

talis



A Platform Approach
to Collaboration

Impact on Union Catalogues

A Talis White Paper

Dave Barker

Version 1.0, March 2006

Contents

Executive summary.....	3
Resource Discovery.....	3
Historical View.....	4
A New Approach.....	5
New Technology.....	5
New Thinking – A Shared Platform Approach.....	5
Physical or Virtual?.....	6
Federated Search.....	6
High barrier to participate.....	6
Incomplete and poor quality search experience.....	7
Performance.....	7
Physical Union.....	7
Cost of Contributing.....	7
Currency of Data.....	8
The Optimal Approach.....	8
Many Views.....	8
Removing the Cost of Contributing.....	8
Contributing to much more than a Union Catalogue.....	9
The Talis Platform.....	9
The Talis Directory.....	9
Data Channels.....	10
Talis Developer Network.....	10
Trust.....	11
Discovery – Not just in Libraries.....	11
Conclusion.....	12
Find out more.....	12

Executive summary

With the launch of Talis Source, a powerful Resource Discovery and sharing system based on the Talis Platform, the landscape for similar systems has changed forever. Talis Source provides entry level Resource Discovery completely free of charge. This has been made possible by significant advances in technology over the past few years, combined with some completely new thinking about the challenges faced by providers of Union Catalogues and related systems.

This paper looks at the issues involved and describes how a new approach makes possible things that were in the past too difficult or too expensive to achieve.

Resource Discovery

For the purpose of this paper, we define Resource Discovery as the process of searching for and locating a particular item, whether it is a physical book, an electronic resource, an audio or visual item (CD, DVD, etc.) or anything else. In the library context this usually means looking for the item in a library or group of libraries, but as we will argue later in this paper, we believe that the search should be broader and include for example on-line retailers and second hand book shops.

With a little thought we can see that this process of discovery is at the heart of many other processes within the library and information domain. Examples include:

An academic performing research might wish to locate an electronic copy of a scientific paper from amongst the University Libraries.

An Inter Library Loan (ILL) Librarian might wish to locate a book for a library user that is not available in the local library, before initiating an ILL request.

A national government might wish to provide a 'one-stop' website to allow members of the public to search for resources within all the libraries within its jurisdiction.

A member of the public may wish to locate a resource in any of the libraries physically close to where he/she lives or works – without having to consult a series of separate OPAC's.

Historical View

Today's systems that attempt to provide Resource Discovery are expensive to build, maintain, and use, are closed, and provide non of the modern functionality expected of Web- based systems by today's generation of users. Those users, brought up on the experience provided by free sites such as Google, Amazon, Flickr, and Technorati, expect a participatory approach with zero cost of entry. There are historical reasons why current systems are what they are, but there is no good reason why they cannot now change.

Until now, systems and solutions designed around collaboration between libraries have been considered independently from many perspectives, including functional, economic, and technical.

As noted above, many different problems have Resource Discovery at their heart. Yet historically each new initiative has 'started from scratch' by considering the problem as one, rather than as a number of component parts. For example, if a regional cross-sectorial grouping of libraries wanted a shared Union Catalogue, they would build one. If a group of public libraries, that might include some of the participants in the Union Catalogue project, wanted to have an inter-lending system, a new project would be built without any re-use of software or data.

From an economic perspective, because most collaborative efforts are sponsored and funded by one-off initiatives, all of the costs are balanced with the specific benefits of the initiative, rather than taking a broader perspective. In most cases this leads to constrained functionality or no project at all.

Because public money is usually involved, many collaborative projects take a very cautious approach to technology, only considering what is already proven. This leads to projects being built on high cost technology that is obsolete by the time the project is complete. There is also very little re-use of technology, again because of the fact that projects are considered in isolation.

A lot of time and money has been spent evaluating the possibilities for future Resource Discovery systems. One such is the UK National Union Catalogue. It is now five years since the final report on the feasibility of such a project, yet nothing has happened (except the serials catalogue). Meanwhile the so called Web2.0 companies (Google, Amazon, etc.) are showing what can be achieved in the broader Resource Discovery field without any input from the library domain at all.

A New Approach

Several factors are now coming together that will change the nature of Resource Discovery forever. New technology, combined with new thinking about the challenges will deliver revolutionary new solutions.

New Technology

Over the past few years the inevitable progress of technology means that now new things are possible at dramatically lower costs. Some of the most important changes include:

- Near universal connectivity through the Internet and World Wide Web.
- Commoditisation of hardware. Whereas once large database systems required large, expensive servers, it is now possible to utilise low cost, commodity hardware to achieve the same thing.
- Dramatically lower software costs. Whilst good, robust software still requires significant resources to build and maintain, new business models in the software industry have significantly lowered the costs to users. The open source community deliver very low cost infrastructure software such as operating systems, web servers, and databases. Furthermore, models such as software as a service, and network software platforms, enable providers to share the costs across many users in a highly scalable way.

