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

1-9 August 2003, Berlin

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

The bibliographic advantages of a centralised union catalogue for ILL and resource sharing

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

Abstract

This paper examines some of the bibliographic advantages of a union catalogue with a central database over a distributed, or "virtual," union catalogue. Such advantages may be worth considering when a library network makes a decision about which system model to adopt as the basis of a document delivery service. The nature of these bibliographic advantages is discussed in the context of interlibrary document delivery, as are the circumstances which produce them, and make them more significant. A brief study of the extent to which two major library catalogues in Singapore have diverged following the adoption of a distributed model, is reported. This indicated that the bibliographic content of a distributed union catalogue may be significantly poorer than that of a central database, and in particular in terms of more: (1) duplication; (2) inconsistency; (3) errors; (4) omissions. There are at least four important reasons why this may be so, since in a centralised system: (a) record duplication is caused only by cataloguer error; (b) it is easier to implement quality control mechanisms; (c) cataloguers are likely to make greater efforts to adhere to agreed standards and policies; (d) records are more likely to get amended and enhanced through the work of other libraries.

Introduction

Union catalogues have always been a very useful tool for ILL and document delivery. They provide "one-stop shopping," to use a current term, simultaneous access to plural collections, whether the collections be physical, virtual or both. It is important to remember that union catalogues need not only be a "tool" in the sense of being used to implement ILL and document delivery: they can also serve to formulate document delivery requests, and often do.

If it is a library's aim for its document delivery service is to be used as fully as resources permit, then one important means for such utilization is the provision of public access to the union catalogue, so that not only library staff but also end-users can identify all the more readily potentially valuable interlibrary loans and other interlibrary transactions. Although one way in which demand, if it outstrips supply, might be "capped" is through deliberate restriction of access to the bibliographic information, such a policy is considered by many to be wrong-headed, for supply does not have to meet demand in the context of most document delivery services: interlibrary loans do not have to be entertained; online document delivery can be refused according to an automated algorithm. If one has to prioritise requests, then in most cases prioritisation can be done more satisfactorily by means other than restricting end-users' access to library catalogues and other databases.

One should note, however, that there are still union catalogues which could be made available to the public but are not, despite the fact that libraries represented by them do offer some interlibrary document delivery services. While there are many reasons why this situation occurs, the fear of being deluged by interlibrary loan requests is one of them. On the other hand, the proportion of union catalogues which are now publicly accessible would appear to be on the increase, with many now freely accessible through the Internet.

The author in this paper assumes that ILL and other document delivery services generally wish to encourage the formulation of potentially valuable requests on the part of users, and welcome the union catalogue as a very helpful tool for doing so. The question of the bibliographic advantages of one type of union catalogue over another still applies even if the union catalogue is not seen in this light, but merely as a tool in the more basic sense of being used to track down requested resources by library staff items. However, the question becomes much more important when the union catalogue is used by end-users *before* they approach the interlibrary loan counter, in order to identify the resources they wish to procure through the document delivery service.

Centralised and distributed union catalogues

There are several ways to design a union catalogue. The traditional way is to establish a centralised database into which records may be contributed either directly (create records in the central database) or indirectly (local catalogue first, then upload to union catalogue), or both. With protocols such as Z39.50, there is now an alternative, namely, to establish a *distributed* union catalogue. In this model, the local catalogues are linked through their respective servers, probably using Z39.50, while the Z39.50 client searches each catalogue in turn, creating "on the fly" a virtual union catalogue. The distributed union catalogue has become a popular concept in some library circles in recent years. In some situations, libraries are more likely to participate in the establishment of such catalogues, since the traditional scheme involving a central database can prove a major inconvenience: the library's cataloguing workflow may be hampered by the need to feed records into the union catalogue, for example.

Advocates of the distributed catalogue also point out how it is more likely to produce *extended* holdings information of a real-time nature. That is, not only will the system inform the user of those libraries holding an item, but it can also inform the user of the availability status of the item in the respective libraries. However, it is in fact perfectly possible for a union catalogue based on a centralised database to also perform additional searches via Z39.50, etc. to provide such real-time information to the user, although such *hybrid* systems

combining both distributed elements (i.e. the extended holdings information) and centralised elements (i.e. the bibliographic data and summary holdings information) have been implemented in few cases as yet.

Since it is feasible to establish centralised union catalogues with one-stop ILL and document delivery facilities, making use, perhaps, of supplementary Z39.50 functionality, the preference for distributed over centralised arrangements should be based more on the convenience that the individual libraries would thus enjoy, as well as the circumstances at the managing agency (such as the national library). However, not all libraries may be inconvenienced by a centralised arrangement; moreover, the Z39.50 servers that the distributed union catalogue requires are still expensive—too expensive for many smaller libraries, and quite possibly too expensive for the national library or central institution to acquire on behalf of each and every one of the participating libraries.

