

STANDARDS FOR SERIALS: BUILDING THE BASIS FOR DISTRIBUTED ACCESS

Rosemary Russell

The article discusses a range of issues related to standards for serials, based on the seminar held in June 1997 and the reports presented. This work was recommended by the first MODELS workshop which identified a number of serials management questions. The discussion takes place within the context of the MODELS project's distributed systems framework which is developing blueprints for more effective management of resources.

Rosemary Russell is MODELS Project Manager, UKOLN (UK Office for Library & Information Networking), ULCC, 20 Guilford St, London WC1N 1DZ
Email: r.russell@ukoln.ac.uk

MODELS background

MODELS¹ (MOVing to Distributed Environments for Library Services) is a UKOLN initiative which is supported by the Electronic Libraries Programme (eLib) and the British Library Research and Innovation Centre (BLRIC). It is developing a framework to manage distributed library services effectively. The project is achieving this through discussion of both strategic and practical issues at MODELS workshops, and developing technical service models based on the issues identified. MODELS has achieved some important results for the management of existing and emerging distributed services, through its recommendations and follow-up work. The clumps² initiative for example, emerged as an outcome of the third workshop.

MODELS Phase I is being completed during the last quarter of 1997, with Phase II commencing at the beginning of 1998. A series of linked workshops form the core of the project, with five held during the first phase, and another three to be held during the second phase.

One of the themes in the first phase of the project was article discovery and request, and a workshop held in December 1995 was based around this subject. The workshop made a series of recommendations for further work which would take service development forward. This included a group of studies on standards for serials³, together with a seminar, to be funded by eLib. The aim was to examine existing standards and assess suitability for use, together with options for further development. UKOLN issued calls for proposals and commissioned the work. A consolidated list of studies, with links to draft and completed reports is available from the MODELS Web pages⁴. This work on standards for serials is contributing to the overall MODELS distributed systems framework.

Standards studies seminar

The recommended follow-up meeting was held on 2 June 1997 at the University of London Computing Centre. Around 50 people

participated in a seminar on standards for the management of serials, organised by UKOLN and BIC (Book Industry Communication). The aim of the meeting was mainly to present preliminary results of the eLib supporting studies on standards for serials; it provided an opportunity to gather feedback on the results and to discuss issues relating to use. The meeting was chaired by Brian Murphy, Director of Academic Information Services at Queen Mary and Westfield College, University of London. Participation was by invitation; representatives were invited from the range of sectors involved: publishers, subscription agents, secondary information services, library managers and systems suppliers.

The trouble with serials... management and metadata problems associated with format

Serials management has always been problematic for library services and the range of agents involved in the information/supply chain. Serials are complicated by their nature, meaning additional difficulties in moving to a distributed environment. Clearly, part of this is due to their publication in parts, together with further division into articles. Physically it is more difficult to catalogue and control multi-part publications (an advantage here for electronic journals) and inevitably parts go missing. This can be at various stages and involves overheads for a range of parties (the number of claims to agents/publishers for non-receipt of journal issues tend to be very high). Permanently missing issues should be recorded in the catalogue, but often this is not carried out and catalogues may not accurately represent actual library holdings.

Another typical problem is title changes: journals have a tendency to change title, sometimes several times, during their publication life. Alternatively, they may cease publication but be replaced by another journal. As a result of these frequent changes, library catalogues sometimes maintain several different records, with associated check-in data, for the same journal. Varying practices for abbreviating titles also cause problems for catalogue maintenance.

There is also the tendency for issue sequences and frequencies to change. Libraries' serials

management systems are not always helpful in facilitating the necessary parameter alterations, to support this and the other changes mentioned.

The problems highlighted in this section serve to illustrate some of the intrinsic difficulties associated with serials transactions, all of which have consequences for serials metadata. While inconsistencies within individual catalogues and databases are frustrating, it is often possible for human users to compensate for this and still access the information they require. However, within a distributed services environment, where systems rely on accuracy and consistency of metadata in order to interwork, it is clearly a far greater problem.

Library management systems are increasingly having to cope with a mixture of print and electronic resources, the so-called 'hybrid library'. While many of the physical management problems may not be present for electronic journals, many of the 'connection' issues remain, since users still have to *discover* relevant articles, even though location and requesting may not involve the same difficulties. (There is of course an additional range of access/licencing and other issues associated with electronic journals, which are not directly relevant to this discussion.)

