

## DIGITIZATION AND PRESERVATION OF HISTORIC US GOVERNMENT PUBLICATIONS

Libraries, federal agencies, and private entities have put great effort, time and expense toward the digitization of historic public domain federal government publications within the Federal Depository Library Program (FDLP). There are both large and small efforts toward this end (e.g., the Google Books Project,<sup>1</sup> Committee on Institutional Cooperation (CIC) Google government documents project,<sup>2</sup> GPO/Library of Congress *Congressional Record* project<sup>3</sup> and the 100+ completed and ongoing digitization projects listed on the FDLP Digitization Projects Registry<sup>4</sup>). The FDLP community must now take collaborative, concerted steps to assure that these disconnected efforts lead to a national plan for the digitization and preservation of historic US government publications.

### Standards:

In order for digitization of historic federal publications to be done properly, technical-, usability-, and metadata standards must be utilized. Federal Agencies Digitization Guidelines Initiative (FADGI)<sup>5</sup> – formed in 2000 under the auspices of the National Digital Information Infrastructure and Preservation Program (NDIIP) and run collaboratively by GPO, NARA, Library of Congress etc. – was organized to “define common guidelines, methods, and practices to digitize historical content in a sustainable manner.” FADGI has done a good job in setting technical standards for digitization for a variety of purposes and uses. But to apply those standards we also need a standard for usability. Jacobs and Jacobs have proposed the *Digital Surrogate Seal of Approval* (DSSOA), a user-oriented “vocabulary for describing the quality of a digital object” to assure that digital reproductions are complete, legible, and accurate in order to “replace analog originals with digital surrogates for purposes of preservation or collection management.”<sup>6</sup>

### Access:

While there are already several public discovery and access points for historic digital government publications (GPO’s Federal Digital System (FDsys),<sup>7</sup> the Internet Archive<sup>8</sup> and HathiTrust<sup>9</sup> among them), such access and discovery can be enhanced in three ways: 1) Expand collaborative cataloging efforts (CIC, ASERL’s Collaborative FDLP project,<sup>10</sup> HathiTrust’s registry of US federal government publications<sup>11</sup> etc.); 2) Allow and encourage the FDLP community to upload their digitized publications and metadata to FDsys; 3) Encourage and enable bulk access to FDsys data and metadata for non-commercial use and reuse, and for collection-building and preservation (e.g., LOCKSS-USDOCS program<sup>12</sup>). Since digitizations differ in their completeness, accuracy, accessibility (e.g., HathiTrust currently only allows single page downloads to non-

---

<sup>1</sup> <https://www.google.com/googlebooks/about/>

<sup>2</sup> <http://www.cic.net/projects/library/book-search/government-documents>

<sup>3</sup> <http://www.gpo.gov/pdfs/news-media/press/11news13.pdf>

<sup>4</sup> <http://registry.fdlp.gov>

<sup>5</sup> <http://www.digitizationguidelines.gov>

<sup>6</sup> Jacobs, James A. and James R. Jacobs. "The Digital-Surrogate Seal of Approval: a Consumer-oriented Standard." *D-Lib Magazine*, March/April 2013, Volume 19, Number 3/4. Available at <http://dx.doi.org/10.1045/march2013-jacobs>

<sup>7</sup> <http://www.gpo.gov/fdsys/>

<sup>8</sup> <http://archive.org>

<sup>9</sup> <http://hathitrust.org>

<sup>10</sup> <http://www.aserl.org/programs/gov-doc/>

<sup>11</sup> [http://www.hathitrust.org/usgovdocs\\_registry](http://www.hathitrust.org/usgovdocs_registry)

