

# Building Web scale for libraries

Robin Murray  
Vice President, Global Product Management  
OCLC  
Robin.murray@oclc.org

**Abstract:**

*Picture a virtual city of key Internet payers. Google has a main street address as do Yahoo! and MSN, but where do libraries reside? Change is afoot, based on a need from libraries and suppliers alike to reduce redundancy and generate what we define as Web scale services. This paper explores what we mean by Web scale and how libraries can achieve it.*

## Introduction

Picture a virtual city made up of key Internet players. Google has a Main Street address. So does Yahoo! and MSN. You can see del.icio.us and Amazon from the corner, as well as eBay, MySpace and YouTube.

Where are libraries? Unfortunately, they are hidden on a back street, away from most of the traffic.

Meanwhile, in the library software environment, questions are being asked of the systems being used and whether they continue to be fit for purpose. Business models are changing as greater economies of scale are being sought by libraries.

Economic and cultural factors are driving the library community to re-evaluate their strategies, to put them in the mainstream traffic flows of the Web.

This paper will define Web scale, what it means for libraries and their users, and describe OCLC's strategy to obtain it. It proposes a plan to build Web scale for libraries, a plan that offers a direction that makes it easy for libraries, collectively and individually, to deliver their services to the network and build a unified, high-value presence on the Web.

## The network rewrites the library

Each day, users conduct 2 billion<sup>i</sup> searches on the Web. About 80 million users<sup>ii</sup> go to eBay, and another 65 million to Amazon<sup>iii</sup>. There are 62 million active users on Facebook<sup>iv</sup>, which grows by 100,000 new users every day. Apple Computer has sold 3 billion songs through its iTunes Music Store<sup>v</sup>, with about 2.5 million songs being downloaded each week.<sup>vi</sup> University portals and learning management systems are network destination sites for faculty, staff and students.

Clearly, as the numbers above suggest, we are in a period of rapid, permanent technological and user behaviour change caused by the World Wide Web. Indeed, it may not be overstatement to say that the Web is the most significant engine dictating change since the advent of the computer. In 10 short years, the Web has become central enough to what people do that it is where they live and where they get things done. More and more activity is migrating to the ubiquitous network.

Why? The following environmental trends document and explain the move to the network.

- We all personally manage large quantities of digital information: documents, images, music. Whether for personal, professional or educational use, people are growing accustomed to being able to 'gather, create and share' digital resources. Facebook

and MySpace illustrate the power of social networking sites as resources are quickly shared through the communities they create.

- People want digital. Where digital is not available, they want predictability and transparency in the supply chain. Compare UPS and Amazon for instance, where items can be tracked and monitored from the moment the order button is pressed. The Google mass digitisation and similar initiatives are changing the landscape of book consumption.
- There has been a radical transformation in how people find and use information over the last five years. Major information and communication hubs—Amazon, iTunes, Google, MySpace, Realtor.com, eBay—exercise strong gravitational pull through a consolidated presence on the network and the ability to find obscure pieces of information more easily than ever before. They account for the majority of access to other content through referrals.
- Information is increasingly created and consumed within integrated environments rather than as a stand-alone activity. Learning, research and personal development activities have entered the network space. ‘Workplace applications’ are being tailored and personalised, generating an increase in the rate in which data is packaged, processed and manipulated. With personalisation comes increased relevance to the audience and a growing desire to share and replicate data.

This fundamental shift in consumer behaviour is exposing weaknesses in the library operating landscape. Whilst libraries have a strong presence in their local geographies, their resources and expertise are largely hidden in a global Web environment dominated by large-scale information hubs. Users must exit their preferred Web environments to come to the library to use its services.

In addition, library Web sites often present an array of complicated ‘islands’ of content — OPACs, licensed e-resources, digitised collections, metasearch engines, and institutional repositories — each with their own user interface and each with very little integration. Libraries need to plan for and build tools that fit new consumer habits, with an emphasis on the flexibility and remixing of their content and services.