'Unconnectedness' of services: the need for standards

Currently users have to connect separately to a varied range of services with different interfaces in order to discover, locate and request individual articles. There is no service framework to link them together and provide a consistent user interface. As mentioned, systems depend on manual intervention to provide the necessary links between them. Library catalogues describe serials at title level only. Abstracting and indexing (A&I) and table of contents services, on the other hand, describe the article content of individual issues. Therefore, a user carrying out a subject search for an unknown item may firstly use an A&I database on CD-ROM to discover an article of interest. In order to locate a copy of the journal issue locally, he/she will then have to write

down or print out the relevant bibliographic details and retype the information again into the local library OPAC. If the library does not hold that title (or individual issue - which may only be discovered by visiting the shelves), the user may want to request an inter-library loan from a regional service. Usually this will have to be done by library staff, who will input the bibliographic details yet again. A fourth repetitive stage may be involved if the item is still unavailable and the user/library staff member finally has to request the item from a document supply service. Even within this stage it may be necessary to repeat the search several times, in order to compare price and other terms of availability across several suppliers.

Therefore, there is usually no way of *directly* telling which articles are available in one's local library. However, on a proprietary level there have been some initiatives to provide the necessary links between contents databases and library holdings, eg. the Swets facility which allows their table of contents data to be downloaded onto a local server. It should also be acknowledged that there are some standards-based developments in this direction, such as Sirsi's *InfoView* client, which links results from Z39.50 searches of A&I databases back to local library holdings, to check availability. While these initiatives are helpful, there is a clear need for a standards-based framework, supported by implementors agreements, to manage the complete range of transactions.

Identifiers

The strong need for a unique identification scheme emerged at the MODELS 1 workshop, and in the context of serials, the SICI⁵ (Serial Item and Contribution Identifier) was identified as particularly relevant. Since it was clear that relatively few people were aware of identifiers in general and of SICI in particular, MODELS 1 recommended the organisation of an awareness-raising seminar on unique identifiers. This took place in March 1997 and was reported in *Ariadne*⁶. However, identifiers play a key role across distributed serials management, and are central to the entire standards discussions.

Descriptive standards for serials metadata

The studies on descriptive standards for serials metadata and standards for terms of availability metadata (described below) were jointly undertaken by David Martin and Mark Bide on behalf of Book Industry Communication. Given the close relationship between the subjects, they were carried out as one piece of work.

The definition of serials metadata adopted for this study was *data in machine-readable form which identifies a journal title, a journal issue, or a journal article*. Content description, subject access and other elements, which might be included in a journal article database, were outside the scope of the study.

The research verified that there is a very wide variety of practice in recording serials metadata. The only two elements on which there turned out to be broad agreement were the title as printed on the journal and the ISSN. However even the title as printed on the journal may not be an unambiguous identifier; there are several instances of different serials with identical titles. Only the ISSN and its derivative identifier, the SICI, have the necessary characteristics to meet the requirements for unambiguous identification of serials at the title, issue and article level. Although SICI is an emerging standard, it has significant support and usage is expected to grow.

The study identified several metadata standards for describing serials including ISDS data elements, MARC, SGML (SSSH - Simplified SGML for Serials Headers) and Dublin Core. It emerged from the survey that the only common denominator across existing practices was the use of the ISSN. Reaching agreement on a standard set of serials descriptive metadata elements is perhaps unrealistic. Instead, it is recommended that the eLib Programme should consider adopting a set of serials data elements, compatible with generating and matching SICI codes, and developing a recommended format for on-screen metadata display.

It is believed that standardisation will almost inevitably lead to *some* loss in the richness of data. However, although standards for serials metadata need to be compatible with lower-level metadata standards, these are not an adequate replacement for richer formats. Since

rich metadata is being produced by an increasing number of serials publishers, in the form of article 'headers', it is recommended that the SSSH should be adopted for communicating rich serial metadata sets through the serials information chain.

The authors conclude that an absolute priority is a standard coded identification for serials articles to enable communication between systems. A key recommendation, therefore, is that the eLib Programme should adopt SICI version 2 as the unique identifier for serial articles. In addition, the ISSN and the ISDS Key Title should be standard for the identification of serial titles. To support this, they recommend that steps should be taken to make the ISDS database available as an authority file to all potential users in the information chain, from publishers through to libraries. It is significant that the EC CASA⁷ (Cooperative Archive of Serials and Articles) project is investigating more effective use of the ISDS database in a networked environment.

Standards for terms of availability metadata

As indicated above, this study was undertaken by David Martin and Mark Bide on behalf of Book Industry Communication, and carried out at the same time as the preceding metadata study.