New Thinking – A Shared Platform Approach

Concepts of sharing and community are native to the library world. Yet the systems in use today – the Library Management Systems, and the shared Resource Discovery and Sharing systems ignore these principles. Most are closed and proprietary. Taking a more open, community approach opens up a host of new opportunities for the future of library systems. Also, by taking a much broader view of the economics involved, it is possible to create an incredible amount of additional value which can be used to lower the cost of participation.

First, the core data needed for Resource Discovery should be regarded as something that should be as freely available as possible. The data is owned by the holding institutions, but it is in their interest to make it visible in as many places as possible. Yet today, bizarrely, institutions may have to pay three times over to make that data available outside their Library Management System – once to extract the data, once to have it loaded into a Union Catalogue, and finally by paying a subscription to use the discovery service. As a principle, we believe that all three of those costs should be removed completely from the system.

New licensing mechanisms such as Creative Commons protect the data owners from commercial exploitation, while maximising the data's reuse for value creation.

As noted above, if all of the costs associated with building a Resource Discovery system are charged back to the users of that system, the result will be high subscription levels and low participation. But by recognising that there is significant additional value that can be created by combining and aggregating data, it is possible to use that value to offer low or zero cost of entry – which will increase participation and in turn increase the value created – a truly virtuous circle.

By taking a platform approach, it is possible to take further costs out of the system. One centrally hosted platform, delivering powerful functionality and rich data, can be used to build many different systems effectively and efficiently. Because the platform performs the 'heavy lifting' (in other words the 'difficult bits'), the applications developer can concentrate his efforts on improving the user experience.

Physical or Virtual?

There has been much debate about which is the best way to build a Union Catalogue – by creating a physical central database, or by using federated search techniques. Five years ago one study concluded, that with respect to cost, there was little difference between the two approaches. Traditionally the drawbacks of each approach have been as follows:

Federated Search

High barrier to participate

The federated search approach pushes the cost, risk and complexity out to every institution that wishes to be searchable. Each institution must buy, install and run a Z39.50 server that enables access to its database. Many public libraries have not bought Z-targets from their vendors and some vendors do not support z-targets at all. The cost and time required in getting a z-target installed and running at every library that wants to participate is not small. Further Education library solutions do not typically support Z.39.50 at all. It should also be noted that hardware upgrades may be required to cope with the extra load placed on the individual institutions.

Incomplete and poor quality search experience

There is a big difference between identifying the book a user wants and identifying where copies of that book might exist. Federated search is very poor at helping the user identify the book of interest. For example, when a user enters a search term and gets no hits, the system is unable to tell the user if the reason for no hits was because no books exist in the world with that title (perhaps because the title was mistyped) or simply that no library in the system has a copy. In addition, the varying cataloguing practices at each library, combined with the varying functionality of each Z-target, results in inconsistent searches.

Performance

If a good user experience is required, then results from the similar sources are usually merged so that instead of getting a separate hit count from each public library, their results are merged. This implies that the system will perform in line with the slowest z-target. The alternative, where results cannot be merged, leads to a poor user experience.

In addition, every single user search on the system results in a search hit on every participating institution, so that, in effect, every institution must scale their systems to support the level of usage traffic the Resource Sharing solution must handle. This is a significant scaling issue should the system become popular.

Physical Union

Cost of Contributing

As noted above, the cost of contributing to physical Union Catalogues has been very high. This is mainly due to the closed, proprietary nature of most Library Management Systems. Vendors had an interest in trying to keep the data within the LMS, so that they could sell additional software – if the data could easily be accessed, then a competitor could provide additional functionality.

Currency of Data

Because of the cost of contributing and the limitations of older networks, data in physical catalogues was often out of date. And it was certainly not possible to obtain 'live holdings' data from a physical catalogue.

The Optimal Approach

We at Talis believe that the optimal solution with today's technology is a hybrid of the two approaches. Start with a physical Union Catalogue. This provides fast, consistent searching, independent of the availability of any local systems. Then remove the costs of contributing – see below. Use new techniques such as RSS feeds to improve the currency of data. And finally provide deep linking into the target OPAC to provide real-time availability and allow the user, if authorised, to perform reservations and other activities.

Many Views

With a large volume of related data held in a centrally hosted database, it becomes very easy to provide many different 'views' of that data, segmented in any way that makes sense. So, for example, one view could be of all collections physically located in a particular geographic region – thus delivering a regional, or national, union catalogue. Another view could show all collections associated with university research libraries – thus creating a powerful Resource Discovery tool for academic researchers.

Removing the Cost of Contributing

As noted above, one large barrier to participation in Resource Discovery solutions is the high cost of contributing. Talis is working hard, through a number of initiatives, to remove both the technical and the economical barriers to contribution. These efforts include:

- Working with others in the industry – LMS vendors, Union Catalogue providers, and standards bodies to agree on common standards for the exchange of holdings information between organisations.
- Until such standards are agreed and implemented, Talis will accept any reasonable format of data. This means that if an institution is already contributing holdings to another Union provider, we can take a copy of the same file, so that the institution does not have to perform multiple extracts.

