être ralenti par la nécessité d'alimenter le catalogue collectif en notices, par exemple.

Ceux qui plaident en faveur d'un catalogue distribué mettent en avant la capacité de ce type de système à produire en temps réel des informations *enrichies* relatives aux exemplaires. C'est-à-dire, non seulement des informations quant à la présence d'un exemplaire dans les collections de telle ou telle bibliothèque, mais aussi quant à sa disponibilité. Mais en réalité il est tout à fait possible, par le biais de recherches supplémentaires via Z39.50, etc., de fournir en temps réel ce type d'informations dans un catalogue collectif reposant sur une base de données centralisée, même si de tels systèmes *hybrides* qui associent des éléments propres aux systèmes distribués (les informations enrichies sur les exemplaires) et aux systèmes centralisés (les données bibliographiques et les informations sommaires sur les exemplaires) n'ont guère encore été implémentés.

Puisqu'il est possible de réaliser des catalogues collectifs centralisés disposant d'outils de PEB et de fourniture de documents, peut-être au prix de nouvelles fonctionnalités Z39.50, la préférence pour les systèmes distribués plutôt que centralisés devrait se justifier par la commodité qu'ils peuvent représenter pour les bibliothèques qui les utilisent ainsi que par le contexte qui prévaut dans l'instance qui les gère (telle qu'une bibliothèque nationale). Mais un système centralisé n'est pas forcément désavantageux pour toutes les bibliothèques ; en outre, les serveurs Z39.50 nécessités par un catalogue collectif distribué demeurent coûteux — trop coûteux pour de petites structures, voire pour que les bibliothèques nationales ou les institutions centrales en fassent l'acquisition pour le compte de toutes les bibliothèques participant au réseau.

Néanmoins, de tels serveurs Z39.50 ne demeureront sans doute pas éternellement coûteux, et au fur et à mesure qu'ils deviendront plus abordables, il faudra se poser la question : y a-t-il d'autres raisons pour préférer le modèle centralisé ou hybride à base de données centralisée à un modèle complètement distribué ? Le présent article passe en revue quelques-unes des raisons *bibliographiques* allant dans ce sens, à savoir que la recherche peut être favorisée dans une base de données centralisée par rapport à un catalogue collectif virtuel, en raison de différences dans le contenu bibliographique. Ce type de recherche intervient notamment dans le contexte du PEB et du partage des ressources, bien qu'il puisse aussi s'agir d'autres utilisations d'un catalogue collectif (le renseignement au public par exemple). Bien que divers autres paramètres — techniques, financiers, politiques, par exemple — aient fait l'objet de mainte étude, ce n'est pas le cas de l'effet respectif qu'a chaque modèle sur les données bibliographiques elles-mêmes.

Il faut garder présent à l'esprit qu'outre les avantages bibliographiques possibles, d'autres raisons peuvent expliquer pourquoi les *utilisateurs* pourraient préférer un modèle centralisé ou hybride ; par exemple : d'énormes systèmes distribués peuvent avoir des temps de réponse extrêmement longs. Il faut noter qu'à l'heure actuelle nombre de clients et de serveurs Z39.50 n'offrent pas encore les fonctionnalités de recherche élaborée dont on peut disposer sur beaucoup de catalogues. Coyle (2000) en souligne le réquisit de base : « Pour qu'un catalogue collectif virtuel soit réalisable, les bases de données qui y participent doivent offrir un ensemble uniforme d'index et de fonctions de recherche qui puissent retrouver dans chaque catalogue des éléments comparables. »

Catalogues collectifs et outils bibliographiques

Ce n'est que récemment que les catalogues collectifs distribués sont devenus réalisables, grâce à l'usage répandu de Z39.50. Auparavant, les catalogues collectifs dépendaient de la création d'une base de données bibliographiques centralisée. Mais de telles bases de données centralisées n'étaient principalement créées qu'à des fins de fourniture de notices, c'est-à-dire pour jeter les bases d'un outil bibliographique ; leur fonction de catalogue collectif n'était que d'importance secondaire. Avec l'essor d'Internet et les progrès des télécommunications au cours des dix dernières années, l'avenir d'un grand nombre d'outils bibliographiques locaux et même nationaux est menacé par le nombre croissant de sources alternatives de notices bibliographiques. Pour cette raison même, peut-être, ou du moins en partie, la fonction de catalogue collectif de beaucoup de ces bases de données centralisées a été mise en avant, ainsi que les autres services reposant sur cette fonction — parmi lesquels, bien sûr, le PEB et la fourniture de documents.

Il y a donc deux raisons pragmatiques de préférer les systèmes centralisés aux systèmes distribués : tout d'abord, il en existe déjà beaucoup, avec des bases de données souvent importantes et alimentées depuis de nombreuses années ; ensuite, bien des agences centrales de réseaux bibliographiques ont déjà investi dans le développement de services de PEB et de fourniture de documents reposant sur leur base de données centralisée, et n'ont peut-être pas envie de voir ces services remplacés.

La plupart des réseaux bibliographiques nationaux passés en revue par Hider (2002) ne semblent pas encore vouloir passer à un modèle distribué, bien que beaucoup d'entre eux aient élaboré ou soient en train d'élaborer une plate-forme automatisée de fourniture de document au dessus de leur base de données centralisée (par exemple, DanBib au Danemark, Te Puna en Nouvelle-Zélande, Kinetica en Australie, LIBRIS en Suède, CCB en Belgique). Certains réseaux nationaux sont suffisamment coordonnés (et modestes) pour avoir évacué le dilemme « centralisé ou distribué ? » en partageant le même SIGB (par exemple, ELINET en Estonie et COBISS en Slovénie).

Dans certaines descriptions de ces réseaux nationaux, on souligne l'importance de maintenir la cohérence interne d'une base de données centralisée. Par exemple, sur le site de COBISS on peut lire :

Un catalogueur vérifie d'abord si la notice bibliographique qu'il veut ajouter dans la base de données locale existe déjà dans la base de données collective COBIB. Si oui, il la décharge dans la base de données locale, et ajoute les données d'exemplaires spécifiques. Si non, il crée la notice dans la base de données locale, d'où elle est ensuite automatiquement transférée dans le catalogue collectif. Le catalogueur n'est pas habilité à modifier les notices dans la base de données collectives COBIB si elles n'ont pas été créées par sa propre bibliothèque ou s'il ne dispose pas d'une habilitation spéciale.

La qualité, l'uniformité et la cohérence des bases de données locales et de la base de données collective COBIB sont assurées de différentes manières : contrôles de doublons, contrôles du programme COMARC, travail sur les notices, utilisation de référentiels communs pour toutes les données codées (pays, langues, CDU...), de référentiels locaux visant à l'uniformisation des données au sein d'une bibliothèque (par exemple, localisations, désignations

internes), de compteurs automatiques (numéros d'inventaires, cotes...), contrôles sur l'identification univoque des publications en série, etc., et surtout formation systématique des créateurs de notices avant de leur décerner l'habilitation de cataloguer...

