

<b>Author:</b>	Sandy Shaw,
<b>Date:</b>	22 September 2004
<b>Programme:</b>	Common Services
<b>Committee Funding:</b>	JCIIE
<b>Structure:</b>	4 pages
<b>Document Reference:</b>	CMSS: Shaw1

## Core Middleware and Shared Services Studies

Institutional Profiling and Terms and Conditions Services Scoping Study

---

# Contents

- Executive Summary .....3
- 1 Introduction .....5
- 2 Background.....6
- 3 Types of data .....7
  - 3.1 Institutional contact data ..... 7
  - 3.2 Institutional telematic service data ..... 7
  - 3.3 Learning and teaching information ..... 9
  - 3.4 Subscriptions data ..... 10
- 4 Sources of data.....11
  - 4.1 Institutional deposition ..... 11
  - 4.2 National authorities..... 11
  - 4.3 Subscription agents and publishers ..... 14
  - 4.4 Catalogue record vendors ..... 14
- 5 Rights management issues ..... 16
- 6 Service development criteria ..... 18
- 7 Service proposals .....19
  - 7.1 Contact data 19
  - 7.2 Telematic service data..... 19
  - 7.3 Learning & teaching data ..... 20
  - 7.4 Subscriptions data ..... 20
- 8 Terms & Conditions .....21
- 9 Conclusions .....23

---

## Executive Summary

This scoping study by EDINA has considered the functional requirements for two components of the JISC Information Environment that have been identified in the Shared Services Development Plan, but have not yet received detailed attention. These components are an *Institutional Profiling* service which resolvers will inspect to discover institutional preferences for OpenURL resolution; and a *Terms & Conditions* service which provides machine-readable information about rights held in resources in the Information Environment (IE).

Four categories of data that might be included in an Institutional Profile were considered:

- institutional contact data, including postal, telephone, web page, and institutional roles;
- institutional telematic service data, including directory access, Z39.50 services, IP range, OpenURL resolver, Shibboleth handle service;
- learning and teaching information, including course data, learning object repository descriptors;
- subscriptions data for electronic resources.

Several sources from which these data may be obtained were identified:

- institutional deposition, for information such as contact data;
- national authorities, such as the JISC Monitoring Unit, JISC Collections, UKERNA, and the IESR;
- subscription agents/publishers, who hold licence information for journal subscriptions;
- commercial vendors of catalogue records for journal aggregations.

While much of the data described can be readily obtained from reliable sources, there are major difficulties in acquiring rights data for electronic resources. As this is a central purpose of the Institutional Profiling and Terms & Conditions services, and concerns the data needed to link users to those services where they enjoy access rights, this deserves closer inspection. There are two broad approaches: *licensee assertion* and *licensor disclosure*. In licensee assertion, the licensee attempts to record details of its subscriptions in a local administrative database. This is a burdensome process, prone to inaccuracy, and represents an ongoing maintenance commitment. In licensor disclosure, the publisher or subscriptions agent discloses the institution's licensing information in machine-readable form. Two approaches are under investigation: firstly, OCLC are developing a Cooperative Rights Database where licensors deposit rights data (licensing authority tables) with a trusted third party; secondly, NISO and EDItEUR are developing data definition standards to describe licence terms (licence agreement records). This would provide the institution with administrative metadata that could be processed by a library automation system.

The main conclusion of the study is that provision of an Institutional Profile service would improve communication with the members and services of an institution. Given the local nature of this data, and the problems of imposing central control, the service should be distributed rather than centralised and treated as a local responsibility. Development could be productively undertaken in the following areas:

- a) ***Institutional contact data***. A schema should be developed that would enable institutions to generate XML data objects describing local contact or telematic services and make this available for machine-to-machine access by external services.

- b) **IP address range.** Open discussion with UKERNA on the feasibility of creating a service that provides data defining the IP address ranges of institutions.
- c) **Subscriptions data.** The two positive developments here, licensing authority tables (OCLC) and licence agreement records (ONIX) should be kept under review.
- d) **IESR.** Possible extension to the scope of the IESR should be considered, both to store additional information required for automated use, and to accommodate additional data types.

A further conclusion is that while a Terms & Conditions service would be of substantial value to institutional librarians, the difficulties in developing the service are equally substantial. Again, the need for co-operation from publishers to adopt standardised terms and conditions is the key factor in resolving the problem of managing large numbers of licences with arbitrary variability and complexity, expressed in legal terminology. The JISC has long recognised the need for standardisation in licence conditions (in the development of the PA/JISC Model Licence, for example) and this work should be continued.

---

# 1 Introduction

This scoping study considers the functional requirements for two components of the Information Environment that have been identified in the Shared Services Development Plan, but have not yet been fully addressed. These components are:

- an Institutional Profiling service which resolvers will inspect to discover institutional preferences for OpenURL resolution;
- a Terms & Conditions service which provides machine-readable information about rights held in resources in the JISC Information Environment (IE).

The main tasks are to re-examine the roles of these services within the shared services model, and formulate concrete proposals for their practical implementation.

The study considers the issues affecting the Institutional Profiling service from various perspectives:

- the types of data that may be included in the services;
- the sources of the data (where the authority for the data resides, and where the ongoing responsibility for its maintenance lies);
- further issues of rights management;
- service provision options.

Specific issues for the Terms & Conditions service are considered in clause 8.

