Clumping towards a UK National Catalogue?

Dennis Nicholson suggests that a distributed approach to the creation of a UK national catalogue is a potentially attractive option despite the difficulties it entails

Overview

This article presents a clumps-oriented perspective on the idea of a UK national catalogue for HE, arguing that a distributed approach based on Z39.50 has a number of attractive features when compared with the alternative physical union catalogue model, but also noting that the many difficulties currently associated with the distributed approach must be resolved before it can itself be regarded as a practical proposition. Dealing with these difficulties requires a mix of further research, some of which is scheduled to take place within existing projects, and - particularly in respect of data-based interoperability problems - additional local and national resourcing. However, it is suggested that the distributed model is sufficiently attractive compared to the physical union model to make the expenditure of additional time, effort and resource worthwhile. 'Dynamic clumping' based on collection level description and other appropriate metadata is seen as the key to user navigation in a distributed national catalogue. Large physical union catalogues like COPAC are assumed to have a role, although updating difficulties and the lack of circulation information may limit its scope.

Dynamic clumping: modelling a distributed national catalogue

In addition to Z39.50 compatibility, intelligent access to a fully distributed national catalogue incorporating every significant catalogue in the country requires a mechanism to reliably narrow the focus of user enquiries to a select few of the total number of servers in the clump. The assumption within CAIRNS [1] (Co-operative Academic Information Retrieval Network for Scotland) is that this mechanism is 'dynamic clumping' (a working demonstration of an early CAIRNS implementation of this kind of mechanism is available - see [2]). Dynamic clumping aims to aid the user by offering a database of subject-based collection strengths, each associated with at least one, but sometimes two or three, servers in the clump. The idea is that the user searches the database by subject, identifies the servers most likely to be of value in his or her search, then searches only the sub-clump, probably taking in other factors that will also reduce the number of servers (e.g. geographical factors, level of material required, language, and so on). This kind of mechanism is likely to be essential in a UK national catalogue based on a distributed model. It will not make sense, either in respect of a user's time, or network bandwidth, or local computing power, or gateway efficiency to search all of the catalogues in what will be a very large clump simultaneously. Dynamic clumping, backed up by active and ongoing collaborative collection management and development, offers a possible mechanism for reducing the number of servers to search in any given instance. This could work in at least two ways in a distributed UK catalogue. The first of these assumes either a single central collection strengths database or a small cross-searchable clump of these based at different regional gateways. This is probably the simplest model, and also arguably

has value in the context of inter-regional collection development collaboration. The problem with it at present, however, is that it assumes that each clump uses either the same or cross-compatible subject schemes to describe its collections. At the moment, this is not the case. However, work is now beginning under the auspices of the SCONE (Scottish Collections Network Extension project - pronounced 'scoon') [3] RSLP (Research Support Libraries Programme) project that could offer a solution to this problem by agreeing a common subject scheme and mapping it to other schemes such as the RAE (Research Assessment Exercise) headings [4] and the Conspectus [5] subject scheme .

The second approach is based on the assumption that regional clumps built around collaborative approaches to collection development such as planned by CAIRNS will:

- Probably want their dynamic clumping collection description databases to include descriptions of major key catalogues elsewhere in the UK (e.g. COPAC) or elsewhere in order to fill in known gaps in the total collection
- Will in the main have constituent catalogues in the clump whose coverage overlaps greatly with those in other regional clumps
- Will therefore only vary significantly from each other in respect of materials or perspectives specific to the region (e.g. CAIRNS will not only specialise in Scottish materials but will offer an environment within which the subject 'Law' (to take one obvious example) will tend to be assumed to mean Scots Law)

If this is true then each regional gateway will in effect offer national coverage at a general level, but with a particular regional slant. It would therefore be possible to envisage a comprehensive central gateway page for a UK national service offering a menu of regional gateways which would be presented as alternative national gateways (giving built-in redundancy). Users requiring a particular regional slant would be directed to the gateway for that region.

The advantage of this second approach is that it is more adaptive to regional requirements and does not seem to require anything major in respect of a central gateway. Further research is required to identify which approach offers the best results in terms of the requirements of all of the stakeholders, including, of course, the users.

