

## Information + access = digitization?

**Myra Michele Brown**

*Information Resource Officer (librarian)*

*U.S. Embassy, Rome*

*Regional officer for: Italy, Greece, Cyprus, Portugal, Spain, Malta and FYROM*

[BrownMM4@state.gov](mailto:BrownMM4@state.gov)

**Co-author: Fabio Iacoella**

*Audience researcher & electronic outreach*

*Information Resource Center*

*U.S. Embassy, Rome*

[iacoellaF@state.gov](mailto:iacoellaF@state.gov)

**"I have always imagined that Paradise will be a kind of library."**

**Jorge Luis Borges - writer, poet and critic (Argentina)**

**"Google's mission is to organize the world's information, and we're excited to be working with libraries to help make this mission a reality." Larry Page, co-founder of Google, 2004**

Information is power ... "information is liberating"<sup>1</sup> ... access to information ... "the Internet is becoming the town square for the global village of tomorrow;"<sup>2</sup> why are these phrases important enough to be the launching point for a serious discussion of digitization? The goal of this paper is to discuss digitization projects objectively. What are the advantages and disadvantages of starting a digitization project? What is the best way to begin digitizing? Finally, a look at some digitization projects on both sides of the Atlantic.

The fundamental definition of librarianship in the U.S. is "to organize information for access." By that definition, information is the business product of librarians and information professionals globally. If information is power, but you are poor or uneducated or do not have a home telephone (which means you cannot have a home computer) where will you go to have access to this powerful information? Suppose the information you want to in an old, rare book about ancient Greece that is in the library of a university in a country far from your home? How will the librarian at your local library organize that information for you so you can access it?

Access is the most compelling argument in favor of digitization. Library collections worldwide are by their nature available to a limited audience. In the U.S., the public library is a building that's very near home where many people I know go to get the latest novel, to read magazines and newspapers and to check out free DVDs. Academic or university libraries serve their communities. Just providing access to Texas Tech University's distance students required special VPN authentication.

How can digitization improve access? An example: I lived in Dona Ana county New Mexico for more than six years. If the collection of Branigan Public Library in Las Cruces, New Mexico was digital and available online, I could sit in my apartment in Rome, Italy reading books about that region.

Digitization for access is not just a geographical issue. Many European countries have ancient materials that are unique, delicate and should be protected for future generations of users. In

May, I was one of three people from the U.S. embassy Rome who were given a tour of the Biblioteca della Camera dei deputati and the Biblioteca del Senato della Repubblica, near the Pantheon in Roma Centro. We were shown a magnificent book from the 13<sup>th</sup> or 14<sup>th</sup> century which predates the printing press. It was handwritten over many decades by generations of people. Items like that book motivate people from all over the world who are interested in history and in literature to travel to Europe to visit your libraries and your museums. However, what about those who are clever and interested but unable to travel here? And what is the best way to make this magnificent book available? It's at least 600 years old. It is irreplaceable, delicate and must be protected so it is not destroyed by excessive handling.

If this book were part of a digitization project at the biblioteche, a potentially enormous group of people worldwide including scholars, historians, students and anyone with an interest in rare books could see this book in digital form. This to me is the defining pro argument for digitizing books, sound recordings, artworks and other items – it makes scholarship more democratic. If  $\frac{1}{4}$  of  $\frac{1}{4}$  of  $\frac{1}{4}$  of one-quarter percent of the world population of 6,602,224,175<sup>3</sup> wanted to view this book, that is almost 258,000 people. It is not unreasonable to postulate there are more than one quarter million people on Earth who would want to view the digital image of such a magnificent, historically significant book. I would never have known existed had I not been fortunate enough to get a job in Rome and to have a colleague who arranged for us to tour those two libraries. Access to information ought not be left to chance; if information is powerful, digitization is a powerful tool that frees music and art and books to be enjoyed not just by those with money, education or connections but to the curious daydreamers amongst us.

There are valid, coherent, legitimate arguments against digitization. Two major ones are cost and labor. These are considerable hurdles.

*"The European Commission has contributed about €60 million, or \$85 million, to develop a digital library system that can be shared by a wide number of national libraries and cultural institutions. But it is not financing basic digitization, which the Commission estimated would cost €250 million over four years. Some major libraries are still pressing for more public financing, but European officials are clearly encouraging private alliances."*<sup>4</sup>

Digitization requires expensive copy machines and staff time. Some early digitization projects actually shipped books to countries with large pools of inexpensive labor. The cost of one premium Kirtas™ robotic book scanner exceeds USD \$150,000. And even the most sophisticated machine still requires human supervision and human intervention. In addition to the cost(s) of one or more book scanners, a digitization project will require sufficient personnel.

Google's advantage over almost every other digital project in the world is the company's extraordinary financial resources. Google founders Larry Page and Sergey Brin have committed Google's resources to their digitization project. The 30 September, 2007 value of Google was in excess of USD \$100 billion.

