The World Digital Library: Lessons Learned and (Most Likely) Not Learned

Paper presented to the National Digital Library of India-UNESCO International Symposium on Engineering for Digital Library Design (KEDL 2019)

New Delhi, India

9-11 December 2019

John Van Oudenaren[1]

In October 2007, when the Library of Congress (LC) and its partner institutions from Brazil, Egypt, and the Russian Federation met at UNESCO headquarters in Paris to launch the prototype of what would become the World Digital Library, the Library's home town paper, the *Washington Post*, ran an enthusiastic story about the event under the title "Checking Out Tomorrow's Library: In Paris, an International Working Group Shows Off the Prototype for a Multilingual 'Intellectual Cathedral' of Digitized Knowledge."[2] The *Post*'s correspondent, John Ward Anderson, began with a bang: "As ideas go, they don't come much bigger: Digitize the accumulated wisdom of humankind, catalogue it, and offer it for free on the Internet in seven languages." Anderson quoted the then-Librarian of Congress, the late James H. Billington, originator of the WDL, to the effect that the project would "promote all kinds of cross-cultural perspectives and understanding."

Anderson also drew on an interview he had conducted with Silicon Valley technology forecaster Paul Saffo. It was Saffo who used the term "intellectual cathedral" that appeared in the subtitle of the story: "In essence, what they are doing is building an intellectual cathedral, and it may never get finished. But this is a good effort even if it fails, because it is going to inspire a lot of other efforts, and if it succeeds it will be a wonderful resource…. The challenges here aren't technological [or financial]…. For me, the issue is the will to make it happen. The people involved in this—will they really see this through?"

More than twelve years have passed since Saffo made these comments, and they are

---

[1] Global Fellow, Kennan Institute for Advanced Russian Studies, Woodrow Wilson International Center for Scholars, Washington, DC, and former Director, World Digital Library, Library of Congress This paper draws extensively on the author's unpublished history of the WDL, *The World Digital Library at the Library of Congress, 2005-2018,* August 2019.

[2] John Ward Anderson, "Checking Out Tomorrow's Library: In Paris, an International Working Group Shows Off the Prototype for a Multilingual 'Intellectual Cathedral' of Digitized Knowledge," *Washington Post*, October 18, 2007.

worth revisiting for the light they shed on the current state of digital libraries and specifically for what they say about the "Human Aspect in Digital Library Service Design," the topic the conference organizers have asked this author to address. For starters, it is clear that WDL *has* failed, a possibility that Saffo admitted but that did not greatly trouble him at the time. The dimensions of and the reasons for this failure can be discussed (and will be touched on briefly in the treatment below), but the facts are clear: The WDL program office at the Library of Congress was shut down in January 2018, with all staff and contractors either terminated or transferred to other positions in the Library. The last content release (releases previously were done every two weeks) was almost two years ago, on February 7, 2018. LC has stopped accepting new partners and ingesting content from contributing institutions. It has been engaged in an effort to find another institution or group of institutions willing to take over the project, but as yet none has formally done so. The one institution that has expressed interest in assuming the project management role has put forward plans that would take WDL in an entirely different direction from the one that Billington envisioned or that Saffo summed up with his phrase "intellectual cathedral." The WDL website – www.wdl.org – is still online, operating under cloud hosting and CDN contracts concluded before the program office was shuttered, but how long this state of affairs will continue is unclear. The site could go dark at some point. In twelve years at the Library of Congress, the WDL chalked up numerous *partial successes*, but the overall effort has failed.

More interesting than the fate of this one project, however, and of greater relevance for this conference, is Saffo's second point, namely that, even if the effort launched in 2007 failed, it would, as Saffo predicted with great confidence, "inspire a lot of other efforts" Has this in fact happened? Are there other intellectual cathedrals out there, or is the digital library community as a whole creating such a cathedral? The answer to this question unfortunately must be no, and the key question is why.