## 2 Background

The architectural model developed for the JISC Information Environment (IE) envisages a set of shared services to provide an infrastructure to support the range of associated services that together provide the link between content and presentation. These shared services are typically at one remove from the user, and are accessed indirectly by means of machine-to-machine interfaces.

The Institutional Profiling and Terms & Conditions services were conceived as placeholders in the JISC IE architecture providing information on institutional rights, service preferences, and holdings, and providing assistance in directing users to sources of service. Since then, the OpenURL model has been developed more fully, and the original definitions for these services now appears misaligned with current practice. Nevertheless, the motivation behind the original concept remains sound and is worthy of reinvestigation.

While the two services which are the subject of this study are logically distinct, in practical terms they are related, and it may be sensible to reconsider the requirements in broader terms. There appear to be two distinct strands: institutional services, and licence data (with Terms & Conditions a special case of the latter). A further consideration is the emergence of the IE Service Registry (IESR), whose function may overlap that of the Institutional Profile service.

## 3 Types of data

Four categories of data may be considered for inclusion in an Institutional Profiling service:

- institutional contact data;
- institutional telematic service data;
- learning and teaching information;
- subscriptions data.

### 3.1 Institutional contact data

The following contact information could be recorded in an Institutional Profile:

- a) **General information.** Formal institution name, postal address, telephone and facsimile numbers, e-mail addresses.
- b) **Web page locations.** Staff list, Library portal, Admissions, Library catalogue.
- c) **Roles.** Names and contact details for the Vice-Chancellor or Principal, Pro-Vice Chancellors or Vice/Assistant Principals, the Secretary, the Librarian, Heads of Departments, and other administrative contacts.

This type of information is already maintained, and published either on the web or in institutional publications. Contact information is used widely on a routine basis, both within and between institutions, but its storage method and management is ad hoc. It is therefore available for human access (if one is prepared to search it out), but is not codified for machine-to-machine access that would simplify its discovery and use.

### 3.2 Institutional telematic service data

Institutions operate a variety of telematic services that may be used for different forms of inter-institutional communication. Access information for these services could be recorded in the Institutional Profile:

- a) **LDAP.** The Lightweight Directory Access Protocol is the predominant access method for directory services, and in theory, the LDAP directory is the most appropriate place for an institution to publish data about its users and services. However, despite efforts in the early 1990s to promote the uptake of LDAP technology, few institutions now use these directories other than for niche services with restricted access. The main obstacles are: the complexity of the technology; the risks involved in storing sensitive information; and concerns over the publication of personal information (such as e-mail addresses) that may be subject to misuse.

There is, however, an enduring requirement for services that provide controlled access to data held on institutional databases, such as the Shibboleth Handle Service or a user's digital certificate or course details, and LDAP directories may yet emerge as the most appropriate general purpose solution for the storage and access of these types of data. Meantime, there is limited demand for access information for institutional LDAP directories.

- 
- b) **Z39.50**. Many institutions provide open-access to their library OPAC by means of the Z39.50 protocol. Whereas Z39.50 may be regarded as a legacy protocol that has been superseded by a variety of more elegant solutions, it remains the predominant technology deployed in commercial library automation systems. Whilst actual connection details for machine-to-machine contact would be properly placed within the IE Service Registry, the detailed licensing arrangements covering terms of use could be referenced from the IESR in the Terms and Conditions service.
- c) **IP range**. The IP range covered by a particular institution's networked computers is of value to both end-resources and other middleware services as a means for determining the institution to which a user belongs. Typically this information is duplicated at each end-resource that requires such information. A central repository of this information could be made accessible as a querying service for end-resources. This could be used to determine:
- whether a user with a specific IP address is allowed access to a given resource;
  - the institution to which the user belongs.

With a central repository of authoritative IP address information, the quality and currency of the data in use would be significantly better than under the present ad hoc arrangements, of multiple repositories with informally recorded IP address ranges. (Note that when a librarian registers the local IP address range with a publisher when concluding a licensing agreement, there is little prospect that any future change to the physical address range will be communicated to the publisher, unless a loss of service results.)

- d) **Credentials for remote services**. Within the JISC IE architecture, it is envisaged that a user will be able to make a request to an intermediate shared service, which acts as a broker in forwarding the request to multiple end-resources. Where common authentication and authorisation credentials (such as those provided by Athens) are used by the end-resources this is relatively straightforward to implement. Unfortunately, many end-resources implement proprietary authentication mechanisms, particularly for machine-to-machine contact, usually in the form of an institutional username and password.

This means that a third-party brokering service, such as a portal, cannot provide access to any such end-resource on behalf of the user, unless these proprietary credentials have been disclosed by the institution to the brokering service. This disclosure has to be made by each institution to each brokering service for each end-resource.

An Institutional Profiling service could act as a repository of proprietary credentials, and be able to securely transmit such details to authorised third-party services upon receipt of standard authentications credentials provided by the end-user, thus allowing an apparently seamless brokering for the user to resources requiring non-standard authentication credentials.