The definition of Terms of Availability (ToA) adopted for the study was:

Data accessible to customers or users which includes any or all of: financial terms of sale; copyright and other conditions of use; and terms for interlibrary loan, photocopy supply or downloading in electronic form.

The survey revealed a strong recognition of the need for international standards. ToA standards are needed primarily for three reasons: interoperability between systems; consistency in human-readable interfaces; interactions between the two (the latter increasingly depends on the former).

However, the study established that there are few existing standards for ToA data, with little implementation. Although EDI transaction messages have highly structured data elements for price, there are none yet for other conditions within the study's terms of reference. The other

families of standards examined were MARC and SGML. Within MARC facilities for recording ToA are limited; for example UKMARC makes little or no explicit provision for terms other than price, unless it is entered as free text into a notes field! This informal approach means that it has little value in system interoperability.

The authors believe that the issue is more systems than standards-related: where the terms of business are well understood and standardised (eg in book sales) EDI standards to support ToA information are well developed. However, it is argued that the development of ToA standards for electronic resources is much more complex. Since there is no broad agreement on standards for the terms of business for licencing electronic resources, it is thought that the development of a ToA standard would be premature. The link with the thorny problem of user authentication is also recognised. Given the continuing work on standardisation in these areas, it is instead recommended that the development of ToA standards should be kept under review.

Standards for serials holdings

Alan Hopkinson of Middlesex University carried out the study on standards for serials holdings; the work was partly based on an earlier briefing paper produced for MODELS 1. The term holdings is used in the study to refer to the metadata representing those parts of the serial which have been acquired by the library or libraries covered by the metadata.

Library practices for recording serials holdings developed when all catalogues were manual, so consistency across libraries was not then an issue. Now that libraries are increasingly looking to share resources, standard structures and levels of detail for recording holdings are of key importance, so that other libraries and external users can easily discover what stock is available. This standardisation is needed both to enable interworking at a systems level, and to provide consistent human-readable information. As discussed above, therefore, standards for serials holdings are needed to enable the unique matching of bibliographic details retrieved for

articles of interest, with a library's holdings. Matching is needed firstly at the title level and then at part level (volume or issue). This study has dealt specifically with the object of a search being a *serial part*, since the Martin/Bide study, discussed above, has covered article records.

From the beginning of the chain, a variety of data formats exist: examination of a range of abstract and indexing database producers on CD-ROM and online revealed a surprising number of variations in recording article 'host statements'⁸. Libraries too reveal a range of formats for recording holdings data. The format used may be partly dictated by the library's serials management system. However, even within individual catalogues a range of formats often exist; this may happen over a number of years as staff practices change and previous styles are not converted. The Anglo-American Cataloguing Rules (AACR) give little guidance apart from a couple of examples for recording holdings statements. Moreover, this rule was only introduced into the 1988 revision of AACR2 - nothing was provided prior to this. ISO 10324⁹ provides an extension of the AACR rules, but does not appear to have been implemented yet in the UK.

A further problem is that since serials have traditionally been catalogued separately from monographs, automated serials control modules have tended to develop separately from library management systems. In recent years there has been more integration, with many systems using the 'check-in' data to create a summary holdings statement, available via the OPAC. However often the automatically-produced statements will not mention missing issues - these may have to be added manually. Additionally, on the basis that serial parts are often not loanable, many systems do not provide fields for recording availability status. In a distributed, resource sharing environment, these are fundamental issues. Different management and system approaches are both clearly required. Again, detailed check-in data in a standard format is needed to support both human decision-making (library staff and end-user), and system interoperability.

The study concludes that the only way to ensure successful uniform matching and retrieval of serials parts data across catalogues,

is for data to be created to a prescribed standard in form and completeness. Provided that a standard could be agreed amongst all the relevant parties, this would be a feasible solution for creating new records. However, existing serials parts data would be a very significant problem. Conversion to a standardised format would be a huge task. It is unclear to what extent algorithms could be utilised to aid conversion, although it seems likely that a large amount of manual effort would also be required.

Hopkinson suggests SICI and MARC as mutually exclusive options for dealing with this standardisation. The first alternative is that SICI could be adopted as the format for the serial item and contribution, although it is recommended that a 'SICI look-alike code should be developed for the format of data for consolidated or compact holdings'. It is suggested that a hyphen could be introduced to indicate spans, with no provision for recording omissions; instead a gap would be indicated by beginning a new span.