Background

First proposed in June 2005, the WDL was very much the product of its time. The United States was in the process of rejoining UNESCO (which it has since again left), and UNESCO was eager to work with LC on the knowledge and information society agenda championed by then-Director-General Koichiro Matsuura. The political wounds in the global order caused by the

U.S.-UK invasion of Iraq in March 2003 were slowly beginning to heal, and President Nicolas Sarkozy of France was eager to improve relations with the United States. (The Bibliothèque nationale de France, after a period of initial skepticism under director Jean-Noël Jeanneney, became one of the WDL's major contributors and a strong supporter.) China was emerging, or so it was thought, as a "responsible stakeholder" in the international order, and the National Library of China was another enthusiastic founding partner. In the post-September 11, 2001 environment, the Arab and Islamic countries and the West were eager to work together in the educational and cultural spheres, where it was thought that greater recognition of the positive achievements of Islamic culture – and greater exposure in the Islamic world to the achievements of other cultures – would help to dampen the kind of radicalism that had culminated in the terror attacks on New York and Washington. Not least, there were as yet positive relations between the technology community, Silicon Valley in particular, and the world of libraries and culture, as reflected in the many meetings and discussions between Billington and Google co-founder Sergey Brin, which resulted in the $3,000,000 planning grant to LC from Google in late 2005 that funded the early work on WDL. Additional funding came from Microsoft, with in-kind support from Intel and Apple.

None of these factors apply any longer. The international political environment has deteriorated from what it was in the early 2000s, as has the relationship between Big Tech and the world of culture – indeed between Big Tech and society in general. After 2009, attempts by the Library of Congress to raise additional funds in Silicon Valley were not successful. Tech was going its own way, leaving the nonprofit cultural sector in the dust. It is unlikely that anything like the WDL – with the support of Silicon Valley and the wide participation that it attracted from most of the major countries of the world – could be started today, certainly not under American leadership and most likely not under the leadership of any other group of countries or national libraries either.

Also missing is the kind of leadership that Billington supplied, which was driven by his faith in what culture joined to technology could accomplish and his belief in the importance of using technology to bring heritage content to end users. As Librarian of Congress, Billington saw his mission as primarily that of bringing the riches of the Library to the American people, through exhibitions, publications, educational outreach, and, with the explosion of the worldwide web after 1993, via the Internet. It was in this spirit that he had launched *American Memory* in

the early 1990s. By the late 1990s and early 2000s he had added an international component to the Library's digital outreach. He wanted to use the web to spark creativity worldwide and to bring peoples and cultures together.

The team that was formed at the Library in early 2006 to plan the WDL was left in no doubt that their target audience was not scholars or librarians, but *end users* among the general public, particularly students and youth. As director of the WDL, I had fixed in my mind's eye as our target user a young man or woman sitting in an Internet café in Lahore, Sao Paolo, Cairo, Moscow or the Bronx, who would stumble upon something in the WDL or be led to it by Google and would find it of interest and be enlightened by it in some way. If we could reach these kinds of people, the rest would be "lesser includeds" – more affluent and better connected users in the United States and Western Europe with their PCs, home and office Internet access, and soon their smart phones and iPads. This focus on reaching end users in the less digitally accessible parts of the world led us to explore early on a number of partnerships, e.g., with the MIT Media Lab's One Laptop Per Child project, with Intel's Classmate, and with the new smartphones that were just emerging in 2007.

The focus on reaching end users also drove the basic decisions that the WDL team made about the project in those early days. Three elements were identified as crucial to meeting our mission: (1) multilingualism; (2) universality, and (3) quality and performance.

Multilingualism was of course essential to reaching a global audience. We chose to offer the interface in seven languages – the six official UN languages plus Portuguese – and hoped to add additional languages in the future, beginning with Hindi and other South and Southeast Asian languages, speakers of which constituted the largest share of the world's population who would be unable to access the site if it remained limited to the initial seven languages. In terms of content, we sought to represent both major languages spoken by tens or hundreds of millions people and languages spoken by smaller numbers of people, including endangered languages – the latter in support of UNESCO's efforts on behalf of these languages.

