

# **The World Digital Library: Project Manager's Report**

## **WDL Executive Council Meeting Alexandria, Egypt November 5, 2015**

This report summarizes the main developments in the World Digital Library since the last meeting of the WDL Executive Council on October 29, 2014.

### **Highlights**

Three major developments relevant to the WDL and the work of the Executive Council occurred in the past year:

- The WDL team released to the public, on November 17, 2014, a new User Interface (UI), the first full redesign of the WDL site since the initial launch in 2009. This resulted in very large increases in usage, as discussed below.
- The Library of Congress provided to the WDL, on a no-cost basis, a dedicated scanner and a full-time staff person for the digitization of Library of Congress content for the WDL. The scanner was installed in the WDL offices in late January 2015.
- Dr. James H. Billington, Librarian of Congress, the initiator of the WDL project, and a member of the WDL Executive Council, retired from the Library of Congress on September 30, 2015. Acting Librarian of Congress David Mao now represents the Library of Congress on the council.

### **Content**

The WDL currently includes 12,465 items, up from 10,778 items at the time of the last EC meeting, and 623,465 images, up from 500,087 images a year ago. These are increases of 16 percent and 25 percent, respectively.

Since the last EC and general partner meetings, content was contributed by 31 partner institutions in 23 countries. The total number of libraries, archives, and museums represented on the WDL with content is 128 institutions in 57 countries.

National libraries newly represented on the WDL are the National Library of Vietnam and the National Library of Albania. National libraries that were already present on the WDL and which contributed new content are the national libraries of Brazil, Canada (Quebec), China, Colombia, Cuba, France, Japan, the Netherlands, Qatar, Russia, Slovakia, Sweden, Uganda, the United Kingdom, and the United States. The municipal library in Porto, Portugal contributed an important manuscript, becoming only the second public library (along with the Intronati Library in Siena, Italy) to be represented on the site.

Content in 128 languages is currently represented on the WDL, up from 116 languages the previous year.

Noteworthy items or collections added in the past year include:

- From the Municipal Library of Porto, Portugal, the only known copy of the journal believed to have been written on board ship during Vasco da Gama's first voyage to India, 1497-99 (a UNESCO Memory of the World item), [www.wdl.org/10068](http://www.wdl.org/10068)
- From the National Library of Catalonia, Barcelona, the oldest surviving document written in Catalan, from the year 1080, [www.wdl.org/14165](http://www.wdl.org/14165)
- From the National Library of Peru, all 39 of the first editions produced between 1584 and 1619 at the press in Lima (the oldest press in South America and the second-oldest in the New World), in Spanish, Latin, Quecha, and Aymara (a UNESCO Memory of the World collection), [www.wdl.org/13760](http://www.wdl.org/13760) through [www.wdl.org/13778](http://www.wdl.org/13778)
- From the National Library of the Netherlands, two of three existing parts of the *Book of Hours of Simon de Varie* (1455) (the third part is in the Getty Museum, California), [www.wdl.org/14414](http://www.wdl.org/14414) and [www.wdl.org/14415](http://www.wdl.org/14415)
- From the National Library of France, all seven of the Mesoamerican codices held by the BnF, including the *Codex Vergara*, *Codex Azcatitlan*, the *Aubin Tonalamatl*, and the *Codex Mexicanus*, [www.wdl.org/15278](http://www.wdl.org/15278) through [www.wdl.org/15284](http://www.wdl.org/15284)
- From the Newberry Library, Chicago, all five of the circa 1540-1574 manuscripts held by the library in the hand of or associated with Fray Bernardino de Sahagún, compiler of the *Florentine Codex*, [www.wdl.org/15015](http://www.wdl.org/15015) through [www.wdl.org/15019](http://www.wdl.org/15019)
- From the Library of Congress, the Library's complete run of the *Cherokee Phoenix* (1828-34), the first Native American newspaper in the United States, [www.wdl.org/15576](http://www.wdl.org/15576) through [www.wdl.org/15567](http://www.wdl.org/15567)
- From the Wellcome Library, 23 Arabic medical manuscripts of a projected contribution of some 400, [www.wdl.org/15263](http://www.wdl.org/15263) through [www.wdl.org/15277](http://www.wdl.org/15277) and [www.wdl.org/15431](http://www.wdl.org/15431) through [www.wdl.org/15440](http://www.wdl.org/15440).

With support from the Qatar National Library, the WDL added 275 items relating to the history of the Arabian Peninsula and the wider Arab world, including books from the nineteenth-century Arab renaissance published in Cairo, Beirut, Istanbul, and other cities. Under the terms of a grant from Carnegie Corporation of New York, the Library of

Congress continued to digitize manuscripts, books, photographs, maps, and other materials relating to the history and culture of Afghanistan and to add these to the WDL. So far approximately 140,000 pages of book, manuscript, and other material have been digitized for this project.

UNESCO Memory of the World collections added to or supplemented in the WDL included the two collections from Peru and Portugal noted above, and additional photographs from the Thereza Christina Maria Collection at the National Library of Brazil.