- Service fragmentation. Consider the complexity of the library systems environment, and how difficult it is to map onto users’ expectations and behaviours. Libraries operate within linear structures, institutional, governmental or corporate, their missions are to service individual communities – student, citizen, employee. But in order to have a significant presence at a network level, libraries must reach beyond these boundaries. By not doing so, library resources are being shallowly represented in the place where supply and demand is increasingly aggregated—on the Web.

- Redundant effort. Libraries are vertically integrated around their collections and carry out routine work, such as cataloguing and acquisitions, in redundant ways, often handling records multiple times.
- Library budget pressure. Library budgets are under pressure: steady-state at best, and regularly declining. The budget is largely absorbed by legacy service management issues, while there is significant pressure for the delivery of new services.

This network transformation has implications for libraries and their service development.

- Brand is the new real estate. A few Internet brands have evolved into central parts of people's lives. And their dominance is growing. Libraries must build a similar, unified, high-value brand on the Web.
- Library services must fit into the user workflow. It used to be the user built their workflow around the library. Now the user is on the network and wants access to library resources but without gateways, and within their existing and familiar workflows. Libraries need to integrate their services and collections with these emerging user Web platforms.
- Attention is scarce and resources are abundant. In pre-network days, attention was abundant and resources were scarce. Today, libraries have to compete for attention. People are living in a network environment where there are abundant resources, and the library is one among many. Library resources have to be easily available in ways that make sense to users.

## **Creating Web scale to meet users at the point of need**

Today's users want the universe of resources, including library collections, at their fingertips, presented as part of their Web experience. To meet this expectation, libraries must place their collections and services within popular, emerging Web platforms — search engines, Internet booksellers, social networking sites, course management systems — to deliver their expertise at the point of need as well as drive traffic back to the local library.

Building such a Web scale model, which complements local and group efforts, requires that libraries develop a global destination and aggregate their presence and services on the network to achieve full capacity. Right now, library databases, collections and services are fragmented, and not in mainstream Web traffic flows — the search engines and Internet services where most people start their search.

The OCLC strategy focuses on building Web traffic flow to libraries by leveraging our unique position as a provider of systems and services to libraries at a local, group and global level, and our unique ability to link them all together so that each one adds value to the other

Throughout its existence, OCLC has been about reducing duplicate effort in libraries by moving functions to the network where appropriate and thereby making libraries more robust through collaboration.

As a global organisation, we can create efficiencies and impact (Web scale) that are not possible by each library acting individually. In 1967, the power of the network inspired OCLC Founder Fred Kilgour to bring together the cataloguing operations and expertise of libraries into a single system that allowed them to improve productivity and share resources. This was the foundation for OCLC members to begin making local, regional, national and global impact. OCLC has created complementary networks for resource sharing, collection analysis, virtual reference and storage and preservation.

Today, the power of the network is familiar to us all due to the birth and expansion of the Internet/Web, the ultimate network. Our strategy is to create Web scale for libraries—to move more and more of the library's resources (collections and services) to the network. We are doing for libraries what eBay has done for auctions.

At the heart of our strategy is the recognition that the assets that a library possesses need to be surfaced at multiple levels. Today, this is being done in a number of ways.

A library has many assets (collections, services, expertise), around which a series of business processes, such as cataloguing, collection development, interlibrary loan and virtual reference, have been developed over many years. See Figure 1. And, these assets are exposed outside of the library through a number of presentational layers, which can be categorised in the following ways:

- **Management Services** help library staff to effectively describe, configure and manage their resources. These services offer an internal window on the assets that are being managed.
- **User Services** connect end-users to library resources, typically through the library's OPAC, but also through group or union catalogues.
- **Network Interface Services** make library resources accessible through machine interfaces so that they can be remixed in other Web environments. At their earliest inception, such interfaces were based on domain-specific protocols, like z39.50, to enable search and retrieval of data. Today, standards have become far more open, enabling adoption beyond the library domain.



Figure 1: Library Assets

The development of network interface services, or Web services as they are now more commonly known, enables seamless presentation and consumption of services across the network and provides the capacity for data and service sharing across nodes: local, group and global. All nodes can interact and leverage services and resources from one another. Functionality resides at the most appropriate node in the network, and all nodes can add value to and gain value from the network.