Architecturally, handling multilingualism was straightforward. The WDL was effectively designed as seven separate websites – one in each of the interface languages – from which it was possible to go back and forth at any level or from any feature of the site. The language of the interface that the user would see was keyed to his or her browser. Thus in Brazil WDL (www.wdl.org/pt) would appear as a Portuguese website, while in Egypt it would show in Arabic

(www.wdl.org/ar), in China in Chinese (www.wdl.org/zh), and so forth. Our aim was to create an "equivalent user experience" in each of the seven languages. The effect of this approach, which was intentional, was to set a very high standard of linguistic excellence for all language interfaces. We were not presenting ourselves as an American, English-language website that was offering other languages sort of as a "bonus" for which the user should be grateful and therefore excuse certain imperfections. Rather, we were seeking to "sell" Spanish in Spain, Arabic in Egypt, Russian in Russia, and so on.

The element of universality related both to content and participation. We wanted to have content *about* every country in the world in the WDL, and ideally we hoped that the best of this content would be contributed by national libraries or other repositories in these countries themselves. Fostering the widest possible participation in the WDL was closely linked to capacity building. There were many countries in which institutions were eager to join the WDL, but these institutions lacked the equipment, trained staff, and software needed to digitize their collections. The Library of Congress had already provided digitization equipment to libraries in Russia and Brazil in support of bilateral projects with these countries, and it made capacity building a central part of its planning for the WDL. This was not, or at least not primarily, an act of altruism. Rather, it was based on a recognition that only by providing assistance to others could the WDL hope to approach anything like universal participation – to gain access to content, much of it at risk, which needed to be represented in order to create a true "world" library.

The WDL established digital conversion centers at several national libraries and it engaged in a number of other capacity building efforts. Overall, however, the capacity building activities of the WDL fell far short of what was envisioned at the start of the project. LC/WDL alone could not carry the burden of capacity building for the whole world, and discussions with other institutions and organizations interested in promoting the digitization of cultural heritage materials in countries where such materials were not being digitized (IFLA, UNESCO, Francophonie, the British Library's Endangered Archives program) did not result in any major international effort. Partly as a consequence of this, the WDL's list of partners – 158 libraries, museums, and archives in 60 countries – while impressive, still omits more than two-thirds of the world's 193 states and leaves large regions of the globe represented sparsely or not at all.

The third element seen as crucial to meeting the WDL's objectives was quality. By

quality we meant a whole range of features that spanned the realms of technology, metadata, and curation. We wanted a carefully selected, high-quality body of content. Working against a set of selection criteria developed by an international committee, we sought cultural content that, to use Billington's phrase, was both "interesting and important."[3] We insisted upon a complete set of metadata for each item in the site, one that could drive the cross-cultural and intertemporal browse and display options that were intended to be the heart of the WDL experience, and that would ensure that search results would be ranked according to the relevance of the item to the search query, and not as an artifact of the quality or completeness of the metadata. The site itself had to be intuitively easy to use, with a variety of search, browse and display options, making use of maps and geographic coordinates to emphasize that this was a *world* digital library.

We insisted upon powerful zoom capabilities that would display our unique and often physically beautiful content to best effect. With rare exceptions, we ensured that all content was made downloadable, for free of course. We assigned a unique and standardized URL to every image file of every item accessible on the WDL, for each of the seven interfaces. This made possible inbound links from millions of websites around the world, including Wikipedia pages (in multiple languages), blogs, and the syllabi for university and school courses – all of which boosted usage and search engine rankings. We introduced full text search in seven languages.

Curation and interpretation were important. For each unique item on the site we required an item-level description, written by an expert curator or scholar, but in clear, jargon-free language that aimed to explain the most important cultural content to a proverbial general audience, including school audiences. Description writers were asked to address two basic questions: "what is this item and why does it matter?"

Not least, we focused on speed and performance. Eager to attract that impatient young user, we could not afford a site that featured slow or broken links. Our goal was speed, speed, speed – of access, everywhere in the world, which we measured through specially commissioned studies that examined response times at different cities through different ISPs.[4]

---

[3] Many things on the web are interesting, even entertaining – think YouTube videos – but they are not important. Other things, Supreme Court cases, for example, are important, but not terribly interesting, at least for the general user. Our job was to find and describe items that were both.