- Working with LMS vendors and individual institutions to reduce the technical complexity of the extract process itself. Wherever possible, through the Talis Developer Network, we encourage the sharing of scripts and other extract devices among institutions. We are researching automated processes for data harvesting from the local systems, and developing data appliances that can be installed alongside the LMS to provide open data feeds.

Contributing to much more than a union catalogue

The power of the platform approach is illustrated very well when you consider that by contributing holdings and other information to the Talis platform, a library is enabling much more than a Union Catalogue. The key is in data aggregation – the fact that data summarised from many sources creates enormous additional value for all participants. One example is borrowing statistics for out of print books – when aggregated nationally this information is extremely valuable to publishers when considering which items they should bring back into print. Comparing borrowing patterns with holdings regionally, can also start to inform re-allocation of stock between institutions.

The Talis Platform

The Talis approach to Resource Discovery opens up many new possibilities while keeping the costs down for participating institutions. It is based on the concept of a shared platform of data and functionality, and critically separates the costs and the benefits. Talis is investing now to deliver free entry level services that will in turn create additional value for all involved.

The Talis Platform has several main components, each of which contribute to changing the way that we think about Resource Discovery and Union Catalogues.

The Talis Directory

The Talis Directory is the link between all the participating nodes in the extended network. Crucially, it maintains details of each collection known to the system and how that collection can be accessed electronically – through a Z-target, OPAC, or another means. The Z-target might be used for harvesting real-time holdings information. The OPAC details are used to provide deep linking into the institution's catalogue.

The Talis Directory has been designed as an open, participatory component of the Talis Platform. The services that it provides are freely available through the Talis Developer Network for others to use within their applications. Maintenance of the data is also open, so that anyone who has information about a collection and its access methods can update the Directory Details.

Data Channels

An important difference between the Talis Platform and more traditional computing platforms is that the Talis Platform delivers both functionality and data. The Talis Platform contains a huge amount of library domain content – bibliographic records, holding records, enrichments, etc. This is a mixture of commercially sourced data and data contributed by the community that is using the Talis Platform and the applications built upon it. The volume of data is growing all the time. This data is then available to application developers through a series of Web Services, enabling them to assemble powerful applications with relative ease.

One way to think of the data channels is to compare them with television and radio channels served up by satellite and cable providers. Many individual channels, often with strong brand identity, are available to consumers in a variety of packages to suit their individual needs. There may be basic packages that are free or carry a low subscription, and then premium content at further cost.

Talis Developer Network

The Talis Platform has been designed to be very open and to offer exceptional opportunities for developers to build their own applications on top. To a developer, the Talis Platform appears as a series of APIs, implemented as Web Services that can easily be assembled and combined with services from other platforms, into powerful applications with relatively little effort. Over the past few months Talis has been demonstrating a powerful prototype application¹ that was built entirely in a browser by one person in less than two weeks.

The Talis Platform is of particular interest to existing Library Management System vendors who can easily incorporate additional functionality and data into their products – providing additional value to their customers, at a much lower cost than doing all of the development themselves.

¹ Whisper – <http://research.talis.com/2005/whisper>

But it is not just LMS vendors who can benefit greatly from Talis' Platform approach. It is now possible, using platform components, to build a regional or consortia Resource Discovery system for very little capital outlay, and very low ongoing costs.

Trust

By contributing data to this collaborative platform, institutions are continuing the tradition of sharing and co-operation that has characterised the library world for generations. Unlike other data aggregators, Talis recognises explicitly that the data contributed remains the property of the contributing institution and is free to remove that data at any time. A simple contribution agreement, modelled on the Creative Commons Licence, clearly describes the rights of both the data owner and the provider of the platform.

Discovery – Not Just in Libraries

As mentioned earlier in this paper, traditional Resource Discovery systems in the library domain search for resources only in libraries. In today's connected world this no longer makes sense. The general purpose search engines such as Google and Yahoo! will search anywhere that they can get to - although in general today that includes only the open Web. A member of the public, or a researcher looking for a particular item, for example a rare book, should be able to search all the possible places that might hold that item. This includes libraries, bookstores (both new and second hand), and even private collections of books belonging to members of the public. The new world of connected platforms makes it possible for one application to be able to search many different sources simultaneously.

But it is important to remember that in many cases Discovery is just the first step. Usually the user will want to obtain the resource - whether that be a physical item or an electronic resource – for their own use. Therefore the platform must deliver additional components and services, or link seamlessly to other platforms, to provide the delivery mechanisms.

Conclusion

What we have attempted to show in this paper is that new technology and new thinking have dramatically changed the world of Resource Discovery and Union Catalogues. The Talis Platform provides powerful functionality and data to drive the next generation of applications. Talis Source, an application that includes free entry level Resource Discovery is a good example of this. The Talis Platform is available now and over the coming months Talis will be announcing additional applications developed by third party providers.

Find out more

If you would like to receive more information about how the Talis Platform will improve and enrich the Resource Discovery and Union Catalogue landscape, then please contact Dave Barker by email: david.barker@talis.com or call 0870 400 5000.



Copyright 2006 © Talis
www.talis.com