The second alternative is using MARC. The author points out, however, that MARC is not in fact a formal national or international standard and there are many different versions (UK MARC, US MARC, OCLC MARC etc.) in use within the UK. Since UK MARC has very limited provision for recording holdings data and US MARC covers a multitude of different circumstances, an option would be to add the relevant US MARC fields and sub fields to UK MARC.

In a distributed environment, where users are carrying out searches of remote catalogues using Z39.50, conformance to profiles is needed to ensure that different libraries are implementing the protocol in the same way, and that search attributes are mapped to the correct fields. Therefore, agreement would need to be reached on a serials profile to ensure systems interoperability.

Standards for document requesting

The document requesting study is co-ordinated by UKOLN, with contributions commissioned from standards experts. It has been examining existing standards approaches to the request

process and sketching the environments in which they are used. For the purposes of the study, document requesting is taken to be the issue of a request for the delivery of an identified journal article. The item may be in print or electronic format and a print article may be returnable (a journal issue) or non-returnable (a photocopy). We are concerned with the entire lifecycle, from a request being placed, to delivery of the item. It is probable that the requesting procedure will involve a series of messages between the requester and the supplier, and will often include an intermediary. Different models will be needed, to cater for different service environments.

In the UK context increased resource sharing is being encouraged from a number of quarters - for example several eLib projects are specifically working towards better exploitation of journal literature and the Anderson Report has made recommendations for methods of improving access to research collections. It is likely that libraries will be using a much wider range of ILL and document delivery providers in future; a standard means of communication to enable interoperability between systems, without re-keying data, is essential for effective management.

The four key 'standards' that are used currently for requesting were identified as HTTP 'GET', ISO ILL protocol, Z39.50 item order and EDIFACT.

Looking, firstly, very briefly at 'GET', this is part of the HTTP protocol developed for accessing documents on the Web. The client sends a GET command for an object named in a URL; after the server transfers a copy of the requested object to the client for display, the connection is closed. Therefore, if the user wishes to return to the server to make a further request, it is necessary to open a new connection and possibly repeat some steps. It was of course designed for a straightforward process and, although improvements (such as HTTP 1.1 developments to keep connections alive) are overcoming some deficiencies, GET is clearly a very basic method of requesting compared to the other specialist protocols examined.

Ruth Moulton of Fretwell-Downing was commissioned to carry out an investigation of

the ISO InterLibrary Loan Protocol (ISO ILL) for the study¹¹. ISO ILL was developed to provide uniform procedures when accessing a library across a network to order copy or loan material, and for carrying out the administrative tasks involved in loan management. It does not, however, provide any services to search for or locate items, nor does it cover document delivery services; although some facilities for billing and accounting are provided, these are fairly simple. The ILL messages exchanged between the requesting and supplying libraries are not designed for direct human consumption but rather for interpretation by the ISO ILL software that will keep track of the state of the ILL transaction and send replies, as and when necessary. The ILL software is likely to be part of a larger management system which would have its own user interfaces.

An ILL transaction starts when an ILL-Request message is sent. One ILL transaction may last for a few seconds, if the supplying library has electronic access to its catalogues and is supplying documents held electronically over the network; on the other hand it may last for months, if a loaned item is renewed.

Development of the profile began in the early 1980s, although the first edition was not published by ISO until 1993, and there are still relatively few operational services. The leading developer has been the National Library of Canada. It has also been implemented by a number of development projects and interest in the protocol has rapidly increased in the last two years - there is now an active international implementors group called IPIG (ILL Protocol Implementors Group). Many organisations such as OCLC are now implementing ISO ILL. The British Library has been investigating the provision of an ISO ILL interface into the ART system and it is expected that this service will be developed. It is notable that relatively few library management system suppliers are implementing the protocol, although an exception is Ameritech which is developing a stand-alone system, designed to work with Horizon, Dynix, or any other library system.

An overview of Z39.50 item order was provided by Denis Lynch of SilverPlatter. The Z39.50¹² Search and Retrieval protocol is an

American National Standard that has been adopted as an ISO standard and is in wide use in the library community. Z39.50's use to access library catalogues and abstract & index databases leads naturally to inclusion of facilities for requesting delivery of items. The Z39.50 protocol makes provision for non-core services (such as ordering items) within a single service called 'Extended Services'. These non-core services involve operations that extend beyond a single session. An item order request will normally be sent after the desired item has been identified in a Z39.50 result set. The item order message is modelled on the ISO ILL message, but it is not completely identical. Requests are translated into the commands appropriate to the specific ILL system in use and this system takes over the process. Item order does not support any messages beyond the initial request for the item — it can only interrogate the status of an active request, or possibly delete or modify the request.