[4] For those technically inclined, it is worth noting the granularity of this data. A worldwide analysis conducted for the Library of Congress by Keynote Consulting determined

All this, needless to say, was incredibly demanding, and required the work of a skilled and dedicated team. In many ways WDL was a vastly overambitious project – a digital moonshot – that was all but destined to come down to earth eventually. But it was an adventure while it lasted. And it was all driven by that attempt to reach the end user.

Architecture and Its Discontents

These three elements – multilingualism, universality, and quality – drove the choice of system architecture. Much of what we wanted to do simply could not have been accomplished, given the technology of the day, with the kind of "decentralized," distributed architecture that then was coming into vogue in the international library community (a simple case in point: the assignment of a unique and standardized URL to each image file of every item accessible on the site could not have been accomplished with a distributed architecture.)

Looking back from the perspective of nearly fifteen years, it is clear that our decision to adopt a "centralized" architecture rather than to build a portal explained both the great successes of the WDL (particularly in attracting end users), but also the resistance the project encountered in the library community. For reasons that are not entirely clear, by the early 2000s the library community in the United States and Western Europe had become interested almost exclusively in metadata aggregation, federated search across platforms, and the development of portals to access distributed (and, as will be emphasized below, heterogeneous) bodies of content. This occurred, ironically, even as the commercial world was moving in the opposite direction – beginning to amass vast amounts of content that each firm "owned" and stored on vast server farms, which it then sought to "monetize" through centralized and highly homogenous display applications, e.g., Flickr, YouTube, iTunes, etc. Many in the library community could never get past the WDL's decision to adopt a "centralized" architecture. Especially in the Anglo-Saxon world, it was the "original sin" that helped to doom the project.

Ironically, those of us who made the basic decisions about the WDL in 2006-2007 did not *think* we were creating a "centralized" architecture. As outlined in the paper presented to a

---

that during the period October 16-October 25, 2009, at 64 separate locations around the world, the download time of all WDL pages was under 2 seconds (in many locations they were 1 second or less), with just three exceptions, all in China (Shanghai China Telecom, 3.19 seconds, Beijing Netcom, 3.23 seconds, Shenyang Netcom, 7.49 seconds). These metrics of course improved further over time.

UNESCO experts group in December 2006, the "WDL system architecture should be distributed in nature, with each participating site contributing to the creation or maintenance of content and/or to ensuring the accessibility and preservation of content, depending upon the site's strengths and capacities."[5] On the access/content delivery side, we envisioned a network of mirror sites around the world, starting with six or seven sites in the major geographic regions. The mirrors were seen as politically important – a way to gain buy-in and participation – but also as a means to improve technical performance, i.e., response times and continuity of access. We took the mirror concept seriously, and in fact had a functioning mirror site in place (at the Bibliotheca Alexandrina in Alexandria, Egypt) when we launched the prototype in October 2007.

On the content creation/processing side, we also envisioned a network of geographically distributed sites that would create and process content, with the aim of decreasing heterogeneity/increasing homogeneity as content moved from digitization at the local site where the physical artifact was held to storage on the central repository and then propagation to the network of mirrors:

> An important issue in the architecture is the management of heterogeneity as content is created, transferred and ingested into the repository. By the time content and associated metadata are ingested into the repository and are made accessible in an efficient and useful manner on mirror sites, they will go through several stages and transformations within sites and between sites. As it is first created in digital form, content may be in several different file and aggregation formats with different naming conventions and qualities from one digitization site to the next. The content sets grow in complexity and volume as they are translated into several other languages, and may be transformed again into other formats at this stage of their lifecycle. Metadata is created and collected from the very beginning and is added to from one stage of the lifecycle to the next. Metadata schemas and their formats and structures thus vary from one stage to the next, generally increasing in complexity. As transfer of content from one stage to the next occurs, transformation into different forms and formats will occur as content is wrapped and packaged to facilitate transfer….
>
> Within the repository (i.e., the site that is replicated on mirror sites), the content formats and structures are normalized to provide uniformity within and between collections, as much as possible. This characteristic will enable effective search within and between diverse collections in ways that heterogeneous information sources could not afford.