Problems with the physical union catalogue model

As is made clear below, many difficulties will have to be resolved before either of these clumps-based models can become a practical working reality that meets the full requirement of users. However, the view taken by those who favour a distributed approach is that it is worth expending further time, effort and resource on, partly because it is felt that, given time and effort, the problems can be resolved, partly because it is felt that the alternative model of a physical union catalogue is at best a less attractive and less practical option that cannot, of itself, successfully meet the requirements of a UK national catalogue for HE.

The following is an admittedly clumps-oriented perspective on the case in favour of a distributed - as opposed to a physical union catalogue based - approach to the issue. If it has no other merit then, hopefully, it will at least provide a stimulus to debate:

Even if a comprehensive physical UK union catalogue for HE could be created and maintained, it is probable, and probably necessary and sensible, that individual organisations will continue to purchase, use, and catalogue onto, their own individual local systems. A range of factors are likely to ensure that this is so - political, funding body divides, the need to maintain local independence because of differing local circumstances (different computing and staffing environments, administrative differences, the need to compete as well as co-operate, and differing requirements generally), the tendering process, the likely temporal spread of replacement system purchases, and so on. This is likely even if the UK catalogue is only to be a catalogue of HE, as opposed to a catalogue for HE. If, as would seem sensible, it is to be a catalogue for HE, the retention of local systems becomes even more likely, because cross-sectoral and cross-domain concerns become additional factors (e.g. in CAIRNS, we are assuming researchers will require the inclusion of specialist collections held in public libraries and of museum-type collections as described in the SCRAN [6] (The Scottish Cultural Resource Access network) database).

This means that:

- The creation of a physical union catalogue as opposed to adopting a clumps-type approach is certain to involve big additional set-up costs and even bigger additional maintenance costs. The latter, presumably, going on forever.
- The creation of a physical union catalogue is certain to involve institutions in significant additional set-up work and costs and in some level of ongoing maintenance work and costs

These, in turn, mean that a clumps-based approach is:

- **More likely to be politically and financially acceptable to the vast majority of organisations** both within and, if applicable, outwith HE, in that it allows them to be independent in terms of their choice of local systems without - potentially at least incurring large and recurrent additional effort and costs that will be seen as simultaneously drawing funds away from local institutions and towards the centre, and adding to their own local costs and workloads. For example, given that libraries are already buying Z39.50 based web interfaces to their catalogues with clumping facilities built in, it is arguably the case that a small simple clump would involve very little in additional set up costs or additional work provided that the various clump standards had been agreed and published beforehand. Removing differences in cataloguing and indexing practices would involve work, of course, but the approach to this can be medium term and can be built into system replacement procedures. A bigger clump would, of course, require a dynamic clumping mechanism and an associated database of subject collection strengths. It would be difficult to entirely distribute this and so there is some central cost and effort involved in setting up and maintaining this. However, if organisations are to be involved in collaborative collection development programmes - arguably both a political and an economic necessity then setting up and maintaining the necessary database would be a task to be undertaken in any case and would not, therefore, involve additional cost and effort.
- **More likely to be sustainable**, in that the long term cost and effort required is likely to be much lower and to be necessary in any case for other reasons.