The only production use of item order so far is OCLC's FirstSearch, which will accept item order with the ILL Request PDU. The request is then forwarded to OCLC's own ILL system. The second known implementation is AT&T's Library System, which has an experimental implementation of item order that is available for interoperability testing.

Finally David Martin of Bookdata covered EDIFACT, the principal international EDI standard for electronic trading communication across all business sectors. It is widely used, so there is a range of software available. EDItEUR¹³ has prepared guidelines specifically for three serials applications: despatch and claims; renewals and quotations; order and order responses. These are clearly designed primarily for subscription trading between libraries, agents and publishers. A number of library management system suppliers have implemented or are implementing EDIFACT messages as part of their acquisitions module; these include BLCMP, Innovative, Ameritech, SIRSI and others. Dynix users, for example, are using EDIFACT messaging for orders to Dawsons. Although it has not been used for document requesting, it has a possible application.

It is notable that the ISO ILL protocol has two encoding options for messages, one of which is an EDIFACT format (albeit based on an outdated version of the syntax). Martin suggests that traditional purchase ordering and ILL could be converging to form a single continuum of document request, for which one messaging standard should be used if possible. It may, therefore, be worth reviewing the EDIFACT format option within ISO ILL to bring it up to date.

Preliminary conclusions point to the desirability of some form of integration of document requesting protocols¹⁴. For example, Z39.50 could be used as a way of transporting ISO ILL messages; it could also be used to encode and transmit EDIFACT messages. At a different level, Web to Z39.50 gateways are commonly used for access to many Z39.50 resources. However, given the current low level of use of ISO ILL and Z39.50 item order, more implementation experience is needed before integration can sensibly be taken forward. The British Library's implementation of ISO ILL is likely to provide impetus for the wider community. In addition, the M25 pilot clump project¹⁵ is focusing on access to a deeper level of serials holdings information.

The ideal scenario: distributed article discovery, location and request

To return to the earlier 'connectedness' discussion, the ideal scenario is for a user to be able to move seamlessly between discovery, locate and request services. For a journal article, this may involve discovering an item of interest (perhaps from the consolidated results of Z39.50 parallel searches of several A&I databases), then moving directly to the local library catalogue to check firstly if the title is held and then (via detailed, standardised holdings information) if the issue is available. If it is on loan (and the date of return is too late), the user will be able to transfer details directly to a relevant document delivery service, such as UnCover, and request the article using item order; based on the user's required speed of delivery and budget, a choice of services may be offered.

In order to achieve these connections between services, consistent metadata needs to be made

available for exchange and use by underlying systems. Implementation of the standards and protocols discussed in this text will help to build a distributed systems framework and enable us to move towards a connected serials environment.

References

1. Further information about MODELS is available from: <URL: <http://www.ukoln.ac.uk/models/>>
2. Clumps information is at: <URL: <http://www.ukoln.ac.uk/models/clumps.html>>
3. The term serial is inclusive of journals, newspapers, periodicals, annuals etc. However the current text is concerned principally with journals, and at the intellectual level, with their content.
4. eLib supporting studies information: <URL: <http://www.ukoln.ac.uk/models/studies/>>
5. SICI version 2 is available from: <URL: <http://sunsite.Berkeley.EDU/SICI/>>
6. Report of BIC/UKOLN seminar on unique identifiers: <URL: <http://www.ariadne.ac.uk/issue8/unique-identifiers/>>
7. CASA project information: <URL: <http://www.cib.unibo.it/casa/>>
8. Hopkinson states that 'host statement' is a term used by ISBD to describe the data relating to the location of the article in the source serial.
9. ISO 10324: Information and documentation: Holdings statements: Summary level
10. <URL: <http://www.ukoln.ac.uk/services/elib/papers/other/anderson/>>
11. ISO ILL Protocol (contribution to eLib study on document requesting standards) is available from: <URL: <http://www.ukoln.ac.uk/models/serials-mtg.html>>
12. The UKOLN Z39.50 pages provide links to a range of information about Z39.50: <URL: <http://www.ukoln.ac.uk/z3950/>>
13. EDItEUR: <URL: <http://www.editeur.org/>>
14. The subject of integration was discussed at 'Standards for electronic document ordering', 20 February 1997, organised by BIC.
15. Summary information about the M25 clump is available at: <URL: <http://www.ukoln.ac.uk/services/elib/background/summary2.html>>