Nevertheless, such Z39.50 servers are unlikely to remain expensive for ever, and as their affordability increases, the question needs to be asked: are there any other reasons why the centralised or hybrid model, which includes a central database, might still be preferred to the completely distributed model? This paper examines some of the *bibliographic* reasons why this might be so, that is, why users' searching might be aided by a central database, compared with a virtual union catalogue, due to differences in bibliographic content. Such searching is placed particularly in the context of ILL and resource sharing, although it may also relate to the use of a union catalogue for other purposes (e.g. reference work). While several other factors—e.g. technological, financial, political—have already been widely discussed, the effect that each model has on the bibliographic data itself has not.

One should bear in mind that in addition to possible bibliographic advantages, there are other reasons why *users* might prefer a centralised or hybrid model, for example, large distributed systems may exhibit significantly slower net response times. It is important to note that many contemporary Z39.50 clients and servers do not yet offer the sophisticated searching functionality that one can now enjoy on many host catalogues. Coyle (2000) underlines the basic requirement: "For a virtual union catalog to be feasible, the participating databases must offer a uniform set of indexes and search functions that retrieve comparable items from each catalog."

Union catalogues and bibliographic utilities

The distributed union catalogue has only become implementable in recent years, through the widespread use of Z39.50. Prior to this telecommunications breakthrough, union catalogues were dependent on the establishment of a central bibliographic database. However, such central databases were usually established primarily for record supply purposes, that is, as the basis of a bibliographic utility; their function as union catalogues was of only secondarily importance. With the rise of the Internet and the advances made in telecommunications over the past decade, the future of many regional and even national bibliographic utilities has been threatened by the growing number of alternative sources of catalogue records. Perhaps because of this, or at least in part, the union catalogue function of many of these central databases has been increasingly emphasised, and supplementary services based on this function offered—including, of course, ILL and document delivery.

There are thus two pragmatic reasons why centralised systems may be preferred over distributed ones: first, many already exist, and their databases are often large and have been

nurtured over a considerable number of years; second, many of the central agencies of bibliographic networks have already invested in the development of ILL and document delivery services based on their central database, and may not wish these services to be superseded.

Most of the national bibliographic networks surveyed by Hider (2002) do not as yet appear to be switching to a distributed model, although many have developed, or are developing, an automated document delivery platform on top of the central database (e.g. DanBib (Denmark), Te Puna (New Zealand), Kinetica (Australia), LIBRIS (Sweden), CCB (Belgium)). Some national networks are coordinated (and small) enough to have circumvented the central versus distributed question by sharing the same library system (e.g. ELINET (Estonia) and COBISS (Slovenia)).

In some of the descriptions of these national networks, the importance of maintaining an internally consistent central database is stressed. For instance, on the COBISS website, we read:

> A cataloguer first checks as to whether the bibliographic record he/she wants to add in the local database, already exists in the COBIB union database. If it does, he/she downloads it to the local database, adding the copy-specific holdings data. If it does not, he/she creates the record in the local database, from which it is then automatically transferred into the union catalogue. The cataloguer is not allowed to edit records in the COBIB union database unless they were created by his/her library, or he/she was granted an official permission to do so.
>
> The quality, uniformity and consistency of the local databases and the COBIB union database are provided in different ways: by duplicates control, by COMARC program controls, by record editing, by common (global) code lists for all standardised data (e.g. countries, languages, UDC), by local codes lists to provide uniformity of data within a library (e.g. locations, internal designations), by automatic counters (e.g. accession numbers, numbering in call numbers), by unique identification control of serials, etc., and above all by providing a systematic training for record creators to obtain the official permission for cataloguing…
>
> A great deal of attention has been given to the uniformity of data entry and local data display intended for end users. The uniformity has been provided to a great extent by the COMARC/H Format for holdings data. Besides, prior to the inclusion of an individual library into the real shared cataloguing environment, the Library Information Service and a respective library agree (by preparing a special Minutes on Holdings Data) on the way of the local data entry. The latter is adapted so as to render an end-user friendly presentation, and at the same time preserve the uniqueness of the work organisation in any of the libraries.

One must bear in mind, however, that there are many other union catalogues, or potential union catalogues, apart from national ones, in many cases based on specific library types, and that many of these do not have the legacy of a well established central database, nor a central agency to take care of one. Even some of those that do, such as COPAC, the union catalogue for CURL (the influential network of British academic library catalogues), have recently considered switching to a distributed model (Crossnet Systems Ltd. 2000).