- **More likely to result in a comprehensive catalogue**, in that it is more likely to result in the inclusion of the catalogues of all relevant UK institutions, particularly if the view is taken that the catalogue must be *for*, rather than *of* HE and must therefore include catalogues that cross sectors and domains. The additional work and costs involved in 'joining' a physical union catalogue, together with other problems such as funding body divides arguably makes it unlikely that a physical union catalogue can ever be comprehensive. Arguably, it is also much more likely that regionally based clumps will identify and recognise the value of relatively unknown research collections in public and other libraries in their region and arrange for them to join the clump by helping to bridge any funding and political barriers that exist for the good of all of the people in the region. There is, moreover, a case for the view that the clumps approach is less likely to encounter such barriers. If an Organisation can join the clump simply by meeting the requirements and informing the other members, it may well be able to side-step such potential barriers.
- **More likely to offer an up-to-date service**, in that it is almost certainly the case that adding catalogue records and other information to the physical union catalogue will involve a delay, whereas a clumps-style approach ensures that the clump is always as up-to-date as the local systems are. Excellent though the service is in other respects, the example of SALSER [7](Scottish Academic Libraries Serials) is a case in point. The majority of libraries aim to up date it every three months but more often than not this period lengthens because it involves local staff in additional tasks that are not seen as high priority. It has not been uncommon for some sites to be six or more months behind in their updates.
- **More likely to offer circulation information**, in that most systems can now present this in 'opac' records sent to Z39.50-based webpacs and so can provide the information more or less immediately in a clumps environment, whereas this is either very difficult or impossible if a physical union catalogue environment where updates are something less than immediate and any circulation information that can be passed on almost certainly well out of date. One of the current CAIRNS gateways [8] reliably returns circulation information.
- **More likely to offer resilience at a lower cost**, in that a physical UK union catalogue could only offer an acceptable level of guaranteed service by having a very up-to-date mirror of the service available at a few hours, if not a few minutes, notice - unavoidably incurring huge additional set-up and maintenance costs, whereas the distributed nature of the clumps approach and the strong likelihood of overlapping coverage arguably makes a similar level of resilience almost free.
- **More likely to be a practical proposition**, in that all of the above points militate against the creation of a politically and financially acceptable, sustainable, comprehensive, up-to-date, resilient physical union catalogue with circulation information being a practical proposition and suggest that a clumps-based approach is much more likely to be practical. Moreover, it is easier to 'grow' a comprehensive national catalogue based on a clumps approach, in that organisations can join the clump by simply meeting the requirements and can be identified and encouraged to join not by one centralised body but by a number of distributed and geographically influential organisations

There is, moreover, an additional argument which says that, because of the different approaches taken in different sectors to things like record format (e.g. the use of GRS-I records in SCRAN in the museums sector), a single physical union catalogue cannot be comprehensive in any case, whereas (if the problems described below can be resolved) a clumps-based approach can - so that, arguably, the case against the physical union catalogue model as viewed from a clumps perspective, is not only that it has the many drawbacks detailed above but also that it cannot meet the need in any case, in that it cannot ever hope to be comprehensive.

Problems with the clumps-based approach

All this having been said, however, even the clumps projects themselves would admit that there are, undoubtedly, many difficulties associated with the distributed model, difficulties which must be resolved if the clumps-based approach is to become a practical proposition. Resolving them requires that additional time, effort and resources be expended on further research in some cases, and on tackling the interoperability problems caused by incompatible and/or incomplete data in legacy systems in others. The following list of problems associated with the clumps-based approach illustrate the point:

Cataloguing and indexing based interoperability problems

Amongst the sites represented within the CAIRNS clump are:

- Libraries whose whole stock is catalogued and others whose stock is only partially covered
- Libraries using UKMARC, libraries using USMARC, libraries using other schemes that map to UK or US MARC, and libraries using a mixture of these and other 'home-grown' formats
- Libraries using one subject scheme, libraries using other schemes, libraries using multiple legacy schemes, libraries using standard schemes with local variations and interpretations, libraries using no scheme at all - with similar differences evident in the use of class schemes
- Libraries using separate author, title and subject keyword indices and libraries offering combined keyword indices
- Libraries indexing two MARC fields in their author indices, whilst others index 6 or 9 or 12 fields, with similar divergent practices in other indices
- Libraries recording and indexing full author surnames and forenames, and libraries recording and indexing only surnames, with similar discrepancies in all indices
- Libraries using national and international authority file headings likely to be relevant in a national or international context and libraries using only local headings

The reasons for these differences are largely historical. The databases were developed, not with the aim of interoperating within a clump, but with the aim of serving specific local user groups, in unique local circumstances (including resourcing circumstances). The effect of the difference, of course, is poor interoperability - which is to say that the results obtained from searching the virtual catalogue are not as good

as they would be if you were searching one single coherent union catalogue with standardised data. For example:

- Zero hits in any given library on an author search can mean either that the library has no items by that author, or that it has but the items have not been catalogued yet, or that it has but that this particular library system will show author hits for surname searches only and show none if the forename is included in the search
- Zero hits in any given library for a subject search can mean either that the library has nothing on that subject, or that it has but has no subject index, or that it has a subject index but does not use that particular subject term, or that it has but that its older records don't have subject terms in them
- Twice as many hits in one library than in another on a title keyword search may mean that the library has twice as many relevant items, or it may just mean that the other library does not index as many potentially relevant fields

- not the kind of helpful results you would hope to get from a union catalogue, virtual or otherwise.

