

**Institution Registry Stakeholders' Meeting  
Jurys Washington Hotel, Washington, DC**

**October 7 2005  
10.00am-3.00pm**

**ATTENDING**

Gabriel Alvaro	University of California Press
Oren Beit-Arie	Ex Libris
Amy Brand	CrossRef
Paul Canning	IEEE
Adam Chandler	Cornell University
Adam Chesler	American Chemical Society
Mike Clarke	American Academy of Pediatrics
Kevin Cohn	Mary Ann Liebert Publishers
Dale Flecker	Harvard University
Martin Halbert	Emory University
Pat Harris	NISO
Barrie Howard	DLF
Tim Ingoldsby	AIP
Tim Jewell	University of Washington
Chuck Koscher	CrossRef
Larry Lannom	CNRI
Tim Lloyd	Alexander Street Press
Sally McCallum	Library of Congress
Greg Malar	Rockefeller University Press
Audrey Melkin	Atypon
James Mouw	University of Chicago
Georgios Papadopoulos	Atypon
Norman Paskin	IDF
Don Pavoni	University of Chicago Press
Ed Pentz	CrossRef
Oliver Pesch	EBSCO
Sarah Porter	JISC
Vicky Reich	Stanford University
Caroline Rothaug	Wiley
David Seaman	DLF
Ralph Shoffner	Ringgold
Mike Teets	OCLC
Gordon Tibbitts	Blackwell Publishing
Michael Winkler	University of Pennsylvania

## SESSION 1

[See PowerPoint presentation for overview of the day]

The meeting began with a roundtable introduction, and various members were able to suggest ways in which the idea (in the abstract at least) of an institution registry fitted with items they were currently working on, including the JISC registry and authentication work (common middleware services for UK institutions); the DLF's Electronic Rights Management Initiative (ERMI); the NSF-funded Ockham web services work; and DOIs.

### Defining the problem

The basic topic under consideration was *whether* and *how* the routine and regular exchange of information between licensing entities (who may be institutions, parts of institutions, or consortia) and service providers could be improved. Much of the morning was spent pulling this apart and examining whether there was a real problem or problems to be solved, what they were, and whether they were different for different stakeholder communities. Questions were tossed out quickly:

- Is the current situation really as complex as we are saying?
- Can a central Registry solve enough to make it worthwhile? Vicky Reich (Stanford) guessed that 80 % of the time Stanford sent on subscriptions went on 20% of providers who were not likely part of this solution.
- Does centralizing IP addresses solve enough of the problem, or does the real problem reside elsewhere – reliable contact information, for example?
- What would it cost to implement and can we quantify the cost of doing nothing?
- Are we looking only at the US and UK markets -- 40% of customer base – and ignoring the needs and local realities of other parts of the world.

There was fairly early agreement that it was time-consuming for libraries to deal individually with publishers who needed authentication information from them. There was less clarity on how a Registry makes sense for the publisher (minimizing the high cost of the migration of subscription data from publisher to publisher when a product or publisher is sold to another was one early suggestion).

Part of the uncertainty was the difficulty in quantifying the cost of IP management currently, although the publishers' discussion soon turned to the value proposition being not the simple management of IP addresses as much as the task of managing subscriptions: the cost for renewals (sending out paper notices) and the lack of a unique name for an institution across the board, which often led to mismatches between a publisher and a subscription agency over the same customer.

Dale Flecker (Harvard) offered that part of the value for a central registry and the extra rigor it could bring was not so much in solving current problems but in looking ahead to where are we going, which seems likely to be a world of many greater machine-to-machine communications of business arrangements within and across institutions.

To summarize what we thought the problems we face are:

**PROBLEM:** communication between libraries and service providers are inefficient, time-consuming, and mostly one-to-one. Access and especially contact information changes frequently.

Libraries maintain records of resources and services providers, etc.

- They need a unique ID and contact information for service providers

- They need to enable machine-to-machine transactions

Publishers maintain systems on subscribers and licensees

- They need a unique ID and contact information for licensing units

**PROPOSED SOLUTION:** A central registry, which holds access and contact information (or which points to locally-held information records).

**CURRENT BENEFITS:** faster transfer of information; potential service improvements; saving of staff time and effort (although balanced by an undetermined cost of implementing the new services by all parties concerned). If all of one's trading partners are in a central place one can download their current information into a local system.