Before we move on, then, to the bibliographic reasons for preferring centralised or distributed systems, we should bear in mind the following point: where a central database already exists, the more comprehensive and carefully built-up it is, the more relevant become the bibliographic advantages that might pertain to it; conversely, where a database has been allowed to deteriorate, in terms of bibliographic quality and comprehensiveness, the less immediately relevant are the advantages. In other words, the bibliographic advantages of a

central database that are highlighted in this paper accrue over time. Centralised systems started in the Eighties that are still well maintained have a very significant head start over a centralised system being established today.

Possible bibliographic differences between models

Amount of duplication

The most obvious difference between the bibliographic content of centralised databases and virtual union catalogues concerns duplication. It is generally more difficult to produce unique citations in the distributed union catalogue. Even if a distributed system includes a sophisticated de-duplication programme which "catches" duplicate records (according to its algorithm) as they are received from the libraries' Z39.50 servers, it is unlikely that it will catch all of them.

While it is in fact possible for a distributed union catalogue to also act as a database for record copy, if each of the libraries is equipped with a Z client as well as a Z server, variant records are still more likely to appear in the local catalogues, and at a greater frequency than in a central database, increasing the risk of failure on the part of the system's de-duplicator (if there is one). The reason for this is that records may be revised in the local catalogue, after being downloaded from another participant's catalogue.

It is conceivable that a very sophisticated distributed system could minimise the problem of duplication by inserting a random control number into the record it downloads from another catalogue (if it does not have one already), both in the downloaded and original copy, but this requires editorial access to the host catalogue, at least in terms of the control number field. It would also be necessary to be able to distinguish between straight copy cataloguing and "cloning" (copying a record in order to edit it into a record for a bibliographically similar item), such that a control number is not generated in the case of the latter. Although a random control number of a long enough sequence would probably be and remain unique, there is also the chance of this being duplicated.

Even if such a sophisticated system were developed, there is still another possible cause of duplication that would probably not apply to the central database model: when one or more of the library servers is "down," the cataloguer may still press ahead and assume there is no existing match on any of the other catalogues (they might not even realise that a server, out of one or two dozen perhaps, cannot be accessed).

There is also a greater danger of records being incorrectly merged by a de-duplicating programme, that is, records being collapsed which do not in fact represent the same item (e.g. they are for items in different formats). While this may happen in a central database on occasion, a de-duplicating programme that matches only on, say, the title and standard number fields would frequently collapse records for different works, let alone different expressions and manifestations of a work. (While the collapsing of records for different expressions and manifestations at the first-level display might be helpful, the collapsing of records for different works would certainly not be.)

One should note that duplication may also occur in a central database if new records are uploaded to it after they have been created locally, due to the time lag between creation and uploading (another library might create their own record for the same item during this period).

However, as long as each library uses the central database for all its copy cataloguing (as is normally the case), and as long as any such time lags do not extend into months and years, this workflow (uploading to central database afterwards) should not produce the kind of duplication that is likely to be encountered in distributed systems.

Consequences of duplication

If we concede that in many situations, a centralised catalogue is less likely to produce duplicate records than is a virtual union catalogue, the next question to ask is: does this matter? This depends on the specific search that a user is carrying out, on the number of catalogues represented, and on the extent of overlap of collections.

For a know-item search, it may well be that only a few duplicate records are displayed (initially in brief citation mode), and these can be prioritised by holding library. However, it is not always the case that even a known-item search retrieves records for only one item. Indeed, it is very often the case that it does not. A known-item search could be based on a standard number, but not in the case of many materials, particularly AV materials, which do not have standard numbers. Even where a standard number does apply, it may retrieve records for both parts and whole item, records for items with a wrongly assigned with standard number, and so on. Moreover, in many situations, the user of a document delivery service will not know the standard number of the item they seek. Another common search for a known item is, of course, title and/or author. Here there is even more chance of duplication, given the non-uniqueness of titles and names.

We have already argued that union catalogues can perform a pre-request role in the document delivery process, that is, that they can provide users (both intermediaries and end-users) with a means to identify what items it is they might wish to request. As such, they would often be searched for unknown items, on a subject, by an author, etc. As we move into the area of subject searching and other unknown-item searching, so the likelihood of a large results set increases, and with it, duplication.

If a user were to examine the full contents of a results set, then a large amount of duplication would not affect the success or otherwise of the search, although it might irritate and distract. However, in reality, users tend to examine only those records appearing in the first one or two pages of a results screen. The problem, of course, is that duplication adds “noise” and reduces the number of relevant items likely to be found in those one or two pages of results.

Duplication and noise also increase as the number of catalogues represented by the virtual union catalogue increases, and as the ratio of holdings to records increases. In the National Union Catalogue of Singapore there are currently about 3 million holding statements attached to about 2 million bibliographic records, with over 50 libraries represented. Holding numbers from a random sample of 1,000 records added to the union catalogue between one and three years ago, were extracted from the database to reveal the distribution shown in Table 1. We observe an approximation of the familiar Zipf function.