There are a number of points that should be noted about this state of affairs, however:

1. For the most part, the differences between the sites are either inherent in the catalogue data itself or, in the case of the indexing differences, are there because the sites in question have attempted to optimise access to materials for local users to help circumvent poor original data or low staffing levels. Any attempt to create a physical union catalogue to replace the virtual one would also have the same problem with data deficiency and would either have to:
   - Improve the data and then build better indices
   - Leave the data as is and cope with the same deficiencies in indices and indexing practice as the virtual catalogue
   - Leave the data as is and build the same indices for all sites but lose the optimisation at the sites with poor data

   In short, these problems are also problems for the physical union catalogue model

2. Although work is required to enable this, it is theoretically possible for a clumping gateway to get as good a result from a local catalogue as would be obtained through the local catalogue itself. If one site is known not to have a subject index and to normally offer its users a title keyword or class search as an alternative, together with advice on how to get the best results, then users of the clumping gateway can be given this information before a search, or in response to no hits from a subject search of that site. Even better perhaps, an automatic alternative search might be run by the system using synonyms if the user chose to do a subject search of the clump that included the site in question (not as simple as it sounds, admittedly). This approach would not solve every problem, but it could provide a valuable interim solution that would provide an acceptable level of service until the interoperability problems themselves could be tackled. CAIRNS plans to attempt to implement and evaluate

mechanisms of this kind during the year 2000, although it will also aim to produce proposals for resolving the base data problems in the longer term.

3. None of these problems with data and indexing are insurmountable. Given the will, the time, and the resources, they are all resolvable, although in some areas the resources required are significant. Many can be solved by rebuilding indexes or reformating data or changing record formats during a system replacement. Others might be tackled as part of retroconversions necessary for other reasons. The increasing necessity for institutions to engage in collaborative collection development initiatives and the encouragement to do so from programmes such as the RSLP is likely to increase pressure on individual institutions to solve such data-based interoperability problems. However, consideration might also be given to implementing a programme of national funding to help deal with some of the more costly problems in this area

Other interoperability problems

Other interoperability problems encountered in the CAIRNS clump and probably echoed elsewhere are:

1. The fact that it is sometimes necessary to send different Z39.50 attribute combinations to different servers in the clump in order to get comparable results and many of the Z39.50 clients available do not support this feature.

   This is not a significant problem in the sense that some Z39.50 clients do support the feature, which means that there are solutions available and that other Z39.50 clients should be able to incorporate the feature at some later date.

2. The fact that many of the servers in the clump send out UK MARC records but indicate to the Z39.50 client that they are sending US MARC records, a fact which can cause problems in respect of field displays if the client assumes and displays a US MARC field that is different in UK MARC (e.g. the field for ISBN)

   Again, this is resolvable in that it is only a programming fix. Moreover, it appears to be possible to design the Z39.50 client in a way that circumvents the problem.. It is not an ideal situation, however, and needs to be resolved by the suppliers concerned.

3. The fact that, currently, the two Z39.50 clients in use in the CAIRNS clump can't deal with all required record formats. CAIRNS wishes to incorporate SCRAN within the clump. SCRAN sends out GRS- I records. Neither Europagate [9] nor the Ameritech NT Webpac client used in the dynamic clumping gateway currently handles this format.

   This also appears to be resolvable in that:

   - It could be resolved by further programming in the clients in use in CAIRNS

- There is a product available called ZAP [10], produced by Indexdata, which appears to handle GRS- I as well as other CAIRNS formats. CAIRNS is investigating this product at the moment with the M25 [11]and SEREN [12](sharing electronic resources in an electronic network) projects.

4. Not all Z39.50 servers in the clump behave in exactly the same way, nor, sometimes do they behave precisely as the standard specifies. This obviously causes inter-operability problems unless spotted and circumvented.

   This is resolvable if the community can succeed in getting Z-client and Z-server developers to adhere to the sub-set of specifications from the Z39.50 standard specified in the draft Bath Profile [13]The various clumps projects are involved in the discussions about this profile and expect that, when finalised, it will play a key role in the eventual resolution of interoperability problems - although it will not, of course, deal with the data problems described earlier.