**FUTURE BENEFITS:** eases machine-to-machine exchange of information, likely to become more and more prevalent.

### **Access and ownership**

Having decided that there was a range of potential problems that the proposed solution could address, we focused in on the central question of what information should be kept in the registry and what should be pointed to from it but kept under local control.

Questions of came up quickly -- some institutions (probably the larger ones?) may well want to manage their information locally, and have the Registry record for that institution refer people to the local place; smaller institutions may well be grateful of a well-run central place to store that and elect to have the central registry contain their information directly.

So we quickly settled on a hybrid model, with some centralized data for all users and for some users some of the additional content held locally. It was noted that technically the whole registry can be a logical construction running through web services and based entirely on distributed data, but the most practical likelihood was of some central place

because you do finally need some single place to look, whether it is the source of the information you seek or only a pointer to it.

## **Core Content**

Attention turned to what should be in an Institution Registry as core content. Any of this information can be held in the central registry if the institution to which it belongs elected to put it there, or it can be pointed to from a minimal record (items 1, 2, and 3?) in the central registry, but with the information itself held at the local institution.

### 1) A unique ID for a licensing unit and for a vendor

There was common agreement that one ID (assumed in the discussion to be a DOI) for an institution – a college or university -- is inadequate (and also inadequate for a large publishing company).

One ID for a licensing unit (an entity to which content is licensed) is possible. There was a question about what institution and publisher ID schemes we now have.

### 2) The official and authoritative name of that licensing unit or vendor.

### 3) Contact address information

### 4) Other ID schemes (e.g. ISIL -- ISO/DIS 15511 Information and documentation -- International Standard Identifier for Libraries and Related Organizations (ISIL))

### 5) URLs by service type

### 6) Authentication information: an identifier for the authentication method with the name of the authentication authority for an institution at the appropriate level.

### 7) Electronic Resource Management system information

There is also the potential for centralizing usage information derived from COUNTER reports and holdings information from Onix holdings records?

## **SESSION 2 -- HIGH LEVEL REQUIREMENTS**

The aim of this session was to see if we could articulate whether there is value in a registry from the various viewpoints represented in the room.

1) From the publishers' (service provider) perspective: Is it enough to get IPs in a better way? No – you probably still have to verify them, so not a “killer app”. Institution identifiers that were used everywhere throughout the supply chain would be a killer app.

Up-to-date and reliable contact information in a central location does have real value for libraries and publishers alike; and there was a clear recognition that the information had to flow both ways – libraries would like to use this service to discover publisher information, contact names, addresses, and a unique ID for the vendor.

A contrary view emerged – a publisher may well prefer that everyone came to their individual publisher website to update information (largely the current situation) – this works well for the publishers and feels like customer service – or at least, customer contact unmediated by a central broker. From a library perspective, one central place is better, but some questions arose about how many publishers would use it.

2) From the library (licensing unit) perspective: the feeling was that a large institution (Harvard; Stanford) probably had a few dozen “licensing units” -- entities who licensed content individually, from a single department to the entire institution (and beyond it into multi-institutional consortia). Each of these would need a unique ID.

If the library was the gatekeeper to the information in the central registry – or the information that is linked to in a locally-controlled space from the central registry – did that mean that the library and its processes would now need to manage lots of subscriber groups that are department or research center based? Do we want to do this?

We were urged to look at how this central registry interfaces with our existing and developing e-resource management systems; how it could enable machine-to-machine communications through a standard form of XML output for systems to consume, and to think of the registry as an index to locally-controlled access information for those institutions able to perform this local function (and therefore have closer control over this material).

**We broke for lunch at 12 noon.**

We resumed after lunch with presentations from Atypon, Ringgold, and JISC.

**ATYPON: Audrey Melkin/Georgios Papadopoulos.**

Atypon has built a registry service that they hope libraries and publishers will adopt – a beta will be out in March 2006 and the service is expected to be released in May 2006. Atypon is not planning on there being any charges to publishers, libraries or users of the system. For Atypon it is a way to serve publishers and benefit customers of the publishers by cutting down customer service calls and costs of access management.

The registry is already populated with content – information on publishers and their customers -- to which they have access from publishers they work with as a software systems vendor. Atypon is migrating metadata into the service now, merged from publisher databases to which they have access, and will be contacting institutions and publishers where they find the need. The rough estimate was that they have about 100 service providers and thousands of libraries in the emerging service.

Libraries will be alerted that they have information in the registry and it will be private until that institution signs off on it, at which point the individual library decides whether to opt out entirely; what to hold centrally and what to hold locally; and to whom to give the DOI pointer that leads to the registry record.