- e) **OAI repository**. The proposal that institutions should create local repositories to enable their staff to self-archive their work has attracted much interest and is likely to result in the deployment of a number of OAI repositories, particularly in the larger institutions. A number of informal mechanisms for announcing the location of institutional repositories are in place for the benefit of harvesting services, but there would clearly be some advantage in a more controlled regime for publishing this information. Again, this may have a more natural home in the IESR than in an Institutional Profile.
- f) **OpenURL resolver**. The OpenURL Router has demonstrated the requirement for a service that enables service providers (“referrers” in OpenURL terminology) to locate the appropriate resolver for

an institution. This information could be registered in various places: in the OpenURL Router itself; in the IESR; or in an Institutional Profile. It is worth noting that the scope of the OpenURL has been broadened by the NISO standard, and we can expect the OpenURL to be deployed for a wide range of services for many different types of information object. It is possible that institutions will make use of more than one resolver, and the task of finding the “appropriate resolver” will become that much more complex.

- g) ***Shibboleth handle service***. Shibboleth has been adopted by the JISC as the preferred framework for authentication and authorisation in the Information Environment and for international secure communication. Each institution which implements Shibboleth must operate a handle service, which verifies a user’s locally provided credentials (typically, a name and password). The scheme requires that the identity of the handle service is known to each federation with which the institution is affiliated (specifically, the federation’s WAYF service).

Clearly, each WAYF service could be separately configured with the location of each federation member’s handle service. Alternatively, the handle service could be registered just once, in the Institutional Profile, and read periodically by WAYF services.

### 3.3 Learning and teaching information

Various forms of Information about courses and teaching could be recorded:

- Number of courses
- Nature and structure of courses
- Reading lists for courses
- Teaching staff
- Number of FTE teaching staff
- Number of FTE students
- Proportion of full-time and part-time students

A requirement has been identified to develop learning objects repositories for the storage of, and controlled access to these materials by students and course developers. The preferred architecture for the management, storage, and access of learning objects is still work in progress, but will probably involve a mixture of institutional, national, and regional repositories with potentially complex access control requirements. Mechanisms will be needed to build catalogues describing the contents of these repositories; this might be based on deposition, or on some harvesting mechanism.

The conditions of re-use of learning objects deposited in a repository are likely be complex. In the JORUM+ project, licence information (which incorporates terminology from the Open Digital Rights Language (ODRL) specification) is embedded within the learning materials and is retrieved along with the materials themselves. There are two licences, which differ according to whether materials may be adapted and modified, and two associated sub-licences for users. Depositors are required to choose a licence and endorse its terms and conditions, and users, in turn, need to agree the terms and conditions of the sub-licence before they can download the materials.

---

It appears that the distinctive requirements for managing learning objects and other information associated with L&T is likely to require a bespoke solution rather than one based on either the Institutional Profile or the IESR.

### 3.4 Subscriptions data

University libraries have long-established procedures for the management of their print journal subscriptions, and library management systems incorporate well-developed tools to assist the librarian in performing each of the bookkeeping and handling tasks necessary. (For each journal, these tasks include evaluation, selection, acquisition, licensing, cataloguing, and other management matters.) Where many subscriptions to print titles exist, academic libraries will frequently use subscription agents, and rely on them to provide the information on holdings, predicted prices and renewal dates. These arrangements, though complex in themselves, have evolved over the years and enable the management of print journals to be handled in an operationally sound manner.

By contrast, the growth in the use of electronic resources has not been matched with the development of tools that enable the librarian to manage these resources satisfactorily. The expenditure on electronic resources has grown substantially in recent years, yet the tools available to the librarian to manage these remain primitive and many serials departments still focus on the more tractable problem of managing the print collection. Some integrated library automation systems are now providing tools to enable librarians to manage electronic resources, but there are significant differences between the administrative metadata and the licensing issues for print and electronic resources which require more than simple extensions to existing administrative workflows. With an electronic resource, it is difficult to know when an issue is missing, or overdue in publication. Worse, as the content of an electronic aggregation will change over time, the librarian cannot even be sure which journals and which issues should be available day-to-day under each licence.

Libraries would benefit from an authoritative source of data on their own subscriptions across the board – both to printed and to digital content. This would assist them with maximising their materials budgets, with decision-making on, for example, the relative value of e-only bundled deals, and in the provision of article access/supply services which ought to be able to draw transparently from both subscribed content and just-in-time document delivery services.

A further issue is that while a print resource, once purchased, is owned by the library (albeit with some restrictions on use in observance of copyright), for electronic resources publishers are free to specify complex licence conditions of arbitrary variability. Often, given the difficulty of determining the Terms & Conditions that apply to a particular resource, the librarian will simply assume that the most restrictive conditions apply.

This problem has been recognised by the Digital Library Federation, and is the subject of the DLF Electronic Resource Management Initiative (see <http://www.diglib.org/standards/dlf-erm02.htm>). This is addressing the need for common specifications for managing licence agreements, administrative information, and library workflow processes for electronic resources. Until the management of subscriptions information is handled within a coherent framework, the provision of a general solution to the appropriate copy problem will remain elusive.

## 4 Sources of data

The types of data identified in clause 3 are candidates for inclusion in an Institutional Profile may be obtained from a variety of sources:

- institutional deposition;
- national authorities;
- subscription agents/publishers;
- commercial vendors.