Questions about the dynamic clumping mechanism

The CAIRNS dynamic clumper [2] is a fully operational facility based on the RCO [14] (Research Collections Online) database of collection strengths in I I Scottish libraries. The subject scheme may appear to some to be unusual in that it is currently based on the Conspectus subject scheme, but any search or browse in the database will produce a dynamically generated sub-clump of CAIRNS libraries which can then be sent a broadcast search and the mechanism would also function with any other subject scheme. This shows that dynamic clumping works at a trivial level - that is, it is possible to use a database of subject strengths to reduce the number of services in the clump offered to the user for searching simultaneously.

Critics, of course, will argue that many questions about the mechanism remain unanswered, and this is true. Further research is required on a number of issues, including, but not necessarily limited to, the following:

1. The navigational effectiveness of the collection strengths database

   Clearly, it narrows down the number of servers to search in an apparently sensible fashion, but does it do so effectively? Are the servers the user is presented with his or her best option or, failing that, his or her best initial option for searching? The logic of the idea appears sound enough. Users looking for items in a particular subject area are perhaps not guaranteed that they will find what they need in catalogues where the institutions are strong in that particular subject area but the probability is that they are more likely to find it in these than in others. Moreover, it is reasonable to assume that as libraries begin working together on describing their distributed joint collections in ways that will best help the user, the dynamic clumping mechanism will gradually become more refined and better able to aid user navigation. It is undeniable, however, that little is currently known about the effectiveness of the mechanism. No tests have yet been carried out, although such tests are planned, both within CAIRNS, which does not complete until December 2000, and within the SCONE RSLP project, which runs till late

2001. What can arguably justifiably be said is that the mechanism *can* be effective. Given good and sufficient data about the users and their needs, good and sufficient data about the collections and their strengths and other characteristics, cross-compatibility of user and collection data, and facilities which allow users to accurately match needs against collections, there can be little doubt that an effective navigational tool can be built. The problem is whether it is possible to reliably and sustainably collect good and sufficient data about users and collections, but particularly about the latter, a question addressed at 5 below.

2. The compatibility of collection strengths data across Scotland and the UK

    Currently, the RCO data is based on the Conspectus subject scheme and was collected using the Conspectus methodology for measuring subject strengths adapted for Scottish use. Other clumps have their own methodologies and their own subject schemes. Under the current circumstances, therefore, an effective dynamic clumper operating across the UK is not a feasible proposition. Moreover, although it is true that the Conspectus subject scheme and versions of the methodology have been used elsewhere (Australia, for example), it has become fairly clear that this approach does not have wide acceptance across either Scotland in particular or the UK in general. It is also, being originally based on the US oriented LC subject scheme, not likely to be widely accepted by UK users. This problem has been recognised and agreement has been reached in principle on a way forward on a common subject scheme and, within Scotland, on a way forward on investigating the methodological question. As with 1 above, it reduces essentially to the question of reliably and sustainably collecting good and sufficient data, the issue dealt with at 5 below.

3. The question of whether or not the dynamic clumping mechanism will scale

    Granted that the mechanism works in the current implementation, reducing 11 servers to (usually) 4 or less, how will it cope with 100, 200, 400 servers or more? This issue also requires further research, some of which will be conducted within the SCONE project. Again, however, it arguably reduces to the question of reliably and sustainably obtaining good and sufficient data dealt with at 5 below. If 3 or 5 or 10 servers is regarded as the optimum number for a dynamically-generated sub-clump, then it is feasible, given sufficiently good data and data structures, to design the system so that it will only produce the optimum number or less, recognising:

    -   That this is a navigational mechanism designed to guide rather than give one comprehensive definitive result
    -   That in any given case, the sub-clump offered would be the first step in an ongoing strategy. If it failed to meet the user's needs, the next best sub-clump would be offered (e.g. libraries with weaker but still significant strengths in the area concerned)

4. The problems associated with the fact that subject schemes in different libraries are different and that all differ from the subject scheme used in the current dynamic clumper