OCLC's unique positioning as a provider of systems and services creates a special value proposition when considering how libraries can present and consume services across the network. We already offer, as an example, data across the network to serve libraries at the local, group and global node, as the cataloguing services testify. The strategy is to extend this proposition; to offer data, network services and reusable software components that can also be exposed on the Web, using industry-wide protocols, enabling applications from a variety of providers to consume them, not just OCLC applications. These are known as WorldCat Grid Services.

The WorldCat Grid is a standards-based platform that makes it possible to quickly link together computer systems across organisations worldwide. It provides an open infrastructure for building a Web-scale library service in a view that is appropriate for each library.

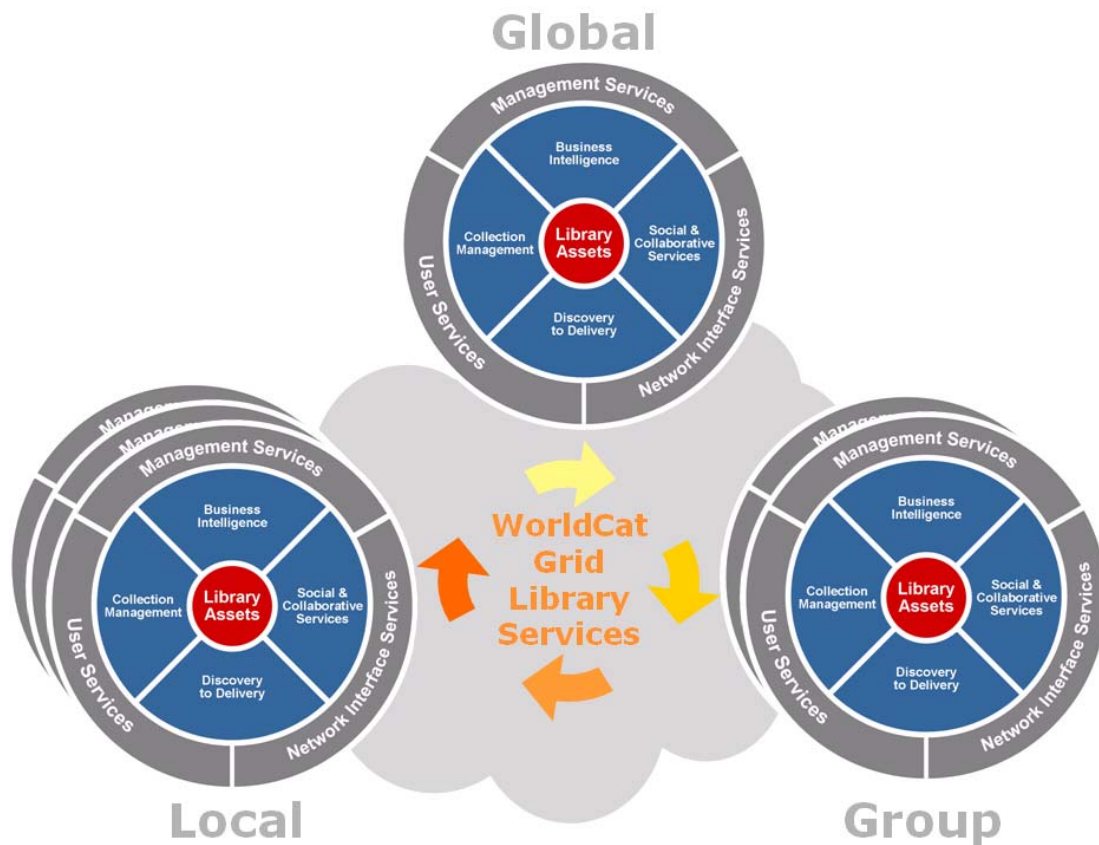


Figure 2: WorldCat Grid Library Services

Applications and services using WorldCat Grid Services will allow libraries to manage and provide their resources at local, group and global levels, and to benefit from scale and network effects. For example, libraries can opt to use a local library management system, a hosted system or a combination of the two. In addition, they can present themselves in local, group and global views to bring searchers to their content and services.