Holdings per record	*f*
1	770
2	117
3	52
4	35
5	7
6	12
7	4
8	2
9	0
…	0
22	1
	1,000

Table 1 : Distribution of Holdings

If a very precise known-item search retrieves only those records that match the sought item, then duplication matters little. However, if a search retrieves records for more than one item, then duplication can matter. It is even possible for it to matter in the case of a known-item search. For instance, if an author/title search retrieves records for several different items, perhaps all different works, a perfectly reasonable outcome if the search words are fairly common and the union catalogue is, in sum, quite large, then the user will miss out if they fail to page to the second screen, and the first screen shows only other items from multiple records (e.g. four items represented by 3+3+2+2 records).

In the case of subject searches, duplication is much more likely to matter. This is because records for more items are often going to be retrieved, and more of these items may well be considered worth requesting through the document delivery service. A user may miss out if he examines only those citations on the first screen, for example, and the chances of duplication causing this possibility are high, if there are records for more than ten items retrieved. We can estimate what the chances are according to the distribution from the sample above, assuming there is no de-duplicator, that all duplicates are retrieved on the search and displayed together, as

$$p \approx 1 - (0.77^{10} + 0.77^{9}.(1\text{-}0.77)) = 0.905$$

Level of consistency

The amount of duplication is not the only bibliographic difference between distributed and centralised union catalogues. Another important difference is one that is likely in some circumstances and almost certain in others. This is the difference of the level of consistency. Although in theory, even in a distributed arrangement, libraries may adhere to the same bibliographic standards and policies, there are two ways in which bibliographic consistency is likely to be reduced in the case of a distributed system.

First, even with a virtual union catalogue to hand, cataloguers are much less likely to create new records in the context of it than they would be if they were cataloguing into a central database, or even if they were creating records locally, but had a central database to hand. There are several reasons for this: it is likely to be considerably more time-consuming

interrogating a virtual union catalogue than a central database; there is no authority file that a central database may offer; cataloguers may not be able to "clone" similar records from other libraries if they are not able to download records from the virtual union catalogue (they need their own Z client appended to their local system).

Second, the monitoring of adherence to the agreed standards and policies is likely to be much more stringent in a centralised system than in a distributed one. Indeed, a centralised system allows for a mandatory review of records, which may be of particular value with respect to contributions from smaller libraries without full-time cataloguing staff.

One should note that cataloguing standards and policies change anyway, and that their application varies within a catalogue, but nevertheless, a central union catalogue is likely to exhibit *less* inconsistency than is a distributed union catalogue for the reasons mentioned above, and the difference could be significant.

Consequences of inconsistency

There are at least four types of bibliographic inconsistency, that which is:

(1) due to an error on the part of one or more cataloguers
(2) due to cataloguers applying different standards or policies
(3) due to cataloguers interpreting the same standards and policies differently
(4) due to a different interpretation of the information derived from the item.

The first type of inconsistency, we shall address in the following sections on differences in record quality. All four types may be exacerbated by a distributed system.

The last three types of inconsistency all affect retrieval, but types (2) and (3) also affect the *reading* of a record once retrieved, which in turn may result in an ILL request when it should not have been, or vice-versa.

While cataloguers are less likely to interpret standards and policies differently if they have reference to a shared database, it may be generally assumed that such divergence is in any case of little consequence as far as the description of the item is concerned—otherwise this divergence would, or should, be addressed in an appropriate extension of the said standards and policies. For example, a cataloguing code might not define a certain form of illustration that should be recorded when it is present—if this is considered a significant issue, then whether in a centralised or distributed situation, the code can be extended with an agreed definition.

However, type (2) inconsistency, if pronounced enough, may well result in significant mis-interpretations of full record displays. For example, the user may see a bibliographic term used in one way on one catalogue record, but then in a different way on the next, but interpret it according to the way used in the first record. In another case, one cataloguing policy may call for the recording of a particular feature, while another policy does not. This might lead a user to interpret the absence of the recording of the particular feature as absence of this feature, when in fact it might nevertheless be present (simply not recorded by the cataloguer following the second policy).

Perhaps more importantly, inconsistency of index terms often affects retrieval. A lack of consistency may well make for incomplete retrieval, that is, only some (and not all) records for an item, or for an expression or work, are retrieved following a search. This may be unfortunate if an ILL is only available (either now or in the future) for the item represented by record(s) not retrieved. It may also be unfortunate if the record(s) for the preferred expression (e.g. latest edition) is/are not retrieved. This of course assumes that users do not follow up with another search to check that additional records for an item/work do not exist, as is in fact often the case—the thought that there might be a later edition, say, might not even occur to them.