Even if the current subject strengths database is a reliable way of accurately focusing the users attention on those services most likely to be of relevance to their needs, there is currently no direct link between the subject terms used in the RCO database and the items in the source libraries identified in RCO as strong in a particular subject area. The libraries in the clump do not subject index the items in their databases using the Conspectus subject scheme. Those libraries that do use subject schemes, use schemes that differ from the Conspectus scheme and from each other's schemes, and some libraries do not subject index at all. This does not mean that no useful work has been done in identifying the libraries concerned as being those most likely to be most useful to the user. This may still offer a useful outcome in respect of the resulting sub-clump and, having identified the libraries, the user may not wish to search them by subject in any case, but by author or title or ISBN. Nor does it mean, necessarily, that retrieval by subject from these libraries is impossible. Different strategies and terminologies may be required for different libraries and, in some, title keywords may be the only option. Accurate and comprehensive subject retrieval from the sub-clump will be difficult - although not essentially more difficult than in the individual catalogues themselves - but it will not be impossible. Once again, however, the situation as it currently stands is far from ideal, and, once again, the accuracy and reliability of the data - the topic covered in section 5 below lies at the root of the problem.

5. The problem, alluded to in 1-4 above, of reliably and sustainably collecting good and sufficient data on collections and their strengths and on users and their needs

    Some of the work required here is scheduled within CAIRNS, which will seek to evaluate the existing user interface and RCO database with a view to improving it early in 2000, and within SCONE, the associated SOEID (Scottish Office Education and Industry Department) project, and the increasingly important, cross-sectoral PAIRTS [15](Public Access to Information, Research and Teaching in Scotland) initiative, which between them will look at:

    - Extending the existing RCO database to include more sites and services and different types of collection (e.g. datasets)
    - Examining alternatives to the Conspectus methodology for measuring collections and their strengths '
    - Interfacing the database with collections data from Scottish public, special and other libraries collected by SLIC (Scottish Library and Information Council) and made available via the SLAINTE [16] service
    - Mapping the Conspectus subject scheme to other schemes such as those used by the M25, RIDING [17] and Music Libraries Online [18] clumps, to RAE headings, to the work of NGFL (Scotland) and, in particular, to the UK-oriented but Dewey and LC based BUBL [19] subject scheme, the aim being to produce a common high-level subject scheme that it is hoped will be widely adopted across the UK

It is possible, if unlikely, that this work will resolve all outstanding issues with regard to the problem of reliably and sustainably collecting good and sufficient data on collections and their strengths and on users and their needs. It may, for example:

- Show that the navigational effectiveness of the existing collection strengths database is adequate to the task of guiding user activity successfully in a distributed catalogue
- Provide an accepted standard approach to the measurement and description of collection strengths data across Scotland and the UK (either by validating the Conspectus approaches or offering something better)
- Provide, through the addition of SCONE, SLAINTE (Scottish Libraries Across the Internet) and SOEID data a big enough database to prove that the approach will scale
- Either show that the discrepancy between the central and local subject schemes does not appreciably effect the navigational effectiveness of dynamic clumping or offer an alternative subject scheme that institutions will agree to add to new records added to their databases (so that, in time, the central and local schemes will be the same)

It is, however, more likely that it will only answer some or some parts of these questions and that it will result in the formulation of a set of additional questions or a refinement of the existing ones, with the following being some examples of questions likely to require further research:

- Who are the users or user groups that a UK national catalogue will have to serve?
- What specifically are user requirements in respect of a UK national catalogue?
- Do they add up to a need for a single UK national catalogue, whether virtual or physical, or simply to a list of functions that might be served by a number of function or user-group specific gateways operating in a distributed environment?
- How many servers are there likely to be in a comprehensive UK national catalogue and how, given this, can we establish whether or not the dynamic clumping approach scales?
- In what circumstances does the collection strengths database provide good results and in what circumstances are they less good and what can be done to improve the areas where the results are poor?
- Is collection strengths data sufficient in itself to provide navigational effectiveness or is additional data required?