On a fait très attention à l'uniformité dans la saisie et l'affichage local des données à destination de l'utilisateur final. Cette uniformité a été atteinte en grande partie par le format COMARC/H pour les données d'exemplaires. En outre, avant d'intégrer toute bibliothèque au véritable environnement de catalogage partagé, le Service d'information des bibliothèques et la bibliothèque concernée passent un accord sur la manière de saisir les données locales, qui est adaptée afin d'aboutir à une présentation compréhensible par l'utilisateur final tout en préservant la spécificité de l'organisation du travail dans chaque bibliothèque.

Il faut toutefois garder présent à l'esprit qu'il existe beaucoup d'autres catalogues collectifs, réels ou potentiels, en dehors des catalogues nationaux, et qui regroupent souvent des bibliothèques appartenant à un même type, et que beaucoup d'entre eux n'héritent pas d'une base de données centralisée bien établie, ni d'une agence centrale qui puisse s'en occuper. Même ceux pour qui c'est le cas, comme COPAC, le catalogue collectif du CURL (l'influent réseau de catalogues des bibliothèques universitaires de Grande-Bretagne), ont récemment envisagé de passer à un modèle distribué (Crossnet Systems Ltd. 2000).

Avant que nous n'abordions les raisons bibliographiques de préférer les systèmes centralisés ou distribués, nous devrions rappeler le point suivant : là où une base de données centralisée existe déjà, plus elle est volumineuse et soigneusement élaborée, plus elle offre de bénéfices bibliographiques pertinents ; à l'inverse, là où on a laissé une base de données se détériorer, en terme de qualité bibliographique et de couverture, moins elle offre de bénéfices immédiatement pertinents. En d'autres termes, les bénéfices bibliographiques d'une base de données centralisée soulignés dans le présent article augmentent avec le temps. Les systèmes centralisés qui ont démarré dans les années quatre-vingts et dont on assure toujours la maintenance ont donc un sérieux avantage sur les systèmes centralisés créés aujourd'hui.

Risques de différences bibliographiques entre les deux modèles

Risques de doublons

La différence la plus évidente entre contenu bibliographique de bases de données centralisées et contenu bibliographique de catalogues collectifs virtuels porte sur les doublons. En règle générale, il est plus difficile de produire des notices uniques dans un catalogue collectif distribué. Même si un système distribué comporte un programme sophistiqué de dédoublonnage qui « traque » les doublons selon un algorithme au fur et à mesure qu'ils arrivent des serveurs Z39.50 des bibliothèques, il est peu probable que ce programme les « traque » absolument tous.

Alors même qu'un catalogue collectif distribué permet en fait également de copier des notices comme dans une base de données, si chacune des bibliothèques est équipée d'un client Z et d'un serveur Z, il est plus probable de voir apparaître des notices dissemblables dans les catalogues locaux, et plus souvent, que dans une base de données centralisée, ce qui augmente le risque d'aboutir à un échec dans le programme de dédoublonnage du système (s'il en existe

un). Cela tient au fait que les notices peuvent être modifiées dans le catalogue local, après avoir été déchargées du catalogue d'une autre bibliothèque.

On peut concevoir qu'un système distribué très sophistiqué réduise ce problème en introduisant un numéro de contrôle aléatoire dans les notices qu'il décharge d'un autre catalogue (si ce n'est pas déjà le cas), à la fois dans l'original et la copie, mais cela nécessite un accès éditorial au catalogue hôte, au moins en termes de champ du numéro de contrôle. Il faudrait en outre pouvoir distinguer entre catalogage par copie de notice et par « clonage » (c'est-à-dire en copiant une notice pour pouvoir la modifier en une autre notice pour un document bibliographiquement similaire), de sorte qu'un numéro de contrôle ne soit pas généré dans le second cas. Même si un numéro de contrôle aléatoire suffisamment long serait et demeurerait probablement unique, le risque de doublon existerait encore.

Même si ce type de système sophistiqué était développé, il y a une autre cause potentielle de doublon, qui n'existerait sans doute pas dans le modèle de base de données centralisée : lorsqu'un ou plusieurs serveurs bibliothéconomiques sont « plantés », le catalogueur risque de continuer quand même et de supposer que la notice n'existe pas déjà sur aucun des autres catalogues (il pourrait ne même pas se rendre compte qu'un des 12 ou 24 serveurs n'est plus accessible).

Le danger de fusionner à tort deux notices dans un programme de dédoublonnage (c'est-à-dire, d'écraser des notices qui ne représentent pas en fait le même document, par exemple deux documents de format différent) est également plus grand. Même si cela peut parfois se produire dans une base de données centralisée, un programme de dédoublonnage qui ne tourne que sur les champs titre et numéro normalisé écraserait souvent des notices correspondant à des œuvres différentes, sans parler des diverses expressions et manifestations d'une œuvre. (Le fait d'écraser des notices de diverses expressions et manifestations à un premier niveau d'affichage pourrait présenter un certain intérêt, mais ce ne serait sûrement pas le cas si l'on écrasait les notices correspondant à des œuvres différentes).

Il faut noter que les doublons peuvent également se produire dans une base de données centralisée si de nouvelles notices y sont chargées après avoir été créées localement, en raison du laps de temps qui s'écoule entre la création et le chargement (une autre bibliothèque risque de créer sa propre notice pour le même document pendant cette période).

Cependant, tant que chaque bibliothèque utilise la base de données centralisée pour tout son catalogage par copie de notice (comme c'est normalement le cas), et tant que de tels laps de temps ne se prolongent pas sur des mois et des années, ce flux ne devrait pas produire le type de doublon qu'on est susceptible de rencontrer dans les systèmes distribués.

Conséquences de la présence de doublons

Si nous admettons que dans bien des cas, un catalogue centralisé risque moins de produire des doublons qu'un catalogue collectif virtuel, la question à se poser ensuite est : est-ce que cela a une importance ? Cela dépend de la recherche spécifique qu'un utilisateur est en train de lancer, du nombre de catalogues représentés, et du degré de chevauchement des fonds.