As the WDL developed after 2007, two things happened to change our approach to system architecture, both of which heightened what was perceived as the "centralized" nature of

[5] "The Library of Congress Initiative for a World Digital Library," discussion paper for the UNESCO Experts Meeting, Paris, December 1, 2006.

the site. On the content delivery side, the advent of CDNs (Content Delivery Networks) and the development of the cloud rendered the mirror concept obsolete. Worried about the kind of crash that had crippled Europeana at its launch in October 2008, for the public launch of the WDL on April 21, 2009 we contracted with a commercial CDN to handle what was expected to be a large volume of traffic, given the publicity that would attend the launch. As it happened, on launch day the CDN flawlessly handled some 600,000 visits and 7.1 million page views, most concentrated in the afternoon and early evening of that day. It became difficult to argue that there was value added, at least with regard to performance, in the envisioned six or seven mirrors when Akamai (the firm that we used) by that time already had 45,000 servers at locations around the world. We thus continued to use the CDN, as well as later contracted with a commercial cloud provider, and dropped the mirror concept.

Things were more complicated on the content creation and processing side, where we faced major challenges with image quality, file naming conventions, content transfer, and especially metadata and descriptions. The *best* metadata, in terms of technical quality, we received from partners tended to be Dublin Core or MARC XML files, generally from large libraries in developed countries. The *worst* piece of metadata we ever received, in terms of technical quality, was a JPEG file, sent via email, of a photo taken by a mobile phone of an index card handwritten in Arabic, the only existing catalog record of an Arabic manuscript held at a library in the Middle East. These were the extremes, with everything in between received as well. We asked for and welcomed the good and the best, but we generally accepted the "bad" or "worst" if it advanced our goal of broadening participation to partner institutions that lacked the capacity or the will to create metadata up to the level specified in our project standards.

Although we never abandoned the idea of a distributed content creation network in which partners would upload their content, normalize it to technical and cataloging standards using online tools, authority files and so forth, as a practical matter, particularly in the early days when we were desperate to get a respectable body of content from as wide and diverse a range of partners as possible, it was simply easier to have the content transferred in raw form to the project office in Washington and from there employ staff (catalogers, digital conversion specialists, translators, and subject matter experts) to ensure that the metadata and descriptions were of a uniformly high quality. It was this reliance on processing at a central location that

drove up the cost of the project and held down the number of items we ultimately were able to publish.

The "centralized" architecture of the WDL is thus something of a misnomer, or at least based on a misconception. Centralization was not an end in itself but a means to an end – that end being high-quality unified search, homogeneity of display, and high performance. If we could have kept the content distributed but enforced a uniformity of item display, with an identical and always complete set of metadata informed by the same authority files, an item-level description created to the same high level of quality, translations of uniformly high quality into the seven languages, display via the same page turner and item viewer, all powered by a search algorithm that produced results ranked by relevance and importance rather than as an artifact of quality of the item-level metadata (as is the case with most portal projects) we would have been happy to use a distributed architecture.

Conversely, it should be stressed that those in the library community who insist that the only acceptable approach to building cooperative digital libraries is to embrace decentralization and distributed systems are not really arguing for decentralization as such. They are in fact arguing for heterogeneity in how content is curated, displayed, and (beyond certain minimum levels of standardization needed to power federated search), cataloged.

Librarians and End Users

One can only speculate on the reasons for the fascination that vast, partially unorganized and largely uncurated collections of skeletal metadata with external links to objects have had for the library profession, especially in the Anglo-Saxon world, these last two decades and more. No doubt this fascination has something to do with the sociology of the profession and the incentive structures that apply within it. Data is the realm of the faculty of arts and sciences; metadata that of the schools of library and information science.