A substantial quantity of high-quality content was received and is undergoing processing. This includes additional Arabic medical manuscripts from the Wellcome Library in London, made available through a joint Wellcome-Bibliotheca Alexandrina project, and additional tranches of content from the national libraries of Brazil, France, Japan, Qatar, and Slovakia, and from other libraries and cultural institutions.

With the installation of a dedicated scanner at the WDL, the project has stepped up the digitization of mostly nineteenth century books relating to geography, history, exploration, ethnography, infrastructure development, and related topics. Books are being chosen for their quality and rarity, and are presented in color with high-resolution scans (no microfilm and no greyscale), with all foldouts included.

WDL continues to aspire to a much larger volume of content, with the aim, as emphasized in previous reports, of providing broader and more comprehensive coverage of countries, time periods, cultures, and languages. Doing so with limited resources, while maintaining quality standards, remains a major challenge.

The storage and data management challenges of maintaining and continuing to expand what in technical terms is becoming a very large website continue to be met successfully by the WDL team and the Library of Congress infrastructure that supports it. When the last analysis was performed on October 26, 2015, the parameters of the site were determined to be the following:

- Downloadable files: 634,351
- Image tiles for zoom views: 211,397,911 files (1.8 terabytes)
- Master files for image tiles: 1,170,064 files (17.4 terabytes)
- HTML pages: 4,407,662 (87,255 item detail pages and 4,320,407 view pages)
- Pages of searchable text: 194,299 (730,000 word coordinate and plain text files; 2.2 gigabytes)
- Pages indexed by Google: 358,357
- Words of metadata in the seven interface languages: 32.5 million

Much additional content is in the pipeline, some of it already stored on testing and staging servers.

## **Partner Recruitment**

The WDL currently has 190 partners representing 81 countries, an increase from 185 partners in 81 countries at the time of the last Executive Council session.<sup>1</sup>

The Library of Congress recommends that five new partner institutions be approved for membership:

- Ministry of Awqaf and Religious Affairs, Oman. This institution has offered important Islamic manuscripts and will work with other WDL partners in Oman. It has participated in and was recruited through the Arab Peninsula Regional Group convened by the Qatar National Library.
- Middle East Institute, Washington. This institution has an important collection of books, manuscripts, and maps, mostly in Arabic, Ottoman Turkish, and various Western languages, assembled over the years mostly by U.S. diplomats and other Americans engaged in the Middle East.
- American Jewish Joint Distribution Committee, New York. This institution has a vast archive of photographs relating to conditions in Europe and the Middle East during and in the immediate aftermath of World War I. These photographs will complement other World War I materials from the Library of Congress, the British Library, the archive of the League of Nations, and other partners.
- Universidad de Chile, Santiago. This is the major university library in Chile, with important collections relating to the history and culture of Chile.
- Biblioteca del Congreso de Chile. This is the parliamentary library of Chile, with important collections relating to the early constitutional history of Chile, electoral campaigns, and political parties.

The long-term goal of the WDL remains universal participation, i.e., securing at least one partner from each UN member country. However, consistent with the guidance provided by the Executive Council in recent years, the WDL has focused on adding only a small number of partners that actually have the capacity to contribute high-quality content.

## **Technical Development**

The major technical development was the release to the public of the new User Interface (UI), the first full redesign of the WDL site since the initial launch in 2009.

---

<sup>1</sup> Includes partners recommended for membership by the Project Manager, pending formal Executive Council approval.

The new UI is designed to better accommodate access on mobile devices and tablets as well as on PCs and laptops with increasingly diverse screen sizes. New features include:

- larger reference images
- a larger search box set against the background of archival world maps
- new sections for featured items, featured partners, and highlights
- a map display option for search and browse results to accompany the list and gallery views
- revamped browse options for place, time, topic, type of item (format), contributing institution, language (of the content/document)
- an updated item-detail page with many new features, e.g., an inset map showing location of the item and an option to show nearby items
- a more prominently displayed “View this item on the partner’s site” located just below the reference image on the item-detail page, intended to send more traffic back to partner institutions
- upgraded audio and video display
- timelines, interactive maps, and thematic sections
- updated and revised “About” and “Help” sections.

Timelines and interactive maps were created for the following sets of content: “Chinese Books, Manuscripts, Maps, and Prints”; “Illuminated Manuscripts from Europe”; “United States History”; “World History”; and “Imperial Russia” (map only).

One thematic section, “Arabic and Islamic Science and Its Influence on the Western Scientific Tradition,” was launched with the support of the Qatar National Library.

The timelines, interactive maps, and thematic section are being upgraded and enriched with new content, and additional thematic sections and timelines are in development.

## **Usage and User Engagement**

For Library of Congress Fiscal Year 2015 (October 1, 2014-September 30, 2015), the WDL website recorded 6,806,147 visits, up from 3,761,303 visits in FY 2014. Total page views were 30,823,978, up from 24,106,592 views the previous year. These were increases of 81 percent and 28 percent, respectively. If China is excluded from the calculations, growth in visits was 20 percent, in page views 34 percent. Chinese usage appears to be somewhat anomalous, and is being investigated by the WDL team.

