

Interoperability & the pioneering role of the Library of Congress,

Face-to-face with Jane Mandelbaum and Martha Anderson

“The collections of the Library of Congress constitute the most comprehensive repository of human knowledge in history. Today, extraordinary advances are revolutionizing scientific research and technology discovery at an astonishingly fast rate.”

Peter Young, Library of Congress, U.S.

The Library of Congress (<http://www.loc.gov/index.html>) is playing a pioneering role in tackling such challenges not only on an institutional level but also through the National Digital Information Infrastructure and Preservation Program or NDIIPP for short (<http://www.digitalpreservation.gov/>). Its mission is to develop a national strategy to collect, archive and preserve the burgeoning amounts of digital content for current and future generations. With a preservation network of over 130 partners, the Program is based on the understanding that digital stewardship on a national scale depends on public and private communities working together. DL.org interviewed Jane Mandelbaum, Manager of Special Projects, Information Technology Services and Martha Anderson, Director of NDIIPP, to explore top-level challenges and milestones achieved.

What are the main challenges surrounding interoperability related to the management of digital content?

One of the main challenges the Library of Congress and the U.S. National Digital Information & Preservation Program faces is the very broad range of communities we engage with that produce and preserve content. These communities span commercial organizations, such as movies, TV and radio; universities, research institutions and other cultural institutions that collect and preserve content from more open channels like the Web and the geospatial and geographic communities. Each of these communities has its own approach to describing, preserving and formatting content. This diversity is a major challenge for interoperability.



Main Reading Room, Library of Congress
Image courtesy of the Library of Congress

What are the main requirements for users of the Library of Congress and how does your mission and work aim to respond to these needs?

There are different viewpoints that we need to take on board when it comes to users. The main set of users are the American people who contribute and actively use the content of the Library of Congress meaning that the library nurtures the preservation of different kinds of content. The Library specifically serves the U.S. Congress, the American people and citizens from all over the world. The Library provides digital content, particularly primary source materials, and research products that support public policy research, educational institutions (including teachers and students) as well as researchers and general public with an interest in cultural heritage and public policy. A major challenge for the Library is that all of these people need to access content in way that is user-friendly and that enables users to discover the information that is useful to them, so we seek to understand how we can cater for these needs.

The vision and strategy of the National Digital Information Infrastructure Program is to preserve content for the years to come through a nationally distributed system under the stewardship of distributed organizations. The main aim is to understand how best to bring together the huge collections and understand what's in the content and items, bearing in mind the different institutional needs and capabilities.

What is the overall expected impact with regard to the long-term strategy for preservation?

There is a growing volume of content and diversity of content under national stewardship. As this volume grows so do the expectations of the users, who want to do different things with the data, not only discover it but also use it in their applications. Indeed, in recent years, students have increasingly been using the services for their studies. As mentioned earlier, making this data user friendly and easily accessible to users is a major challenge that the Library is confronting. Furthermore, current access is a prerequisite for the long-term preservation strategy as we seek to build on and improve this moving forward. Engagement with both users and creators of data from a broad range of communities is a vital part in addressing this challenge.

How does the program for interoperability build on existing foundations?

The Library of Congress is a leader in the development and management of standards in the library world (www.loc.gov/standards), and has been leading and nurturing standards over the years. One of the pioneering roles played by libraries over the last 40 to 50 years is the exchange of documents, which has built trust and a good foundation on which to build interoperability. Trust is key to interoperability and an important part of the foundation is when both sides work to make interoperability possible. Trust has been a very important part of the Library's role in the community. The Library strives for agreements with other institutions to build on mutual strengths and develop roles and services to support the management of the digital lifecycle. These agreements are the building blocks for the future and provide a model for developing stewardship in the next generation. Through the Program partners

have affirmed the value of our leadership and role in serving as a neutral broker across industries. In this respect, the Library acts as a convener, helping to connect and catalyze different communities.

From a technical perspective, how are you tackling the metadata challenge?

One of the key issues that has emerged is the need to search and discover information in ways that are useful to the members of different communities, which is not as easy as it may seem. When metadata comes into play, it becomes apparent that it is never perfect and never complete, hence we need to understand what specific metadata is needed by the target users. For decades discussions among cultural heritage institutions have been dominated by the idea of the "perfect set" of metadata that is globally applied by information managers and users alike. Future information environments are likely to allow for variations in metadata implementations and encourage participation rather than enforce it. In these environments, infrastructure provides the tools to mediate between different data, to simplify metadata creation and to raise its quality. Seemingly simple things can turn out to be utterly complex: e.g. how to convert existing long titles of history videos into short titles as required by *YouTube*.

At a workshop in June 2009 led by the Open Grid Forum's Digital Repository Research Group (OGF26, 25-29 May 2009, Chapel Hill, U.S. - http://www.ogf.org/gf/event_schedule/index.php?id=1738), this point resonated with other participants who also agreed that automatic metadata creation/conversion, and encouraging early metadata creation were amongst the key opportunities in their communities.