On the other hand, if different records for the same work are assigned different index terms, this may be considered an advantage: there is more chance that at least one of the records will be retrieved. While this may indeed be preferable when there is no authority control present in a central database, usually standards that are agreed upon include provision for such control, that is, the use of controlled vocabularies by cataloguer and searcher (and the system). In distributed systems, there may still be authority standards applied, such as the use of particular name and subject files, but new name and subject headings may be excluded from these and have no common file in which to be established.

Moreover, the distributed system model does not incorporate (at least not yet) an authority file for the benefit of end-users, who retrieve only a pool of bibliographic records, undifferentiated with respect to their headings. There are no references or automatic links to authorised headings, no scope notes.

However, in the case of type (3) and type (4) inconsistency noted above, authority control does not necessarily provide the answer. This is where inconsistency is due not to a lack of a controlled vocabulary covering a particular name, subject, work, series, etc., but due to due to subjectivity and local interpretations. It is very possible for two cataloguers to assign, say, different Library of Congress Subject Headings for the same item. It may be because of different subject analyses or because the cataloguers identified different subject headings which they considered appropriate.

Again, we need to ask the question: does this matter? Is it not better that a work can be retrieved through more than one set of subject headings? This is, in fact, a complex issue. If different subject headings result from a focus on different topics within the same work, then this might be considered an error of subject analysis—all subject headings might be necessary to cover each topic. This would fall under type (1) inconsistency. However, perhaps more commonly, different cataloguers translate from their natural vocabularies to the controlled vocabulary slightly differently. Since different users translate their search concepts (originating in natural language) into search formulations also slightly differently, we could conclude that types (2) and (3) inconsistency might be useful: the same work would then be indexed to cover different search expressions. However, what we end up with here is in fact the familiar trade off between recall and precision. We attain greater *net* recall, so that different searches both retrieve the relevant work; but at the same time we lose a certain degree of precision. This is because a user also retrieves records with headings assigned according to interpretations which do not accord with their own, but with those of other cataloguers.

It is clearly a matter of judgement what the optimal balance between recall and precision might be within the context of a union catalogue. However, even if we take no position with

respect to the effect types (2) and (3) inconsistency have on this balance, there is still a way in which such inconsistency has a negative impact on search success. This is because a search is not performed in isolation but in the context of other searches a user makes on the system: a user's approach to searching is dynamic. Thus users can adapt, to some extent, to a cataloguer's application of a controlled vocabulary, as they learn this through continued interaction with the system. The more consistent the applications of the vocabulary in the system, the less confusing for users, and the easier it is for them to adapt to the applications.

This author therefore takes the view that all types of inconsistency listed above are generally detrimental to the user of the union catalogue, types (1) and (2) particularly so.

Amount of errors and data

Another bibliographic difference between distributed and centralised union catalogues is the tendency for the latter to exhibit more errors than the former (which would include errors of omission). This equates to type (1) inconsistency mentioned in the preceding section. In the distributed system, libraries would normally not benefit from other libraries amending records for items that they have already catalogued, which they could benefit from automatically with a centralised system. In a centralised system, a library, in the process of its own cataloguing, may correct an error (e.g. a typographical one) on a shared record in the central database, and the revised record can then be uploaded and replace the existing record in local catalogues of other libraries which share the record in the central catalogue.

Indeed, there is also a likelihood of a difference of quantity as well as quality. This is because shared records can be upgraded in the central database with additional data (e.g. tables of contents), as well as amended, whereas not all of the records representing the same item in a distributed system are likely to receive the same treatment. In other words, shared records receive all amendments and enhancements (unless collective policy states otherwise) made by the cataloguers of the libraries which share them, whereas records in a distributed system probably will not.

Although it is possible for amendments and enhancements to be shared in a distributed system, the system does not readily lend itself to this. Cataloguers performing the amendments and enhancements would not usually be granted editorial access to the other catalogues with the record they are copying (and then amending/enhancing); and even if they were, multiple editing on other libraries' catalogues is unlikely to fall within their job scope. Cataloguers might communicate amendments and/or enhancements to other libraries, for their own cataloguers to replicate, but doing so on a routine basis is unlikely to be entertained. If it were carried out, it would make for a significantly less efficient cataloguing process.