Librarians who concern themselves with curation – context and content – are all but destined to play second fiddle to the historians, social scientists, art historians, and scholars in other disciplines who can devote their lives to the study of content. Librarians who work primarily, indeed exclusively, with metadata, in contrast, are likely to go far. This dichotomy between "data" and "data about data" probably has existed since the earliest libraries were established and people were employed to staff them, but it is one that has become especially

pronounced in the digital age. In this new era, librarians have greatly expanded opportunities to link, network, and manipulate metadata via the worldwide web. With these opportunities have come possibilities for collaborative projects, fundraising possibilities, advancement of the status of the library profession (especially in the university setting), chances to travel and attend conferences, higher salaries, and possibilities for personal advancement.

Interest in these vast metadata aggregation projects arguably also has been fostered by and in turn helped to further reinforce the deep-seated strain of anti-intellectualism that exists in the library profession. Anti-intellectualism is perhaps not the best word – content agnosticism may be a better formulation. When Dr. Billington and I used to visit Silicon Valley for fundraising trips in the early 2000s, we would marvel at the content agnosticism of our corporate contacts. A complete digitized version of *The Egyptian Book of the Dead* was worth no more than Mrs. Jones's fourth grade class's PowerPoint on the *Book of the Dead* – both were pieces of content. The important thing was to slap a piece of metadata on them, get them online, and allow them to be monetized in different ways. But these were businessmen, after all, trying to make money, and this content agnosticism was understandable. We librarians, Billington would argue, were different. We cared about content. We employed hundreds of curators who know what was in and cared deeply about the collections – both their physical/digital form and the substance of what they contained – and who were in constant touch with scholars who delved even more deeply into the meaning of this content. Silicon Valley, Billington would argue, should cooperate with (i.e., provide funding to) the digital library projects of the Library of Congress and other cultural institutions, because we were the ones who knew about content and how it was accessed and used by scholars and the public – the constituents that we dealt with on a daily basis.

It is hard to argue that this is any longer the case – that libraries are better at serving their patrons than the commercial firms are their users/customers. Indeed, how much can librarians, some of whom are now talking about aggregating as many as a *billion* records in a mega-metadata site, really care about content and how it is accessed and used by the public? If anything there has been a reversal of positions over the last two decades between the commercial world and professional librarians, with the former increasingly "curating" its content in a way that aims to serve users, while the latter continue to pile up incoherent jumbles of metadata in

semi-random ways.[6] As the commercial world continues to pull away from libraries in the government and non-profit sector, the latter are likely to be progressively marginalized. This will be increasingly the case as search and other access tools are improved using AI, and as the younger generation of users increasingly accesses content, including educational and cultural heritage content, through voice-activated devices such as Alexa, Siri, and Google Assistant, or even through traditional online search. (As the press reported recently, Google search is currently undergoing a transformation from providing lists of links "to determining the right answer without having to even click through to another site. The challenge will increase as queries increasingly move from text to voice-controlled technology."[7] The implication of this, of course, is that fewer users will even be sent to heritage and educational websites, as Google incorporates more content into its own applications – a process that is already well underway.)

Unfortunately there is as yet no reason to think, especially given the unhappy fate of the WDL, that the American or international library community will rethink any time soon its commitment to projects that feature heterogeneous, distributed, and largely uncurated masses of content, i.e., metadata. Sunk investments (financial and psychological) and the sociological features of the profession just noted in all probability will preclude this. And this despite the fact that these projects do not appear to have great appeal to end users and are not a cost effective

[6] This point can be illustrated, for example, by doing a search on a common term, e.g., "William Shakespeare," on Google and then doing the same on any one of the large portal projects or even an internal search on a large institutional website, e.g., www.loc.gov. Google turns up a sensible, high-quality, and highly useful set of links that is likely to appeal to the user. The top ten search results returned to a U.S. user in October 2019 are: articles on Wikipedia, britannica.com, and poets.org, a link to an online version of the complete works of Shakespeare, links to www.williamshakespeare.net, to the Shakespeare birthplace in England, to the Folger Shakespeare Library in Washington, DC, to the Shakespeare section of the Poetry Foundation's website, and a link to the Royal Shakespeare Library. Searches on the library sites, in contrast, turn up such random and unimportant gems as a 19th century engraving of Shakespeare in his room with metadata in German; a 1939 photo of an actor playing *Hamlet* at a production in Paris; a 1971 poster in Polish of a Warsaw production of *Julius Caesar*; a book about Shakespeare in German, translated from the Danish; a cigarette pack with an illustration of Shakespeare writing, etc. In contrast, an internal search on WDL by design goes straight to what is arguably the most important Shakespeare artefact in existence: one of the approximately 240 copies of the *First Folio* of 1623 still extant. On the other hand, of course, the bane of WDL is its small amount of content: only 35 results are returned, compared, for example, with 12,306 on Europeana.