### 4.1 Institutional deposition

Ownership of the data listed in 3.1 and 3.2 rests with various members of the institution, including the MIS, the Library, and Computing Services (though in some cases, there may be no clear owner). If this data is to be maintained (assuring timeliness and accuracy) the responsibilities for its maintenance must be clearly understood. This may be a significant obstacle for the introduction of a profiling service since it requires the adoption of an institutional policy to assign these responsibilities. Ideally, a single responsible person would be appointed, though the problem remains that any changes to the data would have to be reported to this person. Alternatively, the information could be deposited by each of its owners (which would help guarantee its accuracy), but this may compound the problem of its timeliness.

This data would be a very useful resource (for example, to enable a researcher to discover the contact details of University Librarians or Computing Service Directors). The main question is whether institutions would regard the costs of maintaining this data as justified by the benefits, though there may be scope for rationalising this type of data collection with existing reporting requirements.

### 4.2 National authorities

A number of national authorities hold information concerning all institutions, and some commercial organisations generate data relevant to many institutions. There is a need to establish precedence across these sources for the purpose of determining authority.

#### 4.2.1 JISC Monitoring Unit (MU)

The JISC Monitoring Unit (<http://www.mu.jisc.ac.uk/>) holds information about institutions and JISC-funded services and service providers. This information includes performance and usage data for JISC-funded services, service level definitions for the JISC-funded service providers, the results of commissioned surveys, and a list of HE and FE institutions and Research Council sites.

UKERNA also maintains a list of HE and FE institutions eligible for a primary connection to JANET which is used for charging and network maintenance activity and on which the MU list is based. The two lists will correspond in terms of the total number of institutions but there are differences in some of the data held. For example, UKERNA usually uses the 'colloquial' name of the institution whilst MU check and include the official institution name (as considered by the appropriate funding body) in the list they

maintain. MU indicated that UKERNA are starting to view the MU list as a definitive source of institution information which they should consider using. The MU list is available on the web.

Athens also maintains a list of institutions and local Athens administrators. These institution names do not always quite match those on the MU list, although MU believe that Athens uses the MU on-line list. A definitive list is important as funding councils and government often specify eligibility of funding for additional services to the JANET connection on the basis of an institution having a primary connection.

Another potential source of institutional data was thought to be UCISA; however, it is believed that they no longer publish a book of institution titles.

An Institutional Profiling service would require a means of authoritatively identifying institutions in the community. The MU data set appears to have this status within the community and so would provide this.

MU receives quarterly-based information from JISC-funded service providers. EDINA provides this information to MU and so has knowledge of the type of data involved. This includes aggregated login information by institution for each JISC-funded service. Performance-related information is also provided for each service indicating the response times for specified searches, and the details of any downtime. All this data is made available by MU on their web site.

Under the new Athens contract, Athens will supply MU with enhanced data. Athens will be expected to supply information detailing the JISC-funded services to which each institution has subscribed. MU would expect to receive information about new subscriptions during the first month of an institution subscribing to a service. However, it has not been agreed whether this information will be made available on the MU web site due to sensitivity over potential non-use of a service by an institution.

MU are contracted by JISC to carry out a variety of surveys. The annual network infrastructure survey of HE and FE institutions aims to survey every institution once every three years. Other surveys are commissioned from time to time such that MU expect to manage two surveys a year. Examples include JSTOR use (this was carried out by MAU, before the MAU and TAU were unified into the MU) and a TASI-commissioned learning and teaching based survey.

Limited institutional contact details are held by MU and they see themselves as potential users of a service which would provide this contact information. One recent survey was sent out to 'Head of Computing/IT services', <institution name>. MU would prefer to address surveys to a named individual, but would retain the option of addressing contacts by title. Within HE it is usually possible to use a title that ensures a reasonable chance of correct delivery. In FE, however, this is often not the case as individuals will often assume a role but with no assignation of a title. Roles that MU have required for recent surveys include Head of IT, Director of Information Systems, Subject Librarian, Head of Teaching and Learning.

#### **4.2.2 JISC collections**

The JISC Collections Team negotiates with publishers and other rights holders to enable institutions to acquire access to high-quality electronic content at favourable subscription rates, and to obtain other new forms of electronic resource. The details of these institutional subscriptions provide a potential source of information that would be of value to OpenURL resolver services, which would use the

information to direct users to services for which they had subscription rights. Some issues arise concerning the use of this data:

- It is unclear whether any confidentiality issues arise either for institutions or publishers that would constrain use of this data.
- The administration of the subscriptions is managed by an external agent, acting on behalf of JISC, and the form in which the data is held is, therefore, not strictly under JISC's control. In order to make effective use of the data, its format would have to be agreed and any changes to the format made subject to negotiation.
- JISC collections cover only a subset of the full range of electronic resources used by institutions. Many other consortia, subscription agents and publishers are involved in dealing directly with libraries. Therefore, at best, JISC Collections data gives only a partial solution to the problem of acquiring institutional subscriptions data.

### 4.2.3 UKERNA IP data

Many organisations, including Data Centres and other service providers use IP address ranges to identify the institutional affiliation of a user requesting service. This method has clear shortcomings: for the service provider it gives a non-authoritative indication of affiliation, since there is generally not a one-to-one relationship between address ranges and organisational units; for the user, the mechanism is only effective on campus, unless an institutional web proxy is available. In fact, a more reliable and simpler method is already available, namely reverse DNS lookup. A requesting user's IP address is mapped to a DNS name which can be checked against a list of recognised institutions. Commercial providers, however, appear content with the existing IP-based method, despite its shortcomings.