Au cours d'une recherche sur un document déjà connu, il se peut qu'un petit nombre seulement de notices-doublons soient affichées (d'abord en format abrégé), et ces doublons peuvent être présentés dans un ordre de priorité fixé par la bibliothèque détentrice. Mais il

n'arrive pas toujours que même dans ce type de recherche on aboutisse à des notices ne correspondant qu'à un seul document. En fait, c'est même plutôt rare. On pourrait lancer une recherche sur un document déjà connu sur le critère d'un numéro normalisé, mais pas pour tous les types de documents, notamment pas pour l'audiovisuel, qui n'a pas de numéros normalisés. Même avec un numéro normalisé, on peut trouver des notices pour un document complet et pour ses parties, des notices avec un numéro normalisé erroné, etc. En outre, dans bien des cas, l'utilisateur d'un service de fourniture de documents ne connaît pas le numéro normalisé du document qu'il cherche. Un document déjà connu est souvent recherché, bien sûr, par son titre et/ou son auteur. Le risque de doublon est ici encore plus grand, étant donné le caractère non univoque des titres et des noms.

Nous avons déjà dit plus haut que les catalogues collectifs peuvent jouer un rôle en amont du processus de fourniture de document, c'est-à-dire qu'ils peuvent procurer aux utilisateurs (intermédiaires et finaux) un moyen d'identifier les documents qu'ils peuvent vouloir demander. De ce point de vue, ils servent fréquemment à lancer des recherches sur des documents non connus par avance, par sujet, par auteur, etc. Dans ce domaine de la recherche par sujet et des autres types de recherche sur des documents non connus par avance, le risque d'aboutir à de longues listes de réponses augmente et, parallèlement, le risque de doublons.

Lorsqu'un utilisateur doit passer en revue la totalité d'une liste de réponses, un grand nombre de doublons ne compromet certes pas l'issue heureuse de sa recherche, mais présente un caractère agaçant. Mais en réalité les utilisateurs ont tendance à ne regarder que les notices figurant sur la première ou les deux premières pages. Le problème, bien sûr, c'est que les doublons ajoutent du « bruit » et réduisent le nombre de documents pertinents susceptibles d'être trouvés dans cette page ou ces deux pages de résultats.

Bruit et doublons augmentent également au fur et à mesure que se multiplient les catalogues représentés dans un catalogue collectif virtuel, et que croît le taux d'exemplaires par notice. Dans le Catalogue Collectif National de Singapour il y a actuellement environ 3 millions de mentions d'exemplaires associées à environ 2 millions de notices bibliographiques, pour plus de 50 bibliothèques représentées. Le nombre d'exemplaires sur un échantillon aléatoire de 1000 notices ajoutées au catalogue collectif depuis un à trois ans a été extrait de la base de données pour en montrer la répartition (Tableau 1). On observe une approximation de la fonction de Zipf, bien connue.

Exemplaires par notice	
1	770
2	117
3	52
4	35
5	7
6	12
7	4
8	2
9	0
...	0
22	1
1000	

Tableau 1 : Répartition des exemplaires

Si une recherche très précise sur un document déjà connu ne débouche que sur les notices qui correspondent au document recherché, alors les doublons ne posent pas trop de problèmes. En revanche, si une recherche débouche sur des notices correspondant à plus d'un document, alors les doublons peuvent être gênants. Et ce même dans le cas d'une recherche portant sur un document déjà connu. Par exemple, si une recherche auteur/titre aboutit à une liste de notices correspondant à plusieurs documents différents, voire peut-être à des œuvres toutes différentes, hypothèse tout à fait plausible si les mots de la requête sont passablement courants et que le catalogue collectif est globalement très volumineux, alors l'utilisateur manquera quelque chose s'il omet de passer au second écran et que le premier écran ne montre que d'autres documents pour lesquels il y a plusieurs notices (par exemple quatre documents représentés par 3+3+2+2 notices).

Dans le cas des recherches par sujet, les doublons sont potentiellement une nuisance plus grande encore. Et ce parce qu'un plus grand nombre de documents va souvent être trouvé, et qu'un plus grand nombre de ces documents risque d'être demandé via le service de fourniture de documents. Un utilisateur peut manquer quelque chose s'il ne regarde que les notices abrégées du premier écran, par exemple, et les risques de doublons augmentent cette possibilité, s'il y a des notices pour plus de dix documents trouvés. On peut évaluer ce risque par le biais de la répartition trouvée dans l'échantillon ci-dessus, en supposant qu'il n'y a pas de dédoublement, et que tous les doublons sont trouvés et affichés ensemble, à :

$$p \approx 1 - (0,77^{10} + 0,77^9 \cdot (1-0,77)) = 0,905$$

Niveau de cohérence

Le nombre de doublons ne constitue pas la seule différence bibliographique entre catalogues collectifs distribués et centralisés. Il y a une autre différence, importante, qui est seulement probable dans certaines circonstances et presque certaine dans d'autres. Il s'agit de la différence dans le niveau de cohérence. Bien qu'en théorie, même dans un système distribué, les bibliothèques puissent adhérer aux mêmes normes et politiques bibliographiques, la cohérence bibliographique peut être réduite de deux manières dans le cas d'un système distribué.

D'abord, même avec un catalogue collectif virtuel à portée de main, les catalogueurs auront moins tendance à créer de nouvelles notices dans le contexte de ce catalogue collectif que s'ils cataloguaient dans une base de données centralisée, ou même que s'ils créaient leurs notices au niveau local mais pouvaient disposer d'une base de données centralisée. Cela s'explique par plusieurs raisons : il est probablement infiniment plus long d'interroger un catalogue collectif virtuel que d'interroger une base de données centralisée ; il n'y a pas de fichier d'autorité contrairement à ce que peut proposer une base de données centralisée ; les catalogueurs ne sont peut-être pas capables de « cloner » des notices similaires provenant d'autres bibliothèques s'ils ne sont pas capables de télécharger des notices du catalogue collectif virtuel (il leur faut leur propre client Z connecté à leur système local).

Ensuite, dans un système centralisé on peut veiller de plus près à ce que les catalogueurs adhèrent à des normes et à des polices prédéfinies. En fait un système centralisé permet de rendre obligatoire la relecture des notices, ce qui peut être particulièrement intéressant pour de petites bibliothèques sans équipe de catalogueurs à plein temps.

Il faut noter que les normes et politiques de catalogage changent en tout état de cause, et que leur application varie au sein d'un catalogue, mais néanmoins un catalogue collectif centralisé

est susceptible de *moins* d'incohérence qu'un catalogue collectif distribué pour les raisons invoquées ci-dessus, et cette différence pourrait être substantielle.

Conséquences de l'incohérence

Il existe au moins quatre types d'incohérence bibliographique, qui sont dus à :

1. une erreur de la part d'un ou plusieurs catalogueurs,
2. des applications différentes des normes ou des politiques de la part des catalogueurs,
3. des interprétations différentes des normes ou des politiques de la part des catalogueurs,
4. des interprétations différentes des informations trouvées sur le document.