[7] Greg Bensinger, "Google adjusts its algorithm to improve results on sentence-based searches," *Washington Post*, October 26, 2019.

way to bring educational and cultural heritage content to large national audiences, much less to a global audience.

The question of appeal can be best judged from statistics on usage. Unfortunately, most digital libraries tend to be fairly secretive about their usage statistics. Some indication of usage can be gleaned, however, from web analytics firms such as Alexa Internet, Inc., an Amazon subsidiary, or the Israeli company SimilarWeb. The Digital Public Library of America (dp.la), a project launched some years ago with great fanfare and hailed at the time as a non-profit alternative to Google, apparently registers in the range of about 5,000-6,000 users per day, of which about half are "bounces." This means that approximately 3,000 people per day actually look at content on the site via search or browse. As of this writing, DPLA provides access to 35,462,436 items (images, texts, videos, etc.), which means that if each visitor to the site were to look at one item per visit, it would take 11,820 days (32.4 years) for each item in the site to be accessed even once. The fact is that much of the content on these vast metadata aggregation projects is never accessed. These projects are built by librarians for librarians, not for end users.

WDL does a good deal better at actually having its content accessed, although, in fairness, the body of content it contains is vastly smaller. It averages about 20,000-22,000 visitors per day worldwide (even in its current orphaned state), of which about half, as with DPLA, are bounces. This means that in principle, applying the same hypothetical, heuristic framework used for DPLA, every item in the relatively small body of WDL content (19,147 items) can be accessed every two days. We in fact know that most WDL content *is* accessed, with the most popular item being the *Florentine Codex* (created in Mexico in 1577, and held in the Laurentian Library in Florence, Italy), which during a two-year period in 2015-2017 was viewed about 200 times per day (180,000 times over a 2.5 year period), with each visit averaging about nine page views. This, needless to say, is one of WDL's best cases, and not too much should be made of it. But it does bear emphasizing that every aspect of the WDL's presentation of the *Florentine Codex* was undertaken with the end user in mind: (1) selection of the document, which happens to be the most important primary source for the history of pre-Columbian Mexico; (2) digitization, which was funded by WDL and supported by the generous efforts of the staff and leadership of the Laurentian Library; (3) cataloging and description, which was done to the highest level of quality using scholarly work by leading Italian and Mexican scholars and then presented in understandable everyday language; (4) translation of the

metadata in the seven WDL languages, with Spanish of course being key; and (5) strict attention to SEO techniques, so that the presentation ranks high on search engine pages, as well as through the cultivation of numerous inbound links from other sites.[8]

While DPLA (and by extension other portal projects) and WDL both have their strengths, neither really can be deemed as successful – as something that the library community should be proud of, or that can be seen as placing the library community in a good position to meet the educational challenges of the future or to seize upon the opportunities offered by the emerging worlds of AI and 5G. Compared with the universe of users out there, usage for all educational and cultural libraries is abysmally low, not remotely approaching that of Wikipedia, for example, not to mention Facebook, YouTube, Pinterest, and all the other commercial competitors that take up much of the oxygen in the digital space.

DPLA, and even more so Europeana (which provides access to nearly 58 million items) offer vast bodies of content, but it is not clear that they provide enough intellectual and curatorial value-added to attract large numbers of users. WDL offers such value, but does so for a very small body of content, one created at such high cost that it was able to put online only a limited number of items, not truly representative of the "world" indicated in the project's title. The project was in any case abandoned by its sponsoring institution as financially unsustainable.