IP information is often provided by a librarian when making a subscription to a licensed resource, and this may introduce additional errors. Further, in general, it is poor practice to use one source of information when a more authoritative source is available. In this case, the most reliable source of IP addressing information for HE institutions is held by UKERNA. This is currently made available by means of the whois mechanism. Unfortunately, this does not present the data in a form suitable for use by organisations which currently rely on IP address checking. In principal, however, all the required data is available, and the task of generating IP address descriptors for each institution is tractable. This task would probably have to be undertaken by UKERNA, and integrated with their existing administrative procedures. This could be made available as a stand-alone service provided by UKERNA, or could be included in the central Institutional Profiling service. The only advantage of the latter arrangement is that the data would be available in a more public location.

### 4.2.4 Information Environment Service Registry (IESR)

The IESR is an important development that enables other services to discover the existence of institutional resources and services, and for many of the data items identified in this study the IESR appears to be the most appropriate home.

In some cases, however, the IESR as currently defined does not include sufficient details about operational protocols and compatibility to facilitate full machine-to-machine interoperation. For example, the IESR whilst giving an indication of authentication/authorisation methods, does not hold the specific details required for automated use:

- an IESR record indicates whether a particular service uses Athens, but does not provide information such as the Athens resource name which would be needed to construct an Athens authentication enquiry;
- an IESR record indicates whether a particular service uses a proprietary username/login authentication method, but does not store the actual credentials in the IESR.

The terms and conditions of use of a service are also absent from the IESR.

The IESR could be augmented to hold this additional information. Alternatively, if records within the IESR were cross-linked with augmented data held in an Institutional Profiling service (and Terms & Conditions service), the complementing services would enable an automated process to configure an end-resource for access by other middleware services within the JISC IE.

### 4.3 Subscription agents and publishers

It is the business of a publisher or subscription agent to know which institution is subscribed to which titles, bundles or aggregations. It is much less simple for the library itself to keep track of this information, and to populate a local OpenURL resolver service.

The use of Electronic Data Interchange (EDI) is well established within the publishing industry for the exchange of subscriptions information in electronic form between publishers and subscription agents. However, information about subscription rates for specific titles is extremely commercially sensitive, being an important element in any comparison of cost vs. price. Furthermore, collation of this type of information could reveal the proportion of total UK business commanded by a particular agent; another commercially sensitive area. Thus, while widespread co-operation by subscription agents and publishers is the key to a range of interoperability issues within the JISC IE, these organisations are understandably wary of disclosing subscriptions information to any third party.

Another important factor to consider is that the largest subscription agents are no longer simply agents. In an electronic environment, profits for all intermediaries have been squeezed. Those with the resource and capacity to develop beyond their traditional roles have expanded to occupy areas of the value chain that were previously occupied by other organisations. 'Subscription agents' such as Swets and EBSCO now offer linking solutions as part of their electronic journal delivery services. Disclosure of subscriptions information to the OpenURL resolvers of other providers potentially undermines sales of their own resolver modules.

### 4.4 Catalogue record vendors

In an electronic environment, the information required to determine whether a specific institution is subscribed to a specific title has two components: data confirming that the institution is subscribed to a relevant service; and data confirming that the title sought is included in that service. Clearly, the publisher or subscription agent will know the services to which their customers are subscribed, as do the customers themselves. However, the customer does not always know the titles included within a subscribed service. This arises from the nature of the scholarly journal publishing industry.

This industry comprises a few large companies owning scores and sometimes hundreds of titles, and a very large number of organisations that own only a few titles. However, the cost of developing and providing electronic delivery services for scholarly journals is prohibitively high for many of the smaller publishers and learned societies. Intermediaries such as HighWire provide a service that enables these publishers to sell content in electronic form without investing substantial resource in the development of sophisticated delivery systems. Thus, electronic journals from large and small publishers are often available in bundles and aggregations. The content of these bundles and aggregations is fluid. After a subscription is sold, the seller does not necessarily advise the library of changes to the content of a bundle or aggregation, yet this information is required if library catalogues are to accurately reflect the content available to library users. Thus while a serials librarian may know that she has subscriptions to Services A, B and C, she may not know which specific titles are available through those services. The upshot of this uncertainty is the creation of a niche in the value chain. While subscription agents can tell the library which services it is subscribed to, a new service is required to tell the library which titles are included in those subscriptions.

Commercial vendors such as Serials Solutions have emerged to fill the gap between information provided by publishers and agents and information required by libraries to fulfil the information needs of their users. Serials Solutions provides a library with details of the current content of the bundles and aggregations to which that library is subscribed. It also offers catalogue records for each title. A library may provide Serials Solutions with details of the services to which it is subscribed and Serials Solutions provides the records required to keep the local catalogue current. A subscription to Serials Solutions is an additional cost on the licensee institution. The cost of a commercial OpenURL resolver and a subscription to Serials Solutions cannot be afforded by all UK academic libraries.

It would be very desirable to maintain this information in a national institutional profiling service (if a means of acquiring it could be found). This would obviate the need for the commercial solutions described, and ensure that the complex subscription information which defines many institutional digital library and learning services is available to the national middleware infrastructure which can deliver maximum efficiency across UK HE and FE.

