The World Digital Library: Project Manager's Report

WDL Executive Council Meeting
Munich, Germany
November 14, 2011

This report summarizes the main developments in the World Digital Library since the last meeting of the WDL Executive Council in January 2011. Progress is discussed with reference to the priorities outlined in the WDL Business Plan, as revised and approved by the Executive Council at the January meeting.

Medium-Term Priorities

Addition of Content

In January 2011 the WDL contained 1,350 library items from 38 institutions comprising 98,278 images. The medium-term goal specified in the WDL Business Plan is to reach 6,000 library items, a net addition of 4,650 items. By November 14, 2,700 items comprising approximately 115,000 images will have been added to the WDL, bringing the total site size to 4,050 items (approximately 212,000 images). An additional 600-700 items are scheduled for release in the remainder of 2011, bringing the year-end total to more than 4,600 items (approximately 240,000 images). The initial milestone of 6,000 is expected to be reached in early 2012.

Content was contributed by 44 institutions from 26 countries. Collections or individual items have been added in each of the nine areas of emphasis identified in the business plan: (1) Mesoamerican codices; (2) Arabic manuscripts and books (with particular reference to scientific works); (3) Chinese rare books and manuscripts; (4) content from European partners; (5) content from Latin American partners; (6) scientific works; (7) content from WDL-sponsored digitization centers in Iraq and Uganda; (8) items from the international collections of the Library of Congress; and (9) content from and about three large countries (India, Indonesia, Turkey) currently underrepresented on the WDL.

Progress has been especially marked in areas 1, 2, 3, 4, and 8, where efforts have been most concentrated.

Three institutions – the Austrian National Library, the National Library of Norway, and the National Library of South Korea – contributed parts of collections listed on the UNESCO Memory of the World register.

While the size of the WDL increased, the strong emphasis on quality and providing users with added value has been maintained. Supplemental cataloging, expert

curatorial descriptions, enhanced viewing, and presentation in the seven WDL interface languages are provided for all WDL items. The item-level descriptions for several categories of content -- notably Chinese rare books and manuscripts and Arabic scientific books and manuscripts -- include a significant amount of original research undertaken by scholars and expert curators. These experts have produced authoritative metadata and descriptions of important books and manuscripts that heretofore have never been fully cataloged or described.

Content on the WDL is comprised of a combination of selections from larger collections and entire collections. With the addition of the Prokudin-Gorskii collection of early color photographs from the Russian Empire, the WDL production team gained valuable operational experience in processing an entire large collection. Tools developed to lower the cost and speed the pace of image processing, cataloging, translation, and quality assurance now will be applied to other large collections, beginning with photographs from the National Library of Brazil.

Improved Processes and Tools for Content Creation

The Business Plan calls for the development of new tools and processes to facilitate scaling up the WDL in 2012 and beyond, with improvements in transfer, processing, cataloging, and translation. There has been important progress in this area, although work remains to be done and certain functions (high-quality translation of cultural heritage content; cataloging; curatorial description) probably cannot be scaled up solely through process and software improvements.

Efforts were concentrated in three areas: organization and personnel; development of the new metadata application mentioned in the Business Plan; and rewriting the WDL source code to improve performance and the ease with which additions and corrections can be made (WDL 2).

Staff and organization. The WDL team has now reached the level of approximately 23 staff and long-term contractors identified in the Business Plan. The production staff was organized into five teams: Content, Metadata, Descriptions, Translation, and Development. Following an open, international competitive bidding process, new metadata translation contracts with more demanding requirements and specifications were awarded to three firms.

Metadata application. The new WDL metadata application is part of a larger content management system which has been integrated into the production process as part of the WDL 2 effort. Improvements related to creating and maintaining WDL metadata include: controlled vocabulary maintenance in all WDL languages; batch editing to correct or edit multiple records at once; improved methods for applying existing WDL metadata values to new records to ensure optimal integration of new content into the site; and more efficient paths to the translation and publishing phases.

WDL 2. WDL 2 represents a top-to-bottom overhaul of the WDL application to realize improvements in key areas: page load times have been reduced resulting in improved site performance, particularly for users with slow Internet connections; code optimization for improved search engine placement and social sharing; the quality of the images has been improved throughout the site; the public site has migrated from a substantially static architecture to a dynamic content delivery design needed to support future growth; Amazon Web Services have been incorporated to adjust available resources based on traffic and content requirements; internal processes are increasingly being automated to improve efficiency, reliability, and to minimize the amount of time it takes to review, test, and publish content.

Addition of Partners and Capacity Building

In January 2011, the WDL had 117 partners in 64 countries. The medium-term goal set forth in the Business Plan is to have 125 partners from 80 countries by the end of 2011, as a step toward eventual universal or near-universal participation from among UNESCO's 194 member countries. The WDL currently has 136 partners from 72 countries, an increase of 19 partners and 8 countries. A number of applications for membership are pending.