There are in fact two other ways in which the quality of the bibliographic data can be improved in the central database model. First, as noted earlier, there is likely to be more quality control—in many cases carried out by a central agency. Such quality control might in fact be enforced, particularly for libraries without professional or experienced cataloguing staff. Second, the entering of their records into a central database, in full view of colleagues from other libraries and record reviewers carrying out quality control, may well encourage cataloguers (or their libraries) to perform more *self-monitoring.*

We can obtain an indication of how common it is for catalogue records to be amended and/or upgraded by examining some statistics from the cataloguing performed on the SILAS

database in Singapore. SILAS (Singapore Integrated Library Automation Services) functions as the national cataloguing network and hosts a central database which contains the national union catalogue. When a cataloguer from a library edits a record already "owned" (i.e. shared) by one or more other libraries, according to the holdings information in the record, the edited record is automatically sent for review to SILAS staff. (If, on the other hand, a copy cataloguer does not edit the shared bibliographic parts of the record, but instead simply downloads the record as it is, into their dataset, after attaching their holdings statement and any other local fields, then the record is able to bypass the SILAS record reviewers.) For the month of March 2003, 5,073 records were copied without editing, whereas 3,617 records were copied with editing, according to the statistics generated by the SILAS system. On the SILAS database at least, we see that a very significant proportion of records (41.6% for March 2003) are not merely copied, but amended or enhanced (assuming that most of the editing either amends or enhances).

Consequences of errors and data omissions

Although some of the amendments and enhancements that are made to the shared records may also be made by the individual libraries to their own records over time, many would probably not be. It should not be hard for one to imagine some of the consequences of errors and less enhancement. Relevant records will not be retrieved, and sometimes non-relevant records will be retrieved. Additionally, records may be mis-selected, or mis-deselected, for ILL and document delivery, due to erroneous information, or absent information on the record. In some cases, typographical errors, or missing data, may lead to the non-retrieval of the only *pertinent* record for a known-item search. This might happen when the corrected record, which would otherwise be retrieved, no longer exists due to weeding in the library which has the catalogue with the corrected record, or when the corrected record represents an item which, in that particular library, is unavailable for loan, or already on loan.

Although it is true that if the corrected record(s) still exists and represents an item available for document delivery, then all is not lost, a slightly less efficient service may nevertheless result, since the uncorrected record(s) might represent libraries which would take priority, in the context of the document delivery system, over the library (or libraries) represented by the corrected record(s).

Certainly in the case of selection, a user of a distributed union catalogue may well examine only the uncorrected or non-enhanced record, and this may lead to a poor decision to use, or not to use, the document delivery service. For instance, a table of contents might otherwise have informed the user that there was indeed a very relevant chapter that made a document delivery request worthwhile; or perhaps the presence of an erroneous subject heading might persuade the user to try an ILL; or perhaps a false date of publication which is otherwise judged relevant might make the user de-select the item.

There are other "cumulative" effects of errors and omissions. The user comes to trust the database less. Also, errors and omissions are duplicated in catalogues which can result in further problems when retrieving a particular work or series. Such duplication occurs through "cloning." For instance, a pair of records, for the same edition of a work, exists in a distributed union catalogue, where one of the records has an error. That record, and with it the error, is then cloned in its own cataloguing module, since the library acquires a new edition. If the error affects retrieval, the user of the distributed union catalogue may find only the record

without the error, which is for the older edition. The user will probably not follow up to check for a newer edition.

Of course, when a shared record has an error which has not yet been spotted, the cloning of it may also lead to duplication of this error, and in such a case there is no record without this error. However, there is also a good chance that the error will be spotted in the copy cataloguing or in the cloning process, at least at some point down the line, and at the same time the other record(s) with the error would be corrected. The principle here is that the more times a record is touched, the more likely it is for any error to be amended.

Local data

There is another difference that might have a slight impact on the user, namely *local* bibliographic data—not holdings data, but other copy-specific information or catalogue-specific data, such as a note about the physical condition of a copy, or subject headings from an in-house thesaurus, or a URL which represents a subscribed site. In a distributed system, the user is given any local bibliographic data in the record straightaway; in a centralised or hybrid system, on the other hand, the record in the central database will be displayed first, and this may not include such local information.

Many catalogue records, however, do not contain any local data, and most records contain very little, in comparison with the amount, and importance, of their shared data. In any case, in a centralized or hybrid system, the user may still be able to see significant local data in a few moments, upon retrieving the local record. Regarding URLs not penetrable without authorisation, it can be made clear in the shared data whether the record represents an online resource requiring authorization. Moreover, this problem can be circumnavigated by including URLs together with notes about access rights in the central database record. Indeed, all "copy-specific" data can be displayed in the central database record along with the code for the library to which it pertains.

In sum, it is the view of this author that while there may occasionally be a slight inconvenience experienced on the part of a user when local bibliographic information is not initially displayed in a centralized or hybrid system, this should have little bearing, in most situations, on the decision about which type of system to adopt.