## 5 Rights management issues

The Institutional Profiling and Terms & Conditions services were placeholders in the JISC IE model for the then unresolved problem of locating the appropriate copy. This refers to the problem of directing a user who has identified an article of interest to the full text of the article. As there are usually several potential service providers offering the full text, the problem is to identify which service provider the user is authorised to access (if any) by virtue of institutional subscription. Since the JISC IE model was devised, the OpenURL link format has been developed as a means of addressing this problem. The OpenURL link, which contains a reference to an article, is mapped to a service containing the full text of the article by an OpenURL resolver service with knowledge of the user's institutional subscription rights.

With the OpenURL in place, the challenge for the Institutional Profiling and Terms & Conditions services has shifted to one of rights capture, derived from subscriptions data. An institution's subscription information may be made available to its OpenURL resolver in two possible ways: licensor disclosure or licensee assertion.

Licensor disclosure requires each licensor of electronic journal services (which may be a publisher or a subscription agent) to make its licensing information available in machine-readable form so that resolvers can identify the titles to which a specific institution has rights. The licensing information may be disclosed in a number of ways:

- a) **Licensing authority tables.** Here, the licensor deposits details of each licence agreement with a trusted third party which makes the information available to authorised resolvers. EDINA floated this idea in 2000 with the Association of Subscription Agents, where it attracted significant, but insufficient support (the two largest subscription agents declined).

A similar idea is now being tested by OCLC as it develops its Cooperative Rights Database (CRD). OCLC also perceives licensor co-operation to be the primary challenge for the development of licensing authority tables. When last consulted, OCLC was still working to assemble the critical mass of publishers and subscription agents necessary to make the scheme effective. But as an international player in the serials information chain, OCLC has more leverage than EDINA or the JISC is able to bring to bear. Pressure may also be applied through NESLI in community-wide content negotiation, and through awareness-raising by JISC within institutional libraries and procurement offices of the importance of requiring compliance from publishers and agents. Information on the CRD is limited, but see <http://www.library.cornell.edu/abld/abld03/carney.ppt>. If OCLC is successful in launching the CRD in the US, there is some prospect that a similar scheme could be introduced in the UK.

To date, no concrete information has been available on a possible service, under consideration within the British Library, that would direct a user to the Inter-Library Loan service only where the user has no rights to an object by virtue of institutional subscription. This could also be based on the CRD concept, and the development (if approved) will be watched with interest.

- b) **Rights enquiry.** This approach was investigated by EDINA when developing BALSAs, a general-purpose OpenURL resolver for the JISC IE. Rather than gathering and storing licensing information to respond to OpenURL queries (a just-in-case strategy), BALSAs requests licensing information as it is required, (a just-in-time strategy). When the end-user seeks access to an article of interest, BALSAs asks all possible service providers whether this user (from a given institution) would be

authorised to access the article. If so, BALSAs presents the user with a link which takes him directly to the article on the service provider's site. As with the licensing authority tables, this strategy requires co-operation from licensors. EDINA has secured co-operation from some key providers but not a sufficient number to provide a comprehensive service.

- c) **Licence agreement records.** In this scheme, the licensor provides the licensee with a package of metadata that describes the licence terms in machine-readable form. This provides the institution with a package of administrative metadata that can be processed by the institution's OpenURL resolver, and by its administrative management system for electronic resources. This scheme is dependent on the development of a suitable metadata schema, supported by publishers, that accurately models subscription package data.

Development of a set of industry standards for exchanging serials subscriptions information between all parties in the information chain is underway. This is known as ONIX and its development is being directed by a joint working party of NISO and EDItEUR, an international group which coordinates standards and infrastructure for e-commerce in books and serials (see <http://www.editeur.org/onixserials.html>). This is an encouraging development, but may have some way to go before reaching maturity.

Unfortunately, while all of these schemes provide an effective technical solution, they are wholly dependent on the co-operation of the many licensor organisations involved, which past experience shows is difficult to obtain. A scheme may fail if even only one or two of the larger subscription agents declines to co-operate.

An alternative to licensor disclosure as a means of providing licensing information to an institution's OpenURL resolver is licensee assertion. In this scheme (used by most commercial resolver products such as SFX) each institution attempts to record complete details of each of its licence agreements in a local administrative database. This is a substantially weaker approach than licensor disclosure for the following reasons:

- the master databases that actually control access to licensed data are held by the licensor;
- licence documents often provide inadequate descriptions of the rights and holdings they cover;
- local keying of rights to titles is burdensome for the local librarian, and prone to inaccuracy.

These problems can be mitigated if the institution subscribes to a service such as Serials Solutions which provides current information about the content of aggregations and bundles along with catalogue records for each title. However, not all institutions can afford this type of service.

## 6 Service development criteria

Before embarking on the development of an Institutional Profiling service, a number of conditions should first be satisfied for each candidate category of data that might be included:

- a) **Utility.** Reasonable confidence that there is demand for the data and that its use will justify the costs of the service development.
- b) **Practicality.** Confidence that there is clear assignment of responsibility for maintaining each set of data, and expectation that the data will remain accurate and timely. Existence of effective methods for the creation, updating, and transmission of the data. Note that for different types of data, different personnel will be responsible. Potential sources of authoritative data should be ranked for authority, so that a national service will use the most authoritative source for each data stream.
- c) **Cost-effectiveness.** If the first two criteria are satisfied, it must also be shown that the cost/benefit tradeoff of operating the service gives the community good value for money. Each category of data should be considered individually as there are likely to be differences in both utility and administrative overheads.