*"The scanning process involves Google installing proprietary, high-speed scanning stations on-site at the different libraries (with the exception of Stanford, which will send its books to nearby Google headquarters). The proprietary scanning equipment, using tent scanners, has been in development for over a year, according to Smith. Google staff will scan the books and forward image copies to a central facility for quality control checking and OCR (optical character recognition) conversion to text. Harvard's FAQ for the project stated that Google's scanning process was "much gentler with books than other high-speed processes in use today."*<sup>5</sup>

And then there is copyright. According to some observers, that might be the biggest issue proponents of digitization face. Although the irony of the 2005 U.S. federal lawsuit filed by the

Author's Guild against Google and its Google Print Project is the indexing of books actually increases sales:

*"The Google effort to define copying for indexing as a fair use may end in settlement (rather than case law) because the evidence is mounting that indexing books is boosting sales, which of course, affects the fourth fair use factor. More importantly, it effectively takes the wind out of copyright owners' sails. If they continue on, they will be pressing a point of principle, but are in fact practical economically motivated actors in most cases. This suggests that a settlement is more likely than a jury verdict, appealed all the way to the Supreme Court. There are, no doubt, many exceptions to the pragmatic picture of publishers I'm supposing, but I doubt those are the publishers in charge of the litigation."*<sup>6</sup>

Despite the costs, confusion, legal actions and labor issues, there are digitization projects large and small on both sides of the Atlantic. During a tour of the National Library of Naples (Italy)/la Biblioteca Nazionale di Napoli, the library's director, Mauro Giancaspro, told us about his library's digitization partnership with Provo University in Provo, Utah (U.S.) The National Library of Naples' papyrus rolls, which date from 3<sup>rd</sup> century b.c. to 79 a.d., are being digitized using multispectral imaging at Provo University.

The National University Library in Skopje, FYROM is developing a plan to digitize their special collections. Texas Tech University Library, in Lubbock, Texas, launched a project to begin digitizing portions of their collections in 2005. The Library of Congress (U.S.) has initiated a massive digitization project to preserve its more than three million recordings in formats ranging from wax cylinders to disks to CDs. ([http://www.gcn.com/print/25\\_10/40581-1.html](http://www.gcn.com/print/25_10/40581-1.html)). An academic at Beijing University published a paper in 2005 analyzing a decade of library digital activities in Europe (<http://research.dils.tku.edu.tw/joemls/42/42-4/455-469.pdf>).

The author is a huge fan of digitization. So is Michael Keller, head librarian at the prestigious Stanford University in Palo Alto, California:

*Putting the world's most advanced scholarly and scientific knowledge on the Internet has been a long-held ambition for Michael Keller, head librarian at Stanford University. But achieving this goal means digitizing the texts of millions of books, journals and magazines -- a slow process that involves turning each page, flattening it and scanning the words into a computer database. ... "Think about the power of bringing our library to little schools in the middle of Africa," Mr. Keller said. "Would it make a difference for those who now have their minds closed to the idea of democracy?"*<sup>7</sup>

## Footnotes

1 Kofi Annan, 7<sup>th</sup> Secretary General of the United Nations  
[http://thinkexist.com/quotation/knowledge\\_is\\_power-information\\_is\\_liberating/151771.html](http://thinkexist.com/quotation/knowledge_is_power-information_is_liberating/151771.html)

2 Bill Gates, founder and chairman of Microsoft and philanthropist  
[http://thinkexist.com/quotes/bill\\_gates/3.html](http://thinkexist.com/quotes/bill_gates/3.html)

3 Wikipedia, online encyclopedia [http://en.wikipedia.org/wiki/World\\_population](http://en.wikipedia.org/wiki/World_population)

4 International Herald Tribune, 28/10/07  
<http://www.ihf.com/articles/2007/10/28/business/LIBRARY29.php>

5 Press release for Google digitization project with five library systems in the U.S. and the U.K.  
<http://newsbreaks.infotoday.com/nbreader.asp?ArticleID=16307>

6 Harper, Georgia. "The effect of mass digitization projects on copyright law and policy"  
<http://blogs.tdl.org/digitize/2007/10/11/effect-of-mass-digitization>

7 Markoff, John. "The Evelyn Wood of Digitized Book Scanners"  
<http://query.nytimes.com/gst/fullpage.html?res=9A02E4D6123FF931A25756C0A9659C8B63&sec=&spon=&pagewanted=1>

### **Further reading**

1. Jones, Trevor (2001), "An Introduction to Digital Projects for Libraries, Museums and Archives", available in <http://images.library.uiuc.edu/resources/introduction.htm>, accessed on 17-12-2007.

2. European Navigator, published by le "Centre Virtuel de la Connaissance sur l'Europe", available in <http://www.ena.lu>, accessed on 17-12-2007 (a great example of knowledge digitization)