For some partners, membership in the WDL is more nominal than real, as they lack the equipment and trained staff to digitize their collections for inclusion in the WDL and for building their own digital libraries. Other prospective partners have chosen not to join the WDL because they lack the capacity to effectively participate.

Capacity building thus remains a major challenge, not just for the WDL, but for the achievement of the goals set forth in the *IFLA Manifesto for Digital Libraries* and other international declarations. The WDL has underwritten the establishment of digitization centers at the National Library and Archives of Egypt, the Iraqi National Library and Archives, and the National Library of Uganda. It continues to work with these centers to upgrade capacity, and has provided small-scale assistance to a number of other WDL partners. The Central Library, Qatar Foundation and the Library of Congress are co-hosting a meeting and training workshops in Doha, Qatar in February 2012 for WDL partners and prospective partners from the Arabian Peninsula and neighboring African countries, following up on an initial meeting in Doha in December 2010 that resulted in the recruitment of a number of new partners.

In order to scale up capacity building efforts to a level that would meet the needs of libraries and archives in many developing countries, additional contributions, both financial and in-kind, will be required from many partners besides the Library of Congress and the Qatar Foundation. These issues will be discussed in greater detail in the comprehensive needs assessment and capacity building strategy that is called for in the Business Plan and that is now in preparation.

Longer-Term Priorities

Scaling Up the Addition of Content

The Business Plan calls for expanding the WDL to 100,000 items by the end of 2015. The experience of the last year, and particularly the progress in developing new tools and processes for content creation, suggests that this goal is eminently feasible. Meeting it will depend, however, on continued support for a core staff, continued improvements in the inputs received from partners, and sustained support for translation into the seven WDL languages.

Toward Universal Participation

Reaching universal or near-universal participation remains a laudable goal, but one that will depend on developing a credible capacity building strategy and securing the resources for its implementation.

Adding Interface Languages

The Business Plan stresses the desirability of adding at least three languages – Hindi, Japanese, and German – to expand the WDL's theoretical reach to 53 percent of the world's people in their native languages (68 percent in native and non-native languages combined). Technically, there is no bar to adding these (and other) languages. However, addition of these three languages would increase per-item translation costs by 50 percent (not counting retrospective translation of the existing WDL) and thus cannot be considered until long-term funding for the initial seven languages is secure.

Developing Mobile Applications

There has been some progress in this area. The rewriting of the WDL code was partly undertaken to facilitate the development of such applications, and has already resulted in an improved experience for mobile users. Several members of the James Madison Council of the Library of Congress have pledged to underwrite the development of a WDL iPad app. Work on this is expected to begin in 2012.

New Web Site Features

Users generally are satisfied with the look and feel and functionality of the WDL site, so changes so far have been limited, but some enhancements regarded as essential have been made or are in process. Under a contract with the Library of Congress, the Bibliotheca Alexandrina is developing a full-text search for printed books. Faceted browsing has been improved, with the addition of facets for sub-national locations (states and provinces), additional subjects, and language (of content).

In May 2011, the WDL launched a Twitter account, which is being used to generate usage, draw attention to particular items or collections of content, and

disseminate news about the project. now has approximately 3,000 followers and has resulted in increased attention to the WDL project, partners, and content.

Developing Areas of Thematic Focus

The Business Plan mentions development of thematic areas of concentration as a possible future area of focus, e.g., "systems of writing and their origins, geographic exploration, slavery and the slave trade, science and innovation, founding documents, material culture, and regional themes." There has been no direct progress in this area, but the expansion of the volume of content and various process improvements are laying the groundwork for big potential payoffs, for example with regard to Arab science and Chinese history and literature.

<div align="center">

Resources

</div>

The Business Plan included a five-year estimated total cost of $34,736,000 to achieve all of the medium- and long-term goals specified in the plan. This figure may in fact underestimate the full costs of meeting these objectives, as it is based on a flat-lining of translation costs at $2,500,000 per year. Translation costs could in fact go higher, depending on the amount and character of the content added and the number of interface languages provided on the WDL.

The WDL was able to meet its medium-term objectives for the addition of content in 2011, but this was done by deferring all new capacity building initiatives, cutting back on travel and meeting costs, and achieving significant economies in translation. Continued progress in expanding the WDL to its ultimate target size will depend on keeping the WDL production team intact and on sustained support for translation at higher levels of production.

Meeting the longer-term objectives specified in the Business Plan, particularly with regard to capacity building and the addition of languages, will require significant additional financial and in-kind resources. The Library of Congress is undertaking various fund-raising efforts toward this end and is seeking the renewal of several large gifts and grants. The fund-raising environment remains difficult, however, owing to worldwide economic conditions and the existence of numerous other national and regional projects that compete for resources.