The top countries in the world by number of visitors were China, the United States, Brazil, Mexico, Spain, France, the United Kingdom, Germany, Russian Federation, and Canada.

The top countries in the world by pages viewed were the United States, China, Spain, Mexico, Brazil, France, Russian Federation, Germany, United Kingdom, and Italy.

Year-on-year increases for individual countries were, to give a number of examples, China, 1338 percent; Italy, 370 percent; Canada, 156 percent; France, 128 percent; Saudi Arabia, 70 percent; Egypt, 66 percent; United States, 44 percent; Russia, 37 percent; Germany, 31 percent. These trends were offset somewhat by declines in Spain (-19 percent) and other Spanish-speaking countries (although not in Mexico, where traffic was up 13 percent) and in Brazil (-1.5 percent).

Usage by interface language in FY 2015 (in page views) was 38.1 percent Spanish, 23.3 percent English, 11.2 percent Chinese, 9.0 percent Arabic, 7.2 percent Portuguese, 6.0 percent French, and 5.1 percent Russian. Gains in both absolute and relative terms were particularly notable for the Arabic and Chinese interfaces.

The WDL website registered 245,497 ReadSpeaker activations (text-to-voice conversion) in the fiscal year, and 1,546,342 downloads.

In October 2015 WDL had 3.5 million inbound links from other web pages (per Google Webmasters Tools).

The number of followers on Twitter (@WDLorg) increased to 40,600 (as of October 20, 2015), up from just over 31,000 a year earlier. Tweets featuring WDL content were issued on an (almost) daily basis. Tweets were both in English and in the other six WDL languages.

## **Capacity Building**

The WDL continued to process content from digital conversion centers in two developing country libraries established with the financial support of the Library of Congress: the National Library and Archives of Egypt (NLAE) in Cairo and the National Library of Uganda (NLU) in Kampala. Processing continued on the more than 170 manuscripts digitized at the NLAE and on the approximately 46,000 images scanned at the NLU in the period 2009-2013.

No additional content was received from the WDL-funded digitization center at the Iraqi National Library and Archives (INLA) in Baghdad, owing to unsettled conditions in that country.

In cooperation with the Qatar National Library, the WDL organized a symposium, “Arabic Authority Control—A Capacity Building Training,” which was held in Doha on May 13-14, 2015. This was the fifth jointly organized workshop since 2010. Participants included 55 staff members from fifteen partner and prospective partner institutions in eleven countries (Bahrain, Egypt, Iraq, Jordan, Kuwait, Oman, Qatar, Saudi Arabia, Sudan, the United Kingdom, and the United States). Topics covered in the symposium included Arabic authority control, multilingual access to digital resources using linked

data, and the challenges of Arabic name authority in the descriptions and cataloging processes of the WDL.

## **Implementation of Executive Council and Partner Suggestions and Proposals**

The WDL has completed or in some cases made progress on implementing ideas and suggestions put forward by or discussed in previous Executive Council and general partner meetings. The following has been accomplished:

**Translation Experiment.** A controlled experiment has begun with the students in the Master's program in English to French translation at Université Charles de Gaulle (Lille3), France. This has involved working out an arrangement with the instructors, setting up mechanisms to import and export files, and establishing a workflow and quality review mechanisms.

**Partner Dashboard.** A draft version 1.0 of a partner dashboard that allows partners to see the status of their content at all stages of the production process and to view the overall production schedule is complete and ready for review and testing. It can be seen at <http://partners.wdl.org>.

**Collection Statistics.** A collection statistics page is now live and part of the WDL site, at the bottom with other information such as “About,” “Help,” “Contact,” and so forth. This page gives the total number of pages digitized and the file totals in different categories, and is designed in part as an offset to the prominent item count, which tends to understate the size of the site. The URL is <http://www.wdl.org/en/statistics/>. This page is available in the seven interface languages.

**Institutional Pages.** The template for an institutional page that will provide comprehensive information about each institution and its collections is ready. The template needs to be reviewed. Partners will be asked to provide the descriptive information for each of these pages. The translation burden/cost of creating these pages will be considerable, so their added value needs to be considered carefully. For an example of the template, see <http://www.wdl.org/en/institution/library-of-congress/>

**API Page.** To encourage innovation and facilitate cooperation with other projects, an API page has been launched and is part of the WDL site. For the moment this page is only in English, on the assumption that most members of the developer community using the site will know English. Partners could be encouraged to take responsibility for translating this page into the other interface languages if doing so adds value. The link is <http://api.wdl.org/>.

**Review of content and content selection guidelines.** After some delay, the technical platform for running an online forum has been established, using Slack, a commercial off-the-shelf product. See <https://slack.com/is> for further information. Users must be invited to sign up for the forum; it does not accommodate self-selection or nomination.

## **Priorities for the Coming Year**

Priorities for the coming year are:

- Add content to reach a target level of 15,500 items by the next partner meeting
- Sustain and build on the recent increases in usage and user engagement
- Recruit new partners, especially in parts of the world in which the WDL is underrepresented
- Continue to develop timelines, themes, and other interpretive materials
- Work to ensure the financial sustainability and growth of the WDL.