The WorldCat xISBN service, which supplies International Standard Book Numbers (ISBN) associated with individual works in WorldCat, is a good example of a grid service. Any organisation interested in linking an application to it can do so. A number of libraries are using this service to weave their collections into Web bookstores, by linking ISBNs from the bookseller site to the library catalogue. If the library owns the book, or any associated edition, that the user is looking for, a record or list of records appears on the screen complete with shelf status information.

The WorldCat Registry is a web-based directory for libraries and library consortia. Its proposition is to provide the mechanism for managing an institution's identity centrally in order that it can be surfaced and used in multiple service scenarios. Profiles are managed and created centrally and information sharing with vendors and others is automated to ensure that up-to-date information can be made available instantly.

The Grid enables data and services submitted or created centrally to be reconstituted for use at a local and group level. By doing this, the value of the network can be realised in the individual operational activities of institutions. A correlation starts to exist between libraries that participate in network-level initiatives facilitated by the Grid and savings in staff time and cost.

## Delivering Web scale through participation and momentum

With an open infrastructure and grid services that support seamless interoperability and system linking, libraries have a shared network platform to collaboratively build a Web presence with global impact. The platform can put libraries at the point of need and can make the network work for libraries. It can make possible a library experience for users that match the experience of the Web. It can create Web scale for libraries.

Momentum builds and impact grows incrementally as more libraries around the world participate. A “flywheel” effect takes over—a virtuous circle of activity that gradually builds Web scale and in which each step supports the next and reinforces the last. See Figure 3.

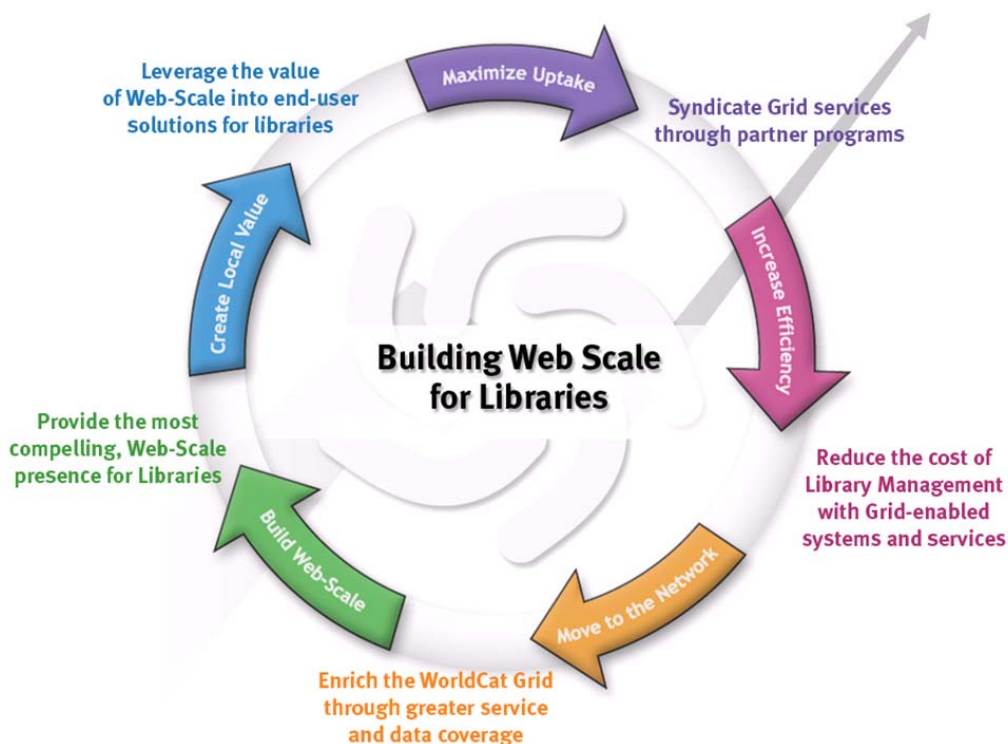


Figure 3: Building Web Scale for Libraries

The concept of creating Web scale starts with an application—an application that demonstrates in actual terms the notion that libraries can provide a compelling Web scale presence.