They are drawing on Onix cataloging information and COUNTER statistics reports.

In response to questions after the presentation, Atypon replied that the registry as conceived does allow a library to download publisher information (there would be DOIs to lead to service provider information), and the Registry can either contain all the information for an institution, or it can function as a pointer from a small central record to content that is held locally at a publisher or library. The DOI that is the identifier one uses to get to a record in the registry has to be given by the owner of the information, as it can't be guessed -- it is a random number.

Other questions: Would it be an asset of Atypon? No -- Atypon does not own this information. In order for the information they now have to be part of the Registry it would have to be "signed off on" by a responsible party at the institution that the data pertains to -- but how will Atypon find that responsible party? They responded that the publisher data they now have gives them a person's name and contact information for the licensed collections they know about.

Later in the meeting Georgios Papadopoulos (Atypon) reiterated that the start-up costs are quite low for a lightweight, functional, central registry and are being borne by Atypon, using existing software, and that they saw it as worthwhile promotion of their name and services.

Independent governance role: what plan is in place to prevent changes if Atypon is purchased, for example? The governance would be vested in an outside Board of people from the library and publisher/vendor constituencies. Privacy and security would be very important for the system.

Even if there were no fees charged by Atypon, libraries and service providers would have costs in adapting their systems.

### **RINGGOLD -- Ralph Shoffner**

Ringgold's core interest is in promoting a standard for the description of organizations -- names and identifiers -- rather than a registry per se, and Ringgold called on NISO to develop such a standard. Identification of organizations is central to their business.

Their IDENTIFY service can be seen at <http://www.ringgold-directory.com/identify/>. The service includes independently verified data on 20,000 institutions obtained from publishers' subscription databases.

Ringgold expect publishers to pay for this electronic directory, and the services they can derive from it. It is an information exchange in which each party owns its own information, and shares only with permission from the information's owner (as with Atypon). It will be open and available to be updated by those who own it. There is no use of DOIs. The service sustains itself only if it is profitable or Ringgold.

In closing, Ralph Shoffner urged that standards for the identification of organizations is needed, rather than a registry.

When asked how this differs from the Atypon plan, the response was that Ringgold "are the creatures of the publishers and are led by their needs." Ringgold provides information about institutions, cleans up data and groups customers. Ringgold is paid by the publishers. If publishers want them to store IP addresses they will. Ringgold independently checks and verifies all information entered by institutions.

External governance – the current focus for Ringgold is not to set up a registry but to focus on standards. However, Ringgold would be willing to consider external governance.

### **JISC -- Sarah Porter**

[See PowerPoint slides]

Following up on a conversation in the morning, Sarah Porter gave us a quick overview of the work that JISC has done in the UK on this issue, where they concluded that a distributed model was more appropriate for their needs than a single centralized one. JISC has a role to help publishers communicate with the universities that JISC serves (and for whom it often licenses content) and their information environment service registry (IESR – Information Environment Service Registry) records collections and the services that provide access to them, as well as information on the "agents" involved in that exchange -- publishers and libraries.

JISC did an institutional profiling study – [http://www.jisc.ac.uk/uploaded\\_documents/CMSS-Shaw1.pdf](http://www.jisc.ac.uk/uploaded_documents/CMSS-Shaw1.pdf). What they have is an institutional profile, listing contact information, telematics (z39.50 IP addresses, etc), learning and teaching system information, and subscription information. There is no central place where this resides but rather an xml schema held locally at each institution that can be accessed by web services – a machine to machine interface. Sarah thought that this suited the much more coordinated and centrally funded university landscape in the UK, but did not translate well to the US situation. She could imagine this local material being harvested from institutions to a central place. The JISC service registry is open now with material that JISC has created.

## **Shibboleth**

The session concluded with a conversation about Shibboleth, and whether it makes the central registry notion less compelling, as it does away with the need to share IP addresses in order to allow and control access.

Those who knew it best pointed out that although adoption has picked up recently, there were still only 20 institutions in the US using it seriously. OCLC will be launching a Shibboleth site soon. It is the right direction to be going but the opinion of the group was that we will have IP addresses to deal with for a long time, and that Shibboleth does not deal at all with the other information-sharing needs we had identified (contact information, for example).

At the closing of the meeting there was a good sense that we had surfaced issues from a variety of perspectives that were certainly not evident to all parties when we started in the morning; and that there was a value both to libraries and to publishers in a central registry under community ownership that made it easier to access a range of information about a licensing unit and a vendor.