Le premier type d'incohérence, nous l'aborderons dans les sections suivantes, qui ont trait aux différences dans la qualité des notices. Les quatre types d'erreurs sont exacerbés dans un système distribué.

Les trois derniers types d'incohérence affectent tous la recherche, mais les types 2 et 3 affectent aussi la *lecture* d'une notice une fois qu'elle a été trouvée, et par voie de conséquence peuvent déboucher sur une demande incongrue de PEB, ou une absence de demande de PEB alors que cela aurait été pertinent.

Si les divergences dans l'interprétation des normes et des politiques sont moins probables quand les catalogueurs peuvent se référer à une base de données partagée, on peut généralement supposer que ces divergences n'ont en tout état de cause que peu de conséquences pour autant qu'il s'agit de la description du document — sans quoi elles seraient, ou devraient être, traitées dans une révision desdites normes et politiques. Par exemple, un code de catalogage peut se trouver ne pas définir un type d'illustrations qui devrait être signalé si on le trouve sur la publication — si ce point est considéré comme suffisamment important, alors le code peut être augmenté d'une définition adéquate, que l'on se trouve dans un contexte centralisé ou distribué.

Cependant, l'incohérence de type 2, si elle est assez prononcée, peut déboucher sur de graves erreurs d'interprétation de la liste de notices qui s'affiche. Par exemple, un utilisateur peut voir un terme bibliographique employé d'une manière dans une notice, et d'une autre manière dans la suivante, où il l'interprétera de la première manière. Dans d'autres cas, une politique de catalogage peut réclamer le signalement d'une caractéristique donnée, et pas une autre politique de catalogage. Cela peut conduire un utilisateur à interpréter l'absence de signalement de ladite caractéristique comme indiquant l'absence de la caractéristique elle-même, alors qu'en fait elle peut être présente (mais simplement pas signalée par le catalogueur qui suit la seconde politique).

Plus important peut-être, l'incohérence des termes des index affecte souvent la recherche. Un défaut de cohérence peut aboutir à du silence, c'est-à-dire que certaines notices seulement (et non pas toutes) correspondant à un document, ou à une expression, ou à une œuvre, sont retrouvées à l'issue d'une recherche. Cela peut être fort fâcheux si le seul exemplaire disponible (sur le moment ou à l'avenir) pour le PEB correspond à la ou les notices non trouvées. Cela peut être également fort fâcheux si la ou les notices correspondant à l'expression la plus pertinente (par exemple, la dernière édition) ne sont pas trouvées. Cette remarque part de l'hypothèse, bien sûr, que les utilisateurs n'enchaînent pas directement avec

une autre recherche afin de s'assurer qu'il n'existe pas d'autres notices pour un document ou une œuvre, comme c'est souvent le cas — l'idée qu'il puisse exister une autre édition postérieure ne les effleure même pas.

D'un autre côté, si différentes notices pour la même œuvre reçoivent différents termes d'index, on peut y voir un avantage : cela accroît les chances de retrouver au moins une des notices. Si cela peut effectivement être préférable quand il n'y a pas de contrôle des autorités dans une base de données centralisée, les normes couramment adoptées comportent ce type de contrôle, c'est-à-dire, l'utilisation de vocabulaires par le catalogueur et par le chercheur (et par le système). Dans les systèmes distribués, on peut encore mettre en œuvre des normes sur les autorités, comme l'emploi de fichiers nom et sujet particuliers, mais de nouvelles vedettes nom et sujet peuvent en être exclues et n'avoir aucun fichier commun où elles puissent être créées.

En outre, le modèle de système distribué n'intègre pas (du moins pour le moment) de fichier d'autorité destiné à l'utilisateur final, lequel ne trouve qu'un réservoir de notices bibliographiques, dont les vedettes ne sont pas différenciées. Il n'y a ni renvoi ni liens automatisés vers les vedettes retenues, ni notes.

Toutefois, dans le cas des incohérences de type 3 et 4, le contrôle des autorités ne donne pas forcément la clé. C'est le cas où l'incohérence est due non pas à une carence du vocabulaire contrôlé concernant un nom en particulier, un sujet, une œuvre, une collection, etc., mais à la subjectivité et à des interprétations locales. Il est tout à fait possible que deux catalogueurs attribuent deux vedettes-matières LCSH différentes au même document. Et ce en raison soit d'une différence d'analyse du contenu soit de l'identification par chaque catalogueur de vedettes-matières différentes pour jouer le même rôle.

Derechef, interrogeons-nous : est-ce que tout ça a une quelconque importance ? Ne vaut-il pas mieux pouvoir retrouver une œuvre par plusieurs faisceaux de vedettes-matières ? Question en réalité fort complexe. Si diverses vedettes-matières résultent d'une attention différemment portée sur divers aspects de la thématique abordée par une même œuvre, alors on pourrait éventuellement parler d'erreur dans l'analyse du contenu — toutes les vedettes-matières pourraient être requises pour couvrir chaque aspect de la thématique. Spécimen de l'incohérence de type 1. Toutefois, et peut-être plus couramment, divers catalogueurs introduisent d'infimes nuances dans leur traduction de leur vocabulaire naturel vers le vocabulaire contrôlé. Dans la mesure où divers utilisateurs traduisent leurs concepts de recherche (issus du langage naturel) dans des formulations de requête avec là encore d'infimes nuances, on pourrait conclure à une certaine utilité des incohérences de type 2 et 3 : la même œuvre serait alors indexée de manière à couvrir diverses expressions de la requête. Toutefois, au bout du compte, il s'agit là de la fameuse balance entre réponse et précision. On obtient une meilleure réponse *globale*, puisque différentes requêtes aboutissent à la même œuvre pertinente, mais en même temps on perd en précision. Cela vient de ce que l'utilisateur trouve aussi des notices dont les vedettes sont attribuées selon des interprétations qui divergent de la sienne mais sont en accord avec celles des catalogueurs.

De toute évidence, c'est à chacun de juger de ce que doit être la balance optimale entre réponse et précision dans le contexte d'un catalogue collectif. Mais même si on se garde de prendre position en la matière, et de déterminer les effets que les incohérences de type 2 et 3 peuvent avoir sur cette balance, il existe encore une forme d'impact négatif de ces incohérences sur le bon résultat de la recherche. C'est qu'une recherche n'est pas menée pour

elle-même mais en relation avec d'autres recherches : l'utilisateur a une approche dynamique de la recherche. L'utilisateur peut s'adapter dans une certaine mesure à l'utilisation que le catalogueur fait du vocabulaire contrôlé, en s'y initiant par interactivité permanente avec le système. Plus cette utilisation est cohérente et systématique, moins l'utilisateur est dérouté, et plus il peut s'adapter.