Performance issues associated with the distributed model

In a physical union catalogue, a user's search is run against the database only once, and is run using central computing power, so that it does not require additional memory, processing power and disc space on local machines. In a distributed system, the same search is run several times against some or all of the databases in the clump and does, presumably, require more in respect of local computing resources. Thus,

while the distributed approach appears to reduce costs by making an additional central catalogue unnecessary, there is also a reduction in efficiency which may result in a requirement for additional local computing resources and associated additional costs in that respect. A number of questions here require further research, for example:

- How do the additional costs of local computing power compare with the cost of an additional central system and associated recurrent costs?
- How do the benefits of one or other affect the overall picture of costs against benefits?
- Logically, there will probably be an increased load on local systems, but is this significant in practice?
- Can any such increased load be reliably measured and predicted?
- Can any such load be minimised by an efficient dynamic clumping mechanism?
- Will local sites benefit from increased local computing resources themselves?
- Are there identifiable circumstances in which performance issues indicate that a distributed approach is safe and others where there would be a case for, say, a limited union catalogue which gathers circulation data from local systems once items of interest have been chosen?

Further research and discussion is required in these and other areas if the full significance of performance issues is to be understood.

Conclusion

In summary, then, the clumps perspective on this issue (at least as interpreted by this author) is as follows:

1. A UK national catalogue based on a the physical union catalogue model is not an attractive option. It not only entails significant additional capital and recurrent expenditure and additional ongoing effort from institutions, making it unlikely that it will ever be politically or financially acceptable to most institutions, it also has a range of other drawbacks. For example, it is always likely to be out of date, is unlikely ever to include useful circulation information, does not offer low-cost resilience, and can never offer comprehensive coverage that crosses sectors and domains.
2. As a model, the distributed approach is a more attractive alternative. However, it too has a number of associated difficulties which must be resolved before it can be regarded as a practical proposition on a UK-wide scale: the interoperability problems, navigational and scaling problems and performance issues outlined above
3. Resolving the problems with the distributed approach requires both additional local and national resourcing to resolve interoperability problems caused by incompatible and incomplete data and additional research. Those who favour the clumps approach take the view that the distributed model is sufficiently attractive when compared with the alternative of a UK-wide physical union catalogue to make it worth further investigation and effort.

Whether this perspective is the correct one remains to be seen. Hopefully, this contribution will at least occasion lively debate, and that will lead us all a little closer to enlightenment!

References

1. The CAIRNS main web site is at: http://cairns.lib.gla.ac.uk
2. The CAIRNS dynamic clumper is at: http://wp338.lib.strath.ac.uk/cairns/dynatop.htm
3. The SCONE project proposal is at: http://wp338.lib.strath.ac.uk/scone/sconebid.htm
4. For further information on the RAE and RAE headings (units) see: http://www.niss.ac.uk/education/hefc/rae2001/
5. For further information on Conspectus see the articles at: http://bubl.ac.uk/org/scurl/rcoabout.htm
6. SCRAN is at: http://www.scran.ac.uk/
7. SALSER is at: http://edina.ed.ac.uk/salser/
8. The CAIRNS Ameritech gateway is at: http://130.159.82.15/webpac/wgbroker.exe?new+-dbselect+/
9. The Europagate site is at: http://europagate.dtv.dk
10. The ZAP site is at: http://www.indexdata.dk/yaz/
11. The M25 clumps project is at: http://www.M25lib.ac.uk/M25link/
12. The SEREN project is at: http://seren.newi.ac.uk/user/seren/
13. The Bath profile is at: http://www.ukoln.ac.uk/interop-focus/activities/z3950/int_profile/bath/draft/
14. Research Collections Online is at: http://scurl.bubl.ac.uk/
15. For further information on PAIRTS see: http://www.slainte.org.uk/Pairts/pairts.htm
16. SLAINTE is at: http://www.slainte.org.uk
17. The RIDING clumps project is at: http://www.shef.ac.uk/~riding/
18. The Music Libraries Online clumps project is at: http://www.musiconline.ac.uk/
19. The BUBL Information Service is at: http://bubl.ac.uk/

the next CLUMPS event is: **Library Resource Sharing and Discovery: Catalogues for the 21st Century**. This is a one-day workshop (two locations, London and Glasgow) presented by the eLib Clump Projects and co-ordinated by UKOLN. The London event is on March 3rd, and the Glasgow event happens on 11th April. Further details are available at: http://www.ukoln.ac.uk/events/elib-clumps-2000/intro.html

Author Details

Dennis Nicholson
Director of Research (Directorate of Information Strategy)
Centre for Digital Library Research
University of Strathclyde