For each category of data, once these factors have been evaluated, further choices exist for service delivery:

- d) **No action.** Where the utility is limited, or where substantial operational difficulties exist, the data category would not be considered further. In cases where the balance of interest is less clear cut, a judgement on cost-effectiveness would have to be made.
- e) **Central Institutional Profile.** Where the data user is likely to require the same data item from several institutions (such as when performing cross-searching on institutional OPACs), it would be reasonable to store the data in a centrally-maintained service.
- f) **IE Service registry.** Some of the data identified may belong more sensibly in the IESR.
- g) **Devolved operation.** In other cases, it may make more sense to devolve the operation to local or specialised services.
  - For data generated and maintained within the institution it may be appropriate to store this locally.
  - For other data, only of use to single specialised services, it may be appropriate to store the data with the service. Examples include details of institutional OpenURL resolver services lodged with the national OpenURL Router service, and details of institutional Handle Services lodged with the national Shibboleth WAYF service.

Where data is transmitted from the institution to either a central service, or a devolved specialised service, suitable mechanisms are required to ensure that all communication originates from authorised personnel.

## 7 Service proposals

Having considered the types of information available, their sources, and issues of maintenance and sustainability, specific service options for an Institutional Profiling service are considered for each of the four data types identified.

### 7.1 Contact data

This type of data easily satisfies the test of utility. Contact information is used daily, both within and between institutions by a wide range of individuals and for a wide range of purposes. What is less clear is whether institutions can be successfully encouraged to maintain this information in a timely manner with a remote Institutional Profiling service. (As with most of the data considered here, inaccurate information is worse than no information at all.)

One alternative to the Institutional Profiling service as a home for this data is a devolved service based at institutional level. At its simplest, this could be based on a web form that the institution fills, and amends as necessary, from which is generated an XML object, made available in a predictable location (such as `profile.uni.ac.uk/schema-name`), and available for machine-to-machine processing. The definition for the structure and contents of the data object would be defined nationally, and expressed using a DTD or other suitable specification mechanism. A service could fetch this data on demand for human display, or could harvest it and make it available as a database resource.

This has some advantages over a centralised service: ownership of the data, including the authority to amend it, and the update frequency, is clearly the institution's responsibility. A number of individuals in the institution may be ultimately responsible for different data elements but it is the responsibility of the institution itself to coordinate maintenance activity.

### 7.2 Telematic service data

The various types of data in this set may be handled in the following groups:

- a) **Devolved services.** Some data is of value only to a single service or a small number of services, and may best be handled by registration of the data directly with the services concerned. This includes:
  - LDAP service (probably no takers);
  - Credentials for remote services (Xgrain);
  - OpenURL resolver (OpenURL Router);
  - Shibboleth handle service (WAYF service).
  - For completeness, this information could also be registered with the IESR.
- b) **Z39.50.** Details of Z39.50 access to library OPACs are of interest to portal services, such as GetRef (Xgrain), and should be registered with the IESR.
- c) **OAI repository.** This data is of potential interest to a number of services, and should be registered with the IESR.

- d) **IP ranges.** This data is used by a wide range of services, including JISC services and commercial publishers. The authoritative source of information on IP address ranges is provided by UKERNA, which would be the obvious home for developing a service to make this information generally available. Alternatively, if an Institutional Profiling service was to be established, this should be populated with data supplied by UKERNA rather than by the institutions themselves.

The service definition would make clear the limitations that have already been noted in the use of IP address to determine a user's location and institutional affiliation.

### 7.3 Learning & teaching data

There is an growing need for increased levels of co-operation and information exchange between institutions with regard to L&T information. This includes identification and sharing of learning materials, joint identification of courses, and establishment of virtual (teaching) organisations.

These requirements are not understood in sufficient detail for specific proposals to be made with any confidence, and it is probable that work currently underway in this area will need to mature before a clearer picture emerges of the type of infrastructure support required. This area could usefully be revisited by subsequent L&T studies to identify appropriate standardisation activity and infrastructure design requirements.

### 7.4 Subscriptions data

As has been indicated, this is a complex topic, but an area that requires resolution to realise the wider ambitions for the JISC IE, and the integrated on-line services it is working to create. While subscriptions data lies at the heart of the original proposal for an Institutional Profiling service, the foregoing discussion indicates that the solution for rights management does not lie with the type of central service envisaged. Broadly speaking, the choices are as follows:

- a) **Institutional OpenURL resolvers.** This is the solution provided by commercial resolvers and is known to be effective. Its disadvantages are the cost to the institution, both monetary and in ongoing use of staff time to populate the service database. (This latter task can be reduced, at the cost of subscribing to a service which provides information on the contents of bundles and aggregations, and provides catalogue records for those journals.)
- b) **Licensing authority tables.** The practicality of this solution is dependent on the success of the OCLC efforts to establish the Cooperative Rights Database. Discussion with OCLC indicated that they would be likely to reach a conclusion later in 2004 on whether they would be able to attract a sufficient degree of industry support to proceed.
- c) **Licence agreement records.** The practicality of this solution is dependent on the completion and adoption of the ONIX proposals for standardised serials subscription records. There is no clear timetable stated for completion and proposed roll-out of this work.
- d) **Existing sources.** Subscription data is available from JISC Collections and from the Monitoring Unit, and this could be made available to a national OpenURL resolver service. The main question is coverage — the proportion of UK subscriptions covered by these services is unknown, but it is likely that a significant number operate independently, and the need for a solution to cover these remains.