L'auteur du présent article est donc d'avis que tous les types d'incohérence énumérés ci-dessus sont, de manière générale, funestes à l'utilisateur d'un catalogue collectif, et surtout les incohérences de type 1 et 2.

Quantité d'erreurs et de données

Une autre différence bibliographique qui existe entre catalogues collectifs distribués et centralisés, réside dans la tendance qu'ont ces derniers à présenter plus d'erreurs que les premiers [*sic ; l'auteur veut évidemment dire le contraire, NdT*] (y compris erreurs par omission). Ce qui revient à des incohérences de type 1 signalées dans la section précédente. Dans un système distribué, les bibliothèques ne disposent normalement pas de la possibilité de se faire corriger leurs notices par d'autres bibliothèques pour des documents qu'elles ont déjà catalogués, alors que dans un système centralisé, ça se fait automatiquement. Dans un système centralisé, une bibliothèque, au cours de son propre catalogage, peut corriger une erreur (comme une faute de frappe par exemple) qui figure sur une notice partagée dans la base de données centrale, et la notice corrigée peut alors être rechargée et remplacer la notice préexistante dans les catalogues locaux des autres bibliothèques qui partagent la notice dans le catalogue central.

En fait, il est également probable que l'on constatera une différence non seulement qualitative mais aussi quantitative. Et ce parce que des notices partagées peuvent recevoir dans la base de données centralisées des données supplémentaires (telles que table des matières, par exemple) en plus d'être corrigées, alors que toutes les notices représentant le même document dans un système distribué ne sont pas susceptibles de recevoir le même traitement. En d'autres termes, les notices partagées bénéficient de toutes les corrections et de tous les enrichissements (sauf adoption d'une politique collective contraire) apportés par les catalogueurs des bibliothèques qui les partagent, alors que ce n'est probablement pas le cas des notices d'un système distribué.

Bien qu'il soit possible de partager corrections et enrichissements dans un système distribué, le système ne s'y prête pas vraiment. Les catalogueurs qui procèdent à ces corrections et enrichissements n'ont généralement pas l'habilitation à intervenir sur les autres catalogues disposant de la notice qu'ils sont en train de copier (et de corriger et d'enrichir) ; et quand bien même ils l'auraient, les interventions multiples sur les catalogues des autres bibliothèques font peu vraisemblablement partie de leurs attributions ordinaires. Les catalogueurs peuvent signaler aux autres bibliothèques les corrections et/ou enrichissements auxquels ils ont procédé, afin que leurs propres catalogueurs les recopient, mais il est peu plausible que l'on encourage à mener cette opération de manière courante. Le rendement de la chaîne de catalogage s'en trouverait en effet sans doute quelque peu compromis.

En fait, la qualité des données bibliographiques peut encore être améliorée de deux façons dans un modèle de base de données centralisée. Tout d'abord, comme on l'a déjà dit, la probabilité d'exercer un contrôle qualité y est plus grande — dans beaucoup de cas c'est une agence centrale qui peut s'en charger. Ce contrôle qualité peut être renforcé, notamment pour

les bibliothèques dépourvues de personnel de catalogage professionnel ou expérimenté. Ensuite, le fait de saisir leurs notices dans une base de données centralisée, au vu et au su de leurs collègues des autres bibliothèques et des relecteurs de notices qui procèdent au contrôle qualité, peut pousser les catalogueurs (ou leurs bibliothèques) à pratiquer un peu plus l'*auto-surveillance*.

À quel point la correction et/ou l'enrichissement des notices catalographiques est une pratique courante, on peut s'en faire une idée en jetant un coup d'œil aux statistiques de catalogage de la base de données SILAS à Singapour. SILAS (Singapore Integrated Library Automation Services) joue le rôle de réseau de catalogage national et héberge une base de données centrale qui contient la catalogue collectif national. Quand un catalogueur d'une bibliothèque intervient sur une notice déjà « détenue » (c'est-à-dire, partagée) par une ou plusieurs autres bibliothèques, d'après les données d'exemplaire qui figurent sur la notice, la notice modifiée est automatiquement envoyée pour relecture au personnel du SILAS. (Si, en revanche, un catalogueur par copie ne modifie pas les parties bibliographiques partagées de la notice, mais se contente de décharger la notice telle quelle, après y avoir intégré la mention d'exemplaire et les autres champs de données locales, alors la notice peut échapper aux fourches caudines de la relecture par le SILAS). Pour le mois de mars 2003, 5 073 notices ont été recopiées sans modification, alors que 3 617 notices ont été recopiées avec modification, d'après les statistiques produites par le système SILAS. Sur la base de données SILAS au moins, on voit qu'une proportion non négligeable de notices (41,6 % sur mars 2003) ne sont pas simplement recopiées, mais aussi corrigées ou enrichies (en partant de l'hypothèse que la plupart des modifications apportent soit une correction soit un enrichissement).

Conséquences des erreurs et omissions de données

Certes, parmi les corrections et enrichissements apportés aux notices partagées, il y en a que chacune des bibliothèques pourrait apporter elle-même à ses propres notices au cours du temps, mais beaucoup d'entre elles ne le feraient probablement pas. Il n'est guère difficile de s'imaginer ce que pourraient être les conséquences d'erreurs et d'une diminution des enrichissements. Des notices pertinentes ne sont pas retrouvées, et parfois des notices non pertinentes sont retrouvées. En outre, des notices peuvent être mal sélectionnées, ou mal désélectionnées, dans le contexte du PEB et de la fourniture de documents, en raison d'informations erronées ou de l'absence d'informations sur la notice. Dans certains cas, les fautes de frappe ou l'absence de données peuvent aboutir à ce qu'on ne retrouve pas la seule notice *pertinente* correspondant à un document connu par avance. Cela peut arriver quand la notice corrigée, qui autrement serait retrouvée, n'existe plus en raison d'un désherbage opéré dans la bibliothèque dont le catalogue contient la notice corrigée, ou bien parce que la notice corrigée correspond à un document qui, dans cette bibliothèque, n'est pas disponible pour le prêt, ou bien est déjà en prêt.

Quand bien même il est très vrai que si la ou les notices corrigées existent toujours et correspondent à un document disponible, alors tout n'est pas complètement perdu, il peut en résulter néanmoins un service un tout petit peu moins efficace, puisque la ou les notices non corrigées peuvent représenter des bibliothèques qui, dans le contexte d'un système de fourniture de documents, auraient la préséance sur la bibliothèque (ou les bibliothèques) représentée par la ou les notices corrigées.