A case study: before and after a central database

SILAS hosts a central database which was initially established as the basis of a nationwide library automation project in the mid-1980s. As it grew, so did its second function, which was to host the National Union Catalogue (NUC). Libraries cataloguing online attached their holdings to the records in the database, and those records with one or more holding statements comprised the NUC. The union catalogue has been used to formulate ILL requests for many years, although there is still no formal, nationwide ILL system in place. One reason for this is that ILL is not always necessary when users can themselves travel between libraries relatively easily—no library in Singapore is much more than an hour's drive away. Nevertheless, some categories of users (e.g. academics) do expect a document delivery service, and are commonly provided with one.

Since the late 1990s, however, there has been a movement away from the centralised system and support for a distributed union catalogue, which has come to fruition under the name of

"Tiara," now subsumed under the National Library Board's "eLibraryHub." This service can also, of course, be used for document delivery purposes, providing Z39.50 access to the catalogues of several major Singapore libraries. With the implementation of this Z39.50 technology, some of the major libraries decided to suspend their contributions to the central SILAS database, and instead include their catalogues in distributed union catalogues, such as the one provided by Tiara.

This has led to a new situation where some libraries in Singapore share a bibliographic record on the central database if they acquired the item before the late-1990s, but do not share a record if they (or one of them) have acquired it since then. Instead, although they continue to apply similar bibliographic standards, the different libraries now catalogue their new materials separately, which means that in some cases they will create new records for the same item independently of each other, and in other cases they will derive different records from different sources, for the same item.

While the National Library Board, for instance, continues to use the central SILAS database for its copy cataloguing, and continues to contribute its new records to the central database, the National University of Singapore no longer derives any of its records from the central database, at least not directly, and instead derives them through other means.

The older records in the central SILAS database shared by the National Library Board and the National University of Singapore (and any other institutions) represent a bibliographic unification and, as such, a full consistency. Of course, this does not mean that they are necessarily perfect, but at least they do not allow for the chance of inconsistency, nor for the reduction of quality and quantity, nor for a higher level of duplication, that all may result from the use of a distributed model.

While the amount of duplication in the distributed union catalogue would very much depend on the specifics of a de-duplicating programme that was, or was not, in place, the other bibliographic differences between centralised/hybrid and distributed models would much more depend on the way in which the cataloguers in their respective libraries carried out their cataloguing, in the context of the particular system in which they operated. Hence it would be interesting to see the extent to which bibliographic inconsistency might occur in a distributed system, the number of record enhancements that would otherwise have been captured in a central database, and the number of errors that might have been corrected.

These facets of the bibliographic impact of a switch to a distributed system, lacking a central database, were thus examined in a brief survey of some post-centralised cataloguing performed by the National Library Board (NLB) and the National University of Singapore (NUS). A random sample of twenty items, catalogued originally by NLB in 2001, were identified as belonging also to the NUS collections. The NLB and NUS records for these items were retrieved and displayed in full format, and examined for the following features:-

1. Access points (fields) the same
2. Access points (fields) different
3. Non-indexed description fields the same
4. Non-indexed description fields different
5. Additional indexed fields
6. Additional non-indexed fields
7. Errors in one record but not in the other.

The bibliographic data in only those fields that might affect retrieval (on a standard system), was considered. Differences were discounted if they were judged as unlikely to affect either retrieval, or selection/de-selection for document delivery. For example, the difference between two MARC subfield codes might be ignored because of this; the difference of 1cm between the recorded heights of a book's spine would also be ignored. Again, additional data fields, found in one record but not the other, were only counted if they added value in terms of retrieval and/or selection decisions. Those errors which existed in one record but not in the other, and thus could have been corrected by one or other library, and were judged to be clearly errors and not perhaps caused by differing interpretations, were only counted as such if they were also deemed to represent a potential impact on retrieval and/or selection decisions.

Where there were differences in subject headings and the number of headings in the NLB and NUS records was different, the number of fields with different subject headings was recorded as half of the total number of variant fields. For instance, if there were three subject heading fields in one record, and four in the other, and none of them matched, then the number of different fields was recorded as 3.5.

From Table 2 below, we can see that over one third (35.2%) of fields providing access diverged in one way or another, and a large majority (80%) of non-indexed fields diverged in significant ways. A rate of almost two additional fields per pair of records were found to provide significant data, with a rate of over one additional indexed field per record pair. Of the 195.5 fields in the record pairs considered of potential impact on the use of a document delivery service, 97.5 (50.1%) were either inconsistent, or occurred in only one of the records. Any serious use of a union catalogue is bound to be impacted by such a content differential.