## 8 Terms & Conditions

Terms and conditions are expressed in full licences and vary across services and providers. While each service provider offers standard licence terms and negotiates on the price and the scope of provision, some service providers negotiate on other terms as a means of arriving at an agreeable price. The consequence is that the details of a licence for a given service may be specific to a particular institution. Without standardisation in licence terms, the librarian can either attempt to master the fine print of large numbers of licences, or make the pragmatic assumption that errs on the side of caution, and apply more restrictive conditions than are actually specified.

While commercial solutions are emerging that address the appropriate copy problem and maintain current information about the content of bundles and aggregations, licensee institutions still struggle to know the terms on which content may be used. It can be difficult to maintain an effective system for managing licence documents centrally (different licences tend to be held in different places within the library), let alone keep abreast of their terms and inform end-users about what they may and may not do with licensed content.

The Creative Commons initiative (<http://creativecommons.org/>) demonstrates how important standard terms are if content is to be discovered and used in a networked environment and if the terms of use are to be easily communicated to users. Creative Commons provides standard licences in three forms: a full legal licence written in legalese; a machine-readable licence for searching on licence terms and retrieving and distributing content within the licence terms; and a set of simple symbols that communicate to the end-user the key terms of the licence, e.g. that it may be copied and distributed for educational use but not in a commercial context.

Licences for electronic journals and bundles and aggregations of journals contrast starkly with the Creative Commons scenario. These licences are provided in one form: the full legal document, provided as a print or as a PDF file. It is not available as a structured document consisting of machine-readable terms that would facilitate telematic communication nor is it sufficiently standardised to facilitate simple iconic representation of its terms to end-users. Furthermore, licence terms are not sufficiently standardized to be memorable. A librarian addressing a query about the permitted uses of a specific title must find and read the legal licence for that title, or for the service in which the title is included. Vendors of library management systems are working to address this issue by producing terms and conditions modules for recording, storing and displaying licence terms. Without standardisation of licence terms however, there's no scope for a standard schema in the database and for effective machine-to-machine communication of use terms. Each library must interpret its licence and populate the database with its interpretation of what that means. This free text may then be consulted by librarians and is retrieved and displayed to end-users.

A centralised terms and conditions service would be valuable for librarians but would be premised on a degree of standardisation in licences that does not yet exist. Converting full legal licences into machine-readable form and communicating the terms to end-users would require the co-operation of the Licensor's legal team. Central translation of licence terms for each provider into machine-readable form and into icons to be displayed to users could then be negotiated on behalf of the JISC, and the resulting telematic licences could be used by licensees within the community. This work would be expensive and its success would be wholly contingent on the co-operation of Licensors, who would also be required to meet the expense of negotiating translation of their licences. National deals on content brokered by JISC

on behalf of the community could in time become conditional upon the provision of standardized information from licensors.

## 9 Conclusions

The main conclusion of this study is that the provision of an Institutional Profile service would improve communication with the members and services of institutions and would be an asset to the JISC Information Environment. Given the local nature of this data, and the problems of imposing central control, the service should be distributed rather than centralised and treated as a local responsibility.

The simplest solution may be to devise pro forma that institutions would be invited to populate locally, and publish on the web. This would enable a mechanism for machine processing by other services, including, for example, a service that harvests and integrates data from every participating institution. The scheme would be voluntary, and the motivation for institutions would be a desire to promote communication both internally and between institutions.

While this study concludes that the case has not been proven for the development of a centralised Institutional Profiling service along the lines originally envisaged in the Shared Services Development Plan, development could be productive in a number of areas.

- a) **Institutional contact data.** Develop a data definition for recording institutional contact data. This would be expressed as a DTD, or some other specification format suitable for generating XML data objects. Additional schemas could also be developed, for example, to cover details of certain telematic services, and locally-defined data sets. The purpose of these schemas would be to enable the local publication, in standardised form suitable for machine-to-machine access, of any data the institution wishes to release.
- b) **IP address range.** Open discussion with UKERNA on the feasibility of creating a service that provides data defining the IP address ranges of institutions.
- c) **Subscriptions data.** The two positive developments here, licensing authority tables and licence agreement records should be kept under review. Crucially, both approaches involve participation from the publishing industry (in the latter case the work arose from industry sources), and thus stand a better chance of acceptance than externally developed schemes such as rights enquiry. The use of existing sources of licensing data would be more attractive if coverage was greater but this is never likely to provide more than a partial solution.
- d) **IESR.** Possible extension to the scope of the IESR should be considered, both to store additional information required for automated use, and to accommodate additional data types.

A further conclusion is that while a Terms & Conditions service would be of substantial value to institutional librarians, the difficulties in developing the service are equally substantial. Again, the need for co-operation from publishers to adopt standardised terms and conditions is the key factor in resolving the Gordian Knot of managing large numbers of licences with arbitrary variability and complexity, expressed in legal terminology. The JISC has long recognised the need for standardisation in licence conditions (in the development of the PA/JISC Model Licence, for example) and this work should be continued.