À coup sûr, dans le cas de la sélection de notices, l'utilisateur d'un catalogue collectif distribué peut n'examiner que la notice non corrigée ou non enrichie, et cela peut le mener à

une mauvaise décision lorsqu'il doit choisir d'utiliser ou de ne pas utiliser un service de fourniture de documents. Par exemple, une table des matières aurait pu signaler à l'utilisateur que la présence dans un ouvrage d'un paragraphe primordial aurait rendu la demande de cet ouvrage absolument indispensable ; ou bien, la présence d'une vedette-matière erronée peut lui faire commettre l'erreur fatale de demander un PEB ; ou bien peut-être encore, une erreur de date de publication peut le faire passer à côté d'un document qu'il va funestement désélectionner.

Et ce n'est pas tout, il y a encore d'autres effets « cumulatifs » des erreurs et des omissions. L'utilisateur risque d'en venir à ne plus faire confiance à la base de données. Et puis, les erreurs et les omissions sont recopiées dans les catalogues, et les problèmes se répercutent quand il s'agit de trouver une œuvre ou une collection données. Ce type de duplication survient quand on « clone » les notices. Par exemple, deux notices, dont l'une comporte une erreur, figurent pour la même édition d'une œuvre dans un catalogue collectif distribué. Eh bien, cette notice, et l'erreur dont elle est entachée, va se retrouver clonée dans le module de catalogage, puisque la bibliothèque achète une nouvelle édition ! Et si l'erreur affecte la recherche, l'utilisateur du catalogue collectif distribué risque de tomber sur la notice qui n'a *pas* l'erreur, et qui correspond à l'ancienne édition. Et il ne va sans doute pas chercher s'il y a une nouvelle édition.

Bien sûr, quand une notice partagée comporte une erreur qui n'a pas encore été repérée, en la clonant on clone aussi l'erreur, et alors toutes les notices sont non seulement clonées mais également erronées. Mais toutefois, il y a de bonnes chances quand même qu'on repère l'erreur au cours de la copie ou du clonage, au moins à un moment ou à un autre, et du même coup la ou les autres notices entachées d'erreur seraient corrigées. L'idée, ici, c'est que plus on touche à une notice, plus il y a de chances qu'une erreur soit corrigée.

Données locales

Il y a une autre différence qui pourrait avoir un tant soit peu d'impact sur l'utilisateur, et qui touche les données bibliographiques *locales* — pas les données d'exemplaire, mais les autres informations propres à un exemplaire ou les données propres à un catalogue, comme une note sur l'état de conservation matérielle d'un exemplaire, ou les vedettes-matières d'un thésaurus maison, ou une adresse URL qui pointe vers un site auquel une bibliothèque est abonnée. Dans un système distribué, l'utilisateur reçoit toutes les données bibliographiques locales à la queue leu leu, alors que dans un système centralisé ou hybride, en revanche, la notice qui figure dans la base de données centrale s'affiche en premier, et elle peut ne pas comporter ces informations locales.

Mais beaucoup de notices catalographiques, pourtant, ne comportent pas de données locales du tout, et la plupart d'entre elles en contiennent très peu, en comparaison de la quantité et de l'importance des données qu'elles partagent. Quoi qu'il en soit, dans un système centralisé ou hybride, l'utilisateur est encore à même de voir des données locales d'importance capitale en quelques instants, en recherchant la notice locale. Quant aux adresses URL inutilisables sans autorisation, on peut, dans les données partagées, indiquer clairement si la notice représente une ressource en ligne qui nécessite une autorisation. En outre, on peut contourner ce gravissime problème en intégrant l'URL aux notes sur les droits d'accès dans la notice de la base de données centrale. En fait, toutes les données « propres à un exemplaire » peuvent s'afficher dans la notice de la base de données centrale, avec le code de la bibliothèque auxquelles elles s'appliquent.

Pour résumer, l'auteur de la présente étude est profondément convaincu que, s'il peut être à l'occasion un tantinet gênant pour un utilisateur de ne pas voir s'afficher des informations bibliographiques locales dès le départ dans un système centralisé ou hybride, ce détail ne saurait dans la plupart des cas peser que de façon fort minime sur le choix du type de système à adopter.

Étude de cas : avant et après une base de données centrale [*sic ; l'auteur veut évidemment dire : « avant et après l'abandon d'une base de données centrale », NdT*]

SILAS héberge une base de données centrale qui a d'abord été créée comme pierre de touche du projet d'automatisation des bibliothèques au niveau national au milieu des années 80. En se développant, il a également développé sa deuxième fonction, qui a été d'héberger le Catalogue Collectif National (CCN). Les bibliothèques qui cataloguaient en ligne intégraient leurs exemplaires aux notices de la base de données, et les notices qui comportaient une ou plusieurs mentions d'exemplaires constituaient le CCN. Le catalogue collectif est utilisé pour formuler des demandes de PEB depuis plusieurs années, bien qu'il n'y ait pas de système officiel de PEB qui ait été établi au niveau national. Une des raisons de cet état de fait, c'est que le PEB n'est pas toujours indispensable quand les utilisateurs peuvent aisément se déplacer eux-mêmes de bibliothèque en bibliothèque — à Singapour, aucune bibliothèque ne se trouve à plus d'une heure de voiture. Quoi qu'il en soit, certaines espèces d'utilisateurs (les professeurs d'université par exemple) s'attendent à bénéficier d'un service de fourniture de documents, et c'est souvent le cas.

Depuis la fin des années 1990, cependant, il existe une tendance à rejeter le système centralisé et à militer en faveur d'un système distribué, lequel a finalement vu le jour sous le nom de « Tiara », aujourd'hui subsumé dans le « eLibraryHub » du National Library Board. Ce service peut aussi, bien sûr, être utilisé à des fins de fourniture de document, en donnant un accès Z39.50 aux catalogues de plusieurs grandes bibliothèques de Singapour. Grâce à la technique Z39.50, quelques-unes de ces grandes bibliothèques ont décidé de suspendre leurs contributions à la base de données centrale de SILAS, et d'intégrer leurs catalogues, à la place, dans des catalogues collectifs distribués, comme celui que propose Tiara.

Ce qui a débouché sur une situation inédite, dans laquelle certaines bibliothèques de Singapour partagent une notice bibliographique dans la base de données centrale si elles ont fait l'acquisition du document avant la fin des années 1990, mais pas si elles (ou l'une d'entre elles) en ont fait l'acquisition depuis lors. À la place, bien qu'elles continuent d'appliquer les mêmes normes bibliographiques, ces diverses bibliothèques cataloguent à présent leurs documents séparément, ce qui signifie que dans certains cas elles créent de nouvelles notices pour le même document indépendamment les unes des autres, et que dans d'autres cas elles dérivent différentes notices de différentes sources pour le même document.

Tandis que le National Library Board, par exemple, continue d'utiliser la base de données centrale du SILAS pour son catalogage par copie, et continuer d'alimenter la base de données centrale avec des notices nouvellement créées, l'Université nationale de Singapour ne dérive plus aucune notice de la base de données centrale, du moins pas directement, et les dérive d'autres canaux à la place.