There is a case to be made that the first organization or institution or group of institutions that can combine the strengths of DPLA/Europeana and WDL – i.e., a large body of content combined with a high degree of curation and intellectual added value – would "own" the digital library space for cultural heritage content. This would mean bringing together a large, high-quality, and representative body of content – not necessarily 37 million or 58 million items but, e.g., one or two million items – with something like the level of presentation, curation, and interpretation offered and aspired to by WDL. *This* would be the "intellectual cathedral" to which Saffo referred, the creation of which, he predicted, would be kicked off by WDL's launch.

[8] A Google search of October 28, 2019 ranks the Wikipedia article "The Florentine Codex," number one; number two is the WDL presentation of the codex. www.wdl.org/10096. The Wikipedia article also references and links to the WDL presentation: "The 2012 World Digital Library high-resolution digital version of the manuscript makes it fully accessible online to all those interested in this source for Mexican and Aztec history." Such links of course drive traffic to the WDL.

Given current realities, however, one should not hold one's breath in waiting for such a project to emerge; none seems visible on the horizon. A more likely outcome will be that libraries and other cultural institutions will continue down the path of increasing marginalization that they have been on since the turn of the millennium, as end users increasingly access information, including cultural heritage and educational information, from for-profit sites that offer better performance, more intellectual added value, and are better oriented to the user. This may not be the best outcome for society, and it certainly is not one to be welcomed by those who are committed to libraries (and museums and archives) and their traditional missions, but it does seem to be the one that is most likely at this time.

Lessons and Implications

The National Digital Library of India is an ambitious venture. Whether it aspires to create (or encompass within it) an "intellectual cathedral" – or, to use language more appropriate for India, an "intellectual temple" – is unclear. With its very long history, its many languages and cultures, and its current and historical links to virtually all parts of the world, India is a virtual microcosm of the world that could, if it wanted to, create this kind of resource. Whether it has the interest and the will to do so of course remains to be seen.

Should it desire to move in this direction, however, a few lessons would seem to follow from the WDL experience:

- Recognize that architecture is a red herring. Eventually technology will make everything – or almost everything – possible. The real issue is the management of heterogeneity. Digital libraries are tools that, among other things, mediate between the heterogeneity in which cultural institutions individually provide access to their content and the homogeneity of display, search, and re-use that end users have come to expect from major web projects. The absolute homogeneity that WDL sought to enforce and that was appropriate for its purposes may not be necessary for other projects, but the complete lack of uniformity in how content is cataloged and presented in many cooperative national and international portal projects is also problematic. A balance needs to be struck.
- Open up the process of building digital libraries to new constituencies. Much the way, as it is often said, war is too serious to be left to the generals, digital libraries

are too important to be left to librarians. They have an important supporting role to play, but their professional self-interest and the sociology of the profession too heavily influence the approach they have taken to building digital libraries. A wider and more diverse set of stakeholders needs to be brought into the process, including historians and other scholars, educators, political and religious leaders, and civil society.

- Strike a sensible balance regarding the amount of content the digital library and its various subcomponents provide access to, one appropriate to its purposes. Insufficient quantity in itself becomes a failure of quality, so the amount of content/number of items should be large – in cultural heritage projects enough to represent all time periods, formats, cultures, national languages, religions, geographic regions, ethnic groups and so forth without leaving major gaps. But do not chase numbers for the sake of chasing numbers, aggregating tens of millions of items that duplicate each other and confuse users. Be especially cautious about joining vast international metadata projects that aim to aggregate tens, possibly even hundreds of millions of bibliographic records. While such projects represent a recurring dream of the library profession and provide grist for many an international meeting, there is no evidence that either scholars or the general public see the value of such projects or will ever use them in a way that will repay the investment they will require.
- Pay attention to the end user. This involves both quantitative and qualitative dimensions. Measure and be transparent about how many users are visiting the site, for how long, viewing/downloading/printing, etc. what, and supplement this data with information about how they are using the site, for what purposes, and so forth. Use web metrics and direct techniques such as surveys and focus groups. Only by paying attention to the end user can the digital offerings of libraries, archives, museums, and the educational sector hope to carve out a space for themselves in a digital universe increasingly dominated by commercial firms.