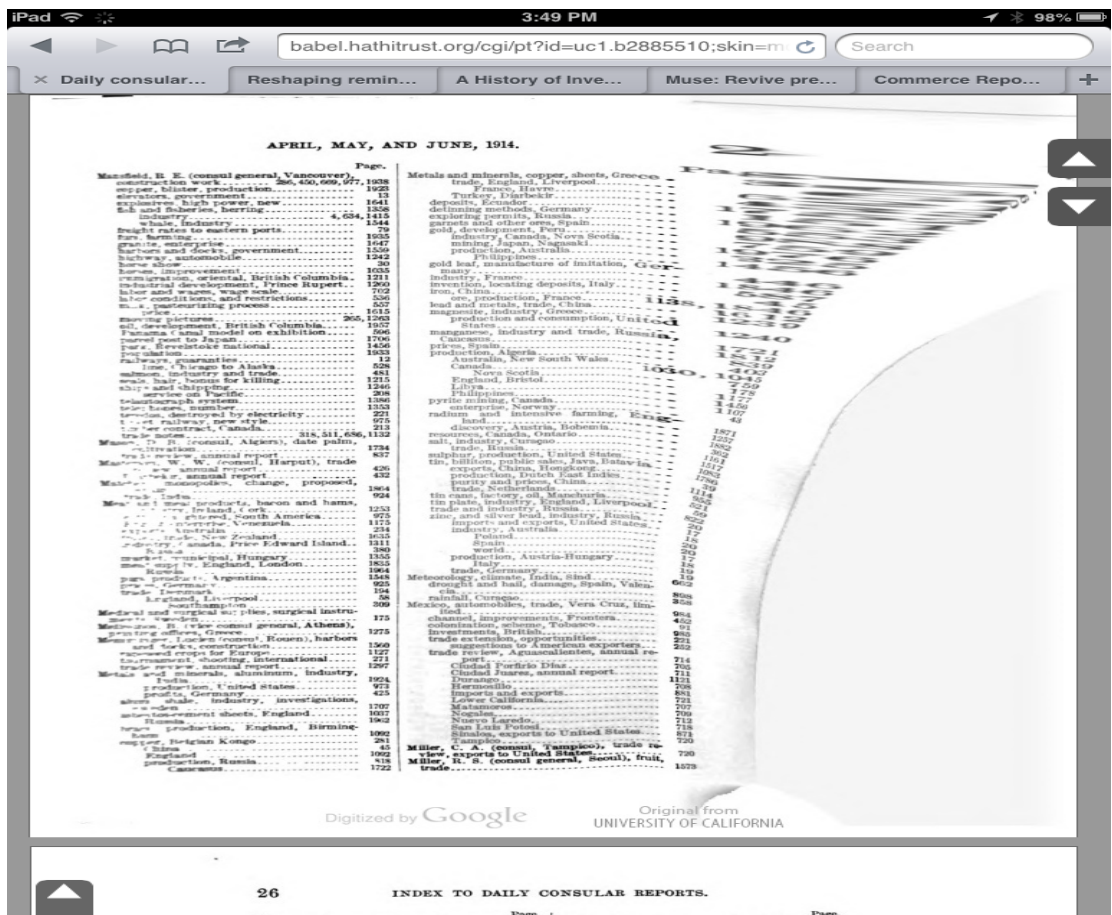
<sup>12</sup> LOCKSS-USDOCS program <http://lockss-usdocs.stanford.edu>.

members), and usability, the community needs metadata that records such differences to enable the building of complete, accurate, freely accessible digital collections.

## Preservation:

Digitization does not necessarily equal preservation.<sup>13</sup> In fact, some digitization projects – including Google Books Project – actually disbind and destroy materials in the process of digitization and there are no quality standards in place to assure OCR accuracy or image/text legibility. The FDLP community should therefore not make the fatal mistake of relying on digitizations of uneven preservation quality (see image below) to manage, weed, and destroy their local physical collections before a national plan is in place that takes into account both digitized and born-digital government information. Digitization needs to be seen holistically as an integral part of a collaborative, integrated national plan for the geographically distributed preservation of born-digital, digitized- and physical historic FDLP collections.

ALA GODORT Legislation Committee: Bernadine Abbott Hoduski (Chair), Kay Cassell, Brett Cloyd, James Jacobs, Kevin McClure, Wilhelmina Randtke, Kevin Reynolds, Melanie Sims, Antoinette W. Satterfield ~ January 2014.



<sup>13</sup> "Wait! Don't Digitize and Discard! A White Paper on ALA COL Discussion Issue #1a." James A. Jacobs and James R. Jacobs. June 2013. <http://freegovinfo.info/node/3961>.

<sup>14</sup> *Daily Consular and Trade Report 1914:2*. Originally scanned by Google and hosted in the HathiTrust digital archive. <http://hdl.handle.net/2027/uc1.b2885510>. See the many other beautifully odd and questionable scans at *The Art of Google Books* tumblr: <http://theartofgooglebooks.tumblr.com>