Les anciennes notices de la base de données centrale du SILAS que partagent le National Library Board et l'Université nationale de Singapour (et n'importe quelle autre institution)

constituent un effort d'unification bibliographique et, en tant que telles, une cohérence absolue. Bien sûr, on se doute bien que cela ne veut pas dire qu'elles soient forcément toutes parfaites, mais au moins si elles comportent des bêtises elles persévèrent toutes dans la même bêtise et n'autorisent ni les incohérences ni les pertes quantitatives et qualitatives ni la haute teneur en doublons qui sont le lot obligé d'un modèle distribué.

Tandis que le nombre de doublons dans le catalogue collectif distribué dépendrait grandement des caractéristiques du programme de dédoublonnage mis en œuvre (ou pas mis en œuvre, d'ailleurs), les autres différences bibliographiques entre modèles centralisé ou hybride et distribué dépendraient encore plus de la manière dont les catalogueurs cataloguent dans leurs bibliothèques respectives, dans le contexte du système spécifique dans lequel ils travaillent. Il serait donc tout à fait passionnant de mesurer le taux d'incohérence bibliographique atteint dans un système distribué, le nombre d'enrichissements de notices auxquels on aurait pu procéder, sinon, dans une base de données centrale, et le nombre d'erreurs qui auraient peut-être pu être corrigées.

Ces facettes des conséquences bibliographiques d'un passage vers un système distribué, sans base de données centrale, ont été examinées dans une rapide étude sur le catalogage post-centralisé accompli par le National Library Board (NLB) et l'Université nationale de Singapour (UNS). Un échantillon aléatoire de vingt documents, originellement catalogués par le NLB en 2001, ont été identifiés comme appartenant aussi aux collections de l'UNS. Les notices du NLB et de l'UNS pour ces documents ont été retrouvées et affichées en format complet, et ont été examinées des points de vue suivants :

1. Identité des (champs) points d'accès
2. Non-identité des (champs) points d'accès
3. Identité des champs descriptifs non indexés
4. Non-identité des champs descriptifs non indexés
5. Champs indexés supplémentaires
6. Champs non-indexés supplémentaires
7. Erreurs figurant dans une notice et pas dans l'autre.

Seules ont été prises en considération les données bibliographiques contenues dans les champs qui peuvent affecter la recherche (dans un système standard). Les différences ont été négligées lorsque l'on estimait qu'elles n'étaient pas susceptibles d'affecter soit la recherche, soit la sélection/dé-sélection aux fins de fourniture de document. Par exemple, la différence entre deux codes de sous-champs MARC a pu être négligée pour cette raison ; une différence de 1 cm dans la hauteur relevée pour le dos de la couverture d'un même livre dans les diverses notices a été jugée négligeable. En outre, les champs de données additionnels, trouvés dans une notice et pas dans une autre, n'ont été pris en compte que s'ils avaient une valeur ajoutée en termes de recherche et/ou d'aide à la décision de sélectionner ou non un ouvrage. Les erreurs qui figuraient dans une notice mais pas dans l'autre, et qui auraient donc pu être corrigées par l'une ou l'autre des bibliothèques, et qui ont été considérées comme des erreurs manifestes et non pas peut-être causées par des différences d'interprétation, n'ont été prises en compte que si l'on estimait qu'elles pouvaient avoir un impact sur la recherche et/ou sur l'aide à la décision de sélectionner ou non un ouvrage.

Là où il y avait des différences dans les vedettes-matières et où le nombre de vedettes n'était pas le même dans les notices du NLB et de l'UNS, le nombre de champs comportant des vedettes-matières différentes a été noté comme la moitié du nombre total des zones variables.

Par exemple, s'il y avait trois vedettes-matières dans une notice, et quatre dans l'autre, et qu'aucune d'entre elles ne concordait, alors le nombre de champs différents était considéré comme étant de 3,5.

Sur le Tableau 2 ci-dessous, on peut constater que plus d'un tiers (35,2 %) des champs permettant un accès divergeaient d'une manière ou d'une autre, et qu'une large majorité (80 %) des champs non indexés divergeaient de manière non négligeable. Un taux d'environ deux champs additionnels par couple de notices comportait des données importantes, avec un taux de plus d'un champ indexé additionnel par couple de notices. Sur les 195,5 champs des couples de notices jugés potentiellement funestes à l'utilisation d'un service de fournitures de documents, 97,5 (50,1 %) étaient soit incohérents, soit présents dans une seule des deux notices. Toute utilisation sérieuse d'un catalogue collectif est vouée à subir les épouvantables conséquences d'une telle disparité des contenus.

Parmi les divergences observées, on notera que :

- (a) beaucoup de vedettes-matières étaient différentes, et dans la plupart des cas les différences n'étaient pas purement syntaxiques, mais sémantiques ; par exemple, des sujets additionnels étaient présents dans une notice et pas dans l'autre
- (b) trois couples de notices comportaient des dates de publication différentes, ce qui serait gênant pour une recherche avec des bornes chronologiques
- (c) deux paires de notices donnaient des ISBN différents ; cela peut être absolument catastrophique si l'utilisateur ne trouve que la notice correspondant à l'exemplaire qui n'est pas disponible pour le PEB
- (d) plusieurs noms n'avaient pas leur vedette forme retenue complète dans une notice, mais l'avaient dans l'autre, ce qui montre à quel point le niveau général du contrôle des autorités peut sombrer dans la plus lamentable des décadences dans un système décentralisé
- (e) une notice comportait un titre de partie, et pas l'autre
- (f) dans tous les couples de notices sauf un, les champs de la description matérielle du document n'étaient pas remplis de la même façon, par exemple, une notice indiquait « nombreuses ill. en coul. » alors que l'autre n'indiquait nullement que la publication consistait essentiellement en un livre d'images
- (g) dans un couple de notices, les mentions d'édition différaient substantiellement, ce qui peut être catastrophique si l'utilisateur recherche avant tout la fraîcheur des informations et que cette différence lui fait rater l'édition la plus récente dans sa demande de PEB
- (h) pour sept des couples de notices, une table des matières figure dans une notice et pas dans l'autre ; et pour trois couples de notices on trouve un résumé dans une notice et pas dans l'autre ; ces champs rendent une notice plus accessible, notamment pour une recherche par sujet, et sont d'une utilité particulièrement précieuse pour juger si un document est utile
- (i) pour deux couples de notices il y avait des champs de variantes de titre dans une notice mais pas dans l'autre
- (j) pour cinq couples de notices il y avait des points d'accès par nom additionnels dans une notice et pas dans l'autre.