Examples of the divergence observed include:

(a) many subject headings were different, and in most cases the differences were not merely syntactic, but semantic; for example, additional subjects were represented on one record that were not so on the other
(b) three record pairs included different recorded dates of publication, which would affect a date-limited search
(c) two record pairs included different ISBNs; this might prove critical if the user retrieves on the record for the copy that is not available for ILL
(d) several names were not assigned their full authorised heading in one record, but were in the other, showing how the overall level of authority control may decline in a decentralised system
(e) one record included a part title, while the corresponding record did not
(f) in all but one record pair, the physical description field differed, for example, one record gave "chiefly col. ill." whereas the corresponding record gave no such indication that the item was essentially a picture book
(g) in one record pair, the edition statements varied (in substance), which might prove detrimental if the user bases their selection on currency, foregoing a more convenient ILL
(h) seven of the record pairs include a table of contents in one record but not the other; and three pairs include a summary in one but not the other; these fields make a record much more accessible, particularly for subject searching, and are especially useful for judging the utility of an item
(i) two records contain variant title fields, while their counterparts do not

(j) five records contain additional name access points, while their counterparts do not.

It may be noted that two of the record pairs (11 and 13 in Table 2) represented editions of a work for which an earlier edition was shared by NLB and NUS in the central database. Their rates of consistency were higher (6/8 and 6/9) than the overall rate for the twenty pairs (58.2%).

More than one error per record, on average, was found—often the records contradicted each other, so it is reasonable to assume that in many cases, one or the other record was correct, and the error may have been corrected (or never have arisen) in a database shared by both NLB and NUS. Even if only half or one third of these errors would not have resulted, this would still represent a significant number of errors according to typical cataloguing quality evaluations. For instance, an (ultimate) avoidance of one third of these errors would mean that the error rate would be reduced by four errors per ten records.

Bibliographic advantages increase according to number of catalogues

It is important to note that if the effect of the decentralisation of cataloguing amongst two library catalogues is significant, its effect amongst ten or twenty or more library catalogues is going to be all the greater: there will be more inconsistency, more non-enhancements and more errors.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	SUM
Access points the same	3	5	4	5	7	6	3	4	0	4	5	4	6	4	8	5	3	8	5	4	93
Access points different	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
Non-indexed description fields same	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	1	0	5
Non-indexed fields different	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	0	1	20
Indexed additional fields	1	0	1	0	2	1	0	2	3	2	0	2	2	1	1	0	1	1	5	1	26
Non-indexed additional fields	0	0	0	0	0	1	1	1	1	0	2	0	1	0	0	0	1	0	1	2	11
Errors in one record but not in the other	1	0	1	2	1	1	2	1	6	2	1	3	1	0	0	0	1	0	0	1	24

Table 2 : Divergences between records post-central database

Conclusions

While document delivery arrangements do exist between some libraries in Singapore, it is this author's view that a more comprehensive service could be put in place, and that such a service would benefit bibliographically from the central database that continues to be maintained by SILAS. Although there are many factors to consider apart from the purely bibliographic ones discussed above, we have seen from only a small sample of records how a central database may improve the use of such a document delivery service. However, there are two very important issues that would need to be addressed in order for most of this improvement to be realised. First, the National Union Catalogue of Singapore would need to be made available to end-users, whereas presently it is available only to staff of SILAS member libraries. Second, the central database needs to be updated, since some key libraries in Singapore ceased contributing to it several years ago.

Most interlibrary services are based on the premise that users' information needs from the different libraries overlap. As such, their bibliographic searching need not be catalogue-specific. That is, they may benefit from searching several catalogues simultaneously, for the same subject, author, work, etc. This paper has contended that users would thus be advantaged where these catalogues exhibit particular bibliographic qualities. Ultimately, one would want the different catalogues merged, so that a user retrieves only one record for each item. One record per item (or manifestation), instead of one record per copy, is a well-established cataloguing principle. Another such principle is consistency. Users would benefit from records for similar items containing similar content, to express these similarities. Users would also benefit from consistent styles of record content, so that it is easier to understand. Accuracy is another important bibliographic quality. Users would generally benefit from encountering fewer errors. And finally, another quality that is particularly important where the catalogue is used for more serious purposes—and this would certainly include where it is used as the basis of a document delivery request that costs both additional time and money—is that of more (useful) content. In other words, users would benefit from encountering as many enhanced records as possible.

It is the view of this author that a central database is likely to produce higher levels of the above bibliographic qualities than is a distributed union catalogue. There are at least four important reasons why this is so: (a) cataloguing into a central database ensures that record duplication is caused only by human cataloguer error; (b) it is easier to implement a quality control mechanism whereby records contributed to and revised in the union catalogue are reviewed, perhaps by a central agency; (c) cataloguers are likely to make greater efforts to adhere to agreed standards and policies, since they are more exposed to quality control and the judgements of fellow cataloguers; (d) records are more likely to get amended and enhanced through the work of other libraries.

References

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.