1. **Providing the most compelling Web-scale presence for libraries.** This means making the Web experience as good as it can be: compelling for libraries to be a part of and compelling for end-users. WorldCat.org is the first step in providing a global discovery tool and a Web-scale presence for all libraries to better serve users' needs. And if users are compelled by a service, it means they are not just passive consumers of it, they want to participate in building it. As we see time again, the Web scale services that draw the biggest crowds are the ones that offer an opportunity to socialise. Socialising in WorldCat.org has begun with the ability to create lists and reviews already under way.
2. **Leveraging the value of Web-scale into end-user solutions for libraries.** The Web-scale presence and value of WorldCat.org can be brought down from the network and leveraged into premium solutions for local and group libraries. This means integrating discovery into distributed end user solutions and driving Web traffic into local environments to deliver what the user needs. WorldCat Local is a new service that brings together local discovery and delivery within the global network. Through a locally branded interface, the service provides libraries the ability to search the entire WorldCat database and present results beginning with items most accessible to the user. These include collections from the home library, collections shared in a consortium and open access collections.
3. **Syndicating WorldCat Grid services through partner programs.** The entire library network needs to be involved in producing Web scale. For our part, we will syndicate our services to libraries and third parties to help them generate high quality services. Sharing data and application components will maximise uptake and contribution for a Web scale presence for libraries. For example, the British Library Document Supply Centre is mixing WorldCat Resource Sharing applications with its local interlibrary loan service.
4. **Reducing the cost of library management with grid-enabled systems and services.** Building Web scale and generating network effects will reduce the rate of increase in cost and increase the efficiency of library management because libraries will operate in a way that takes advantage of and adds value to the network. Using the grid, libraries and their partners will not need to build all software from scratch. They can create applications that mix and build on top of other applications. Redundancies are minimised, economies of scale are realised and libraries focus on creating local value for their users. Baker and Taylor, a leading supplier of materials to libraries, uses the WorldCat Grid to add unique OCLC record numbers to the 4 million items in its database, as well as to contribute new records to WorldCat. This collaboration reduces the cost of cataloguing by making metadata available to libraries earlier in the acquisitions process.

5. **Enriching the WorldCat Grid through greater service and data coverage.** As more libraries join, more services are shared, and more library data is available. Mixing services at the network, group and local level increases the richness of library services and makes Web scale possible. The WorldCat Link Manager brings together functionality from local library systems and the grid to route searchers to their libraries electronic full text collections. This grid-enabled service creates additional capacity for libraries beyond what they could create for themselves. The libraries' Web presence becomes more compelling, and users see library resources wherever they are on the Web.

## **Organising libraries for Web scale: Where do we go from here?**

The Internet has changed the way we work, live and play—how we buy things and choose our entertainment. Why wouldn't it change the way libraries interact with themselves and the public?

The current state of libraries on the Web can best be characterised as a fragmented and redundant approach for bringing the resources of the world's libraries to the Web environment. There are dozens of library communities, hundreds of library portals, and thousands of library catalogues that describe the resources of individual library collections. The strategy of building Web scale creates a unifying force that aggregates the power of libraries on a global scale and allows them to achieve equal footing with other Internet hubs. It will give libraries a Main Street address in the virtual city of the Web.

Every library needs to be part of building Web scale. To get there requires a strategic approach which complements existing effort. By building on the interaction of libraries locally and within formal group arrangements, libraries can aggregate their presence and services on the network to optimal effect.

The OCLC strategy focuses on building Web traffic flow to libraries by leveraging its unique position as a provider of systems and services to libraries at a local, group and global level, and its unique ability to link them all together so that each one adds value to the other. The WorldCat Grid provides the open infrastructure required for achieving this.

## Notes

---

- i <http://www.comscore.com/press/release.asp?press=2018>
- ii <http://www.comscore.com/press/release.asp?press=2000>
- iii <http://www.comscore.com/press/release.asp?press=2000>
- iv <http://www.facebook.com/press/info.php?statistics>
- v <http://www.apple.com/pr/library/2007/10/17itunes.html>
- vi [http://www.news.com/2102-1027\\_3-5173115.html](http://www.news.com/2102-1027_3-5173115.html)