On notera au passage que deux des couples de notices (11 et 13 sur le Tableau 2) représentaient des éditions d'une œuvre dont une édition antérieure était partagée par le NLB et l'UNS dans la base de données centrale. Leur taux de cohérence était supérieur (6/8 et 6/9) que la moyenne générale des vingt couples de notices (58,2 %).

En moyenne, on a donc trouvé plus d'une erreur par notice — souvent les notices se contredisaient les unes les autres, aussi est-il raisonnable de supposer que, dans bien des cas, l'une ou l'autre des notices était correcte, et l'erreur aurait pu être corrigée (ou ne jamais se produire) dans une base de données partagée par le NLB et l'UNS. Même si la moitié seulement ou le tiers de ces erreurs ne se seraient pas produites, cela représenterait quand même encore un nombre d'erreurs important par rapport aux critères usuels de qualité du catalogue. Par exemple, l'élimination (définitive) d'un tiers de ces erreurs voudrait dire que le taux d'erreur pourrait encore être réduit de quatre erreurs pour dix notices.

Les bénéfices bibliographiques augmentent en fonction du nombre de catalogues

Il faut noter que si l'effet de la décentralisation du catalogue est déjà sensible sur deux catalogues de bibliothèques, il ne sera que plus grand sur dix ou vingt catalogues de bibliothèques : il y aura encore plus d'incohérences, d'absences d'enrichissement et d'erreurs.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
Identité des (champs) points d'accès	3	5	4	5	7	6	3	4	0	4	5	4	6	4	8	5	3	8	5	4	93
Non-identité des (champs) points d'accès	4,5	2	2,5	2,5	1	0	8	2,5	4	1,5	1	3,5	2	3	0	3	1,5	0	3	5	50,5
Identité des champs descriptifs non indexés	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	1	0	5
Non-identité des champs descriptifs non indexés	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	0	1	20
Champs indexés supplémentaires	1	0	1	0	2	1	0	2	3	2	0	2	2	1	1	0	1	1	5	1	26
Champs non-indexés supplémentaires	0	0	0	0	0	1	1	1	1	0	2	0	1	0	0	0	1	0	1	2	11
Erreurs figurant dans une notice et pas dans l'autre	1	0	1	2	1	1	2	1	6	2	1	3	1	0	0	0	1	0	0	1	24

Tableau 2 : Divergences entre notices, base de données post-centrale

Conclusions

Si des accords pour la fourniture de documents existent bel et bien entre certaines bibliothèques de Singapour, l'auteur du présent article exprime l'opinion qu'un service plus large devrait être mis en œuvre, et qu'un tel service tirerait de grands avantages bibliographiques d'une base de données centralisée dont le SILAS continuerait d'assurer la maintenance. Bien qu'il faille prendre en compte de multiples paramètres en dehors des questions purement bibliographiques qui ont été examinées ci-dessus, un simple petit échantillon de notices a montré comment une base de données centralisée peut améliorer l'utilisation d'un tel service de fourniture de documents. Toutefois, il y a deux questions tout à fait primordiales qu'il faut prendre en considération pour pouvoir mener à bien l'essentiel de cette amélioration. Tout d'abord, il faudrait rendre le Catalogue Collectif National de Singapour accessible à l'utilisateur final, alors qu'aujourd'hui il n'est accessible qu'au seul personnel des bibliothèques membres du SILAS. Ensuite, il faudrait mettre à jour la base de données centrale, puisque certaines bibliothèques fondamentales de Singapour ont cessé de l'alimenter depuis plusieurs années.

La plupart des services de prêt entre bibliothèques reposent sur le présupposé que les besoins en informations des utilisateurs des différentes bibliothèques se recoupent partiellement. Leurs recherches bibliographiques ne sont donc pas nécessairement liées à un catalogue spécifique. C'est-à-dire qu'ils peuvent tirer bénéfice d'une recherche lancée simultanément sur plusieurs catalogues, sur le même sujet, le même auteur, la même œuvre, etc. Le présent article soutient que les utilisateurs seraient mieux servis par des catalogues présentant certaines qualités bibliographiques. Au bout du compte, on souhaiterait que les différents catalogues fussent fusionnés, afin que l'utilisateur retrouve une et une seule notice pour chaque document. Une notice par document (ou par manifestation), au lieu d'une notice par exemplaire, est un principe catalogographique bien établi. La cohérence en est un autre. Les

utilisateurs tireraient avantage de notices dont le contenu serait semblable pour des documents semblables, afin d'exprimer ces similitudes. Les utilisateurs tireraient également avantage d'une cohérence dans le style du contenu des notices, afin qu'il soit plus facile à comprendre. L'exactitude est encore une autre qualité bibliographique de première importance. Les utilisateurs tireraient généralement avantage à être confrontés à moins d'erreurs. Et pour finir, une autre qualité primordiale lorsque le catalogue est utilisé à des fins plus sérieuses — et parmi ces fins plus sérieuses il y a à coup sûr l'utilisation du catalogue pour répondre aux demandes de fourniture de documents, qui représentent un surcoût et en temps et en argent — réside dans la présence de plus de contenu (utile). En d'autres termes, les utilisateurs tireraient avantage de trouver autant de notices enrichies que possible.

De l'avis de l'auteur du présent article, une base de données centrale a plus de chances de produire les qualités bibliographiques énumérées ci-dessus en plus grande quantité qu'un catalogue collectif distribué. Et ce au moins pour quatre raisons majeures : (a) le fait de cataloguer dans une base de données centrale permet de s'assurer que les doublons de notices ne sont dus qu'à une erreur humaine des catalogueurs ; (b) il est plus facile d'implémenter un mécanisme de contrôle qualité permettant de relire les notices versées et révisées dans le catalogue collectif, peut-être par le biais d'une agence centrale ; (c) les catalogueurs vont sans doute fournir de plus grands efforts pour appliquer les normes et politiques convenues, puisqu'ils sont plus exposés au contrôle qualité et au jugement de leurs pairs ; (d) les notices sont plus susceptibles d'être corrigées et enrichies par le travail des autres bibliothèques.

Références bibliographiques :

COBISS: Co-operative Online Bibliographic System and Services. (25 May 2003). Online.
http://www.cobiss.net/cobiss_platform.htm

Coyle, Karen. (2000). The Virtual Union Catalog: A Comparative Study, *D-Lib Magazine* (accessed online 25 May 2003). <http://www.dlib.org/dlib/march00/coyle/03coyle.html>

Crossnet Systems Ltd. (2000). *CURL Z39.50 feasibility study* (accessed online 25 May 2003).
<http://www.curl.ac.uk/projects/z3950.html>

Hider, Philip. (2002). A Survey of National Union Catalogues, *Singapore Journal of Library & Information Management* v31: 73-78.