Our approach is to focus on specific key elements of metadata and understand how users perceive information. If we take, for example, geographic metadata and online maps, one can provide useful search and discovery of cultural heritage materials in a context of place that people find useful. Specific contexts are useful pointers for users to access information. Time and time-periods and contexts such as events and place are a good case in point. With the broad diversity of descriptive elements of metadata from various communities that the library deals with, time and place are often elements which bridge data diversity enabling users to navigate content. This is part of our focus, along with web-

based tools that can be applied to metadata. While the communities are diverse, we can pinpoint a set of commonalities. These common elements help bridge diversities and help people navigate in the sea of digital content, bearing in mind the different data that people need and fostering broad and interdisciplinary approaches.

In many respects, information from harvested and archived websites is one source that has brought data from unexpected areas. The Web has brought not only documents but also other formats, hence a rich source of information about national events, such as elections or natural disasters. Understanding how this information is brought together has been a valuable asset. Such events serve as an important catalyst, building a community around a set of information stemming from an event and enabling us to work towards coherence. This entails a lot of interaction with library staff, who deal with requests from all over the world. It is important for the library to tap into the global community of library staff in order to address another layer of interoperability.

What targeted improvements has the evaluation of use cases brought to light?

Use cases play an essential role in working towards interoperability. We cannot simply focus on all the requirements of users, so the use cases we have evaluated are very important for pinpointing specific needs and requests from users to which we seek to respond. One example is our work with educational outreach staff dealing with teachers and their need to teach children how to use primary sources, for example on a specific state. By leveraging metadata around specific use cases we can improve the services provided.

What achievements have been made to date and what challenges lie ahead?

One of the important stepping stones has been fostering collaboration on different levels. No single institution can achieve interoperability by itself. Collaboration and mutual trust have proven to be invaluable and will continue into the future. Major achievements have been a series of social successes, as opposed to technical achievements, such as building bridges, creating new communities and providing entry into existing communities, which in

turn ensures engagement and enables us to understand the challenges they face.

Interoperability is the most challenging of all. This approach means we can look beyond the local context towards a more diverse set of content and pinpoint communities of best practices, key to interoperability. Another achievement is fostering better metadata in new ways. We have a use case where photographers are encouraging commercial vendors to improve metadata.

In terms of the challenges ahead, XML has proved to be a valuable tool to bridge diversity but when it comes to growing volumes and mass, such as hundreds of thousands of items, then we are facing a huge challenge. Different communities have different needs. The key is to understand these different roles and foster continued engagement,



LIBRARY OF
CONGRESS



NATIONAL DIGITAL
INFORMATION INFRASTRUCTURE
AND PRESERVATION PROGRAM

moving forward.



Jane Mandelbaum, Manager of Special Projects, Information Technology Services at the Library of Congress, is currently guiding enterprise-wide projects and architecture initiatives for large-scale high-performance digital storage and

archiving. Jane works with users, system engineers, high performance experts and library managers to define, design, develop, implement, operate and model projects and systems for an environment with a multi-petabyte capacity.

Previously at the Library of Congress, Jane served as IT implementation and operations manager for the Library's first integrated library system, the largest known single-library metadata catalog. She led the team to establish and operate Library's LAN and workstation environment; and served as the automation officer for the National Library Service for the Blind and Physically Handicapped.



Martha Anderson, is Director of Program Management for the National Digital Information Infrastructure and Preservation Program (NDIIPP) at the Library of Congress. Early in the program, she participated in the Preservation Architecture Working Group and managed the Archive Ingest and Handling Test (AIHT) with four University partners – Harvard, Stanford, John Hopkins and Old Dominion University. The AIHT, the first practical test of the NDIIPP preservation architectural model, simulated migration and transfer of an archive of web content over time. Martha also manages the archiving of web content in support of the Library’s Digital Strategic Initiatives Program. She serves on the Steering Committee of the International Internet Preservation Consortium (IIPC), an international organization of 37 national libraries and archives dedicated to collecting and archiving significant content from the Web.

DL.org (www.dlorg.eu)

DL.org is aimed at consolidating and enhancing the **DELOS Digital Library Reference Model**; identifying effective methods for achieving interoperability among Digital Library (DL) systems, pinpointing and promoting best practices and successful technological approaches on key aspects of DL systems. DL.org will also deliver a **DL Technology and Methodology Cookbook**, a portfolio of best practices, patterns and solutions to common issues faced when developing large-scale interoperable DL systems, as well as guidelines for selecting the appropriate interoperability techniques, standards and approaches when implementing interoperable DL federated systems, or for describing resources which are shareable across systems.

These objectives are being achieved with the support and active contributions of the international DL research community, specifically through six **Thematic Working Groups** and a **Liaison Group**, thus ensuring participation across the board to advance the frontiers of knowledge on DL interoperability.

DL.org Consortium



National Research Council of Italy,
Institute of Information Science & Technologies
(CNR-ISTI)



University of Athens



University of Glasgow | Humanities Advanced Technology & Information Institute



Trust-IT Services Ltd

DL.org is a two-year Coordination Action (2008-2010) funded by the Commission of the European Union (EC) under the 7th Framework Programme ICT Thematic Area “Digital Libraries and Technology-Enhanced Learning”.

