The Hunt for Digital Oregon Documents:
Building and Populating the Oregon Government Publications Digital Collection

by Jey Wann
Oregon Documents Coordinator,
State Library of Oregon
jey.a.wann@slo.oregon.gov

JEY WANN is Oregon Documents Coordinator at the State Library of Oregon, where she has worked for a long time. She particularly enjoys imagining researchers in the future finding just the right publication, partly due to what she does. When not immersed in the world of state government publishing, she enjoys reading, baking, gardening, walking and bicycling, performing early music, and chocolate. Activities that combine more than one of these are a bonus!

When the Oregon Documents Depository Program (OrDocs) began in 1907, the only way to access an Oregon state government publication was to get a paper copy. The OrDocs program distributed (and still distributes) state government documents to designated depository libraries around the state, enhancing government transparency and citizen engagement.

By the end of the century, however, it was obvious that the old model was no longer sufficient. The internet was becoming an additional avenue, if not the only avenue, for the dissemination of state government information. Libraries nationwide, including the State Library of Oregon (then known as the Oregon State Library), were looking for solutions to the problem of collecting digital documents.

By the early 2000s, the State Library, working with the Department of Administrative Services, OrDocs depository libraries, and the Documents Interest Group of Oregon, had begun planning a way to collect digital OrDocs. The culmination of this effort, in 2006, was the passage of House Bill 2118. The bill required state agencies to provide digital copies of their publications to the State Library, regardless of whether the publications were available online or not (Hulshof-Schmidt, 2006, p. 7).

At the State Library, we were in a good position to work with agency web publishing. We administered the search function on the Oregon.gov state web portal. We established the metadata scheme that agency publishers used, trained them to metatag, and worked closely with the e-government contractor. Most state agencies used a central content management system, which provided a single source for state web publishing.

With the help of our e-government ties, we developed:

- A metatag that agencies could use to indicate if a digital publication should be collected by the Library. (We later abandoned this, as agency compliance was spotty at best.)

- A daily feed from the content management system of all the new and changed documents pages published on https://www.oregon.gov/Pages/index.aspx.

- An in-house tool to review the documents and sort them into workflows for further action.

- An in-house site, the Oregon Documents Repository, to store digital documents and make them available to the public.

We dubbed the in-house tool the Repository Acquisitions Tool, or RAT.

Kyle Banerjee, the RAT's inventor, recalls:

One of the biggest challenges when creating the RAT was identifying which of the thousands of documents created each day on State of Oregon systems were good candidates for archiving—manually doing this was unfeasible. Another challenge was building workflows and structures that supported their processing and use. (K. Banerjee, personal communication, July 2, 2020)

It's difficult to over-emphasize the importance of the RAT. As far as we know, it's unique among state documents depository programs. Similar programs in other states rely on other methods, such as using a web crawler, regularly reviewing agency websites, or relying on agency staff to submit digital documents. Oregon Documents Specialist Angie Jannelli says "Having the RAT do the work of finding documents and acting as a tool to manage our workflow frees up so much staff time that we can then use to catalog digital documents. We are so grateful for it" (A. Jannelli, personal communication, July 8, 2020).

The State Library had been collecting physical Oregon state government documents for over 100 years. We had a proven workflow in place for collecting and cataloging the documents and distributing them to our depository libraries around the state. Now, we were faced with a virtual flood of digital documents, broad statutory definitions, and little previous experience to draw on. Arlene Weible recalls, "Once we were able to see the full range of state agency web publishing, we were overjoyed but also somewhat overwhelmed by the volume of material we now had to review" (A. Weible, personal communication, June 26, 2020).

We established criteria for collecting digital documents, including:

- Generally collect documents as-is, but convert Word, HTML, etc., to PDF

- Break extremely large documents into more manageable chunks

- Combine small documents published in multiple pieces into one file

We decided to do full MARC cataloging for all the digital documents. This made sense, as that's what we did with the physical documents. In addition, depository libraries are required to include records for all OrDocs in their online catalog or other finding aid, and were used to getting catalog records from us.

In the case of PDFs and Word documents, it was fairly easy to decide whether or not to collect something. Web pages were harder: when were they the equivalent of a publication? It was, and still is, not uncommon for something that was previously published as a tri-fold brochure to change to an HTML page. We initially developed an in-house tool to capture these, but it was a cumbersome process. We later decided to convert them to PDFs.

Although the RAT is still our all-purpose tool for identifying documents, hosting the Repository on State Library servers became an increasing problem. We migrated to the Islandora digital asset management system in 2017. The new platform enabled us to expand the scope of our digital collections, eventually including the digitized Oregon Index Online. We changed the name of the Repository to the more descriptive Oregon Government Publications Digital Collection.

Most of the documents we collect are PDFs, or are easily converted to PDF. Other formats can be troublesome. For years, we weren't able to collect digital video; we're now able to with Islandora.

We collect complex HTML documents via our Internet Archive partnership. Unfortunately, we can't upload these files to Islandora, but they are available in our public catalog.

More troublesome are formerly static documents that become interactive. Early on, we discovered that a number of directories had turned into databases. The most recent, notable example of this is the *State of Oregon Telephone Directory,* which lists state employees and agencies. It switched to a database in early 2020, ending a run of directories that started in 1938. The most ironic switch was the *Directory and Statistics of Oregon Libraries,* which ceased publication with the 1996/1997 edition. Recently, a number of statistical reports have also switched to databases. The amount of information available from them is incredible, but we have no mechanism to collect them.

Collecting born-digital publications is one thing. But there is another big thing: our physical Oregon Documents Collection, which may be the most comprehensive collection of Oregon government publications in the world. We want to make the entire collection available digitally, but the task is huge. One estimate, done a few years ago, projected that at our rate and staffing at the time, it would take 50 years to digitize the entire collection. Our staffing and equipment have improved since then, but it's still a monumental task. At the time of this writing, we have approximately 19 percent of the physical collection digitized (20,277 items).

We decided to prioritize digitization based on how widely the title is held, where:

- We only have one copy, and no other library holds it

- We have multiple copies, but no other library holds it

We also digitize publications for other reasons, including:

- Patron requests

- Inclusion in special digital highlights

- Cooperative projects with state agencies, such as anniversaries

- Other special projects based on special handling needs

Cooperative projects with other state agencies include state highway maps (with Department of Transportation), Oregon National Guard newsletters (with the Military Department), or celebrating the centennial of State Fire Marshal. Cooperative projects benefit the State Library, the state agencies involved, and the citizens of Oregon.

Collecting digital publications takes planning, hard work, and attention to detail. But it also can be fun. We never know what we'll encounter, whether reviewing born-digital documents or digitizing historic documents from our collection. From the State Fire Marshal's recent Sasquatch stand-up character (you can dress him in various fire-safety outfits) to Depression-era scrip, the collection is full of surprises. Digitization Specialist Kate Anderson, a Western Oregon University alum, was surprised and amused by rules for women students at her alma mater in 1925.

History

1907

Oregon Documents Depository Program established by the Legislature

1951

State Library begins publishing *Checklist of Official Publications of the State of Oregon*

1970s

State Library becomes the central cataloger for Oregon documents and establishes the OrDocs classification system

1979

Checklist ceases publication

1994

Two-tiered system of Full and Core depository libraries instituted; agencies required to appoint a publications liaison to the State Library

2005

Statutes governing the program changed to include digital publications; number of depository libraries reduced to 10; two-tiered system eliminated

2006

State Library begins collecting digital publications

2017

Digital documents move from an in-house platform to Islandora

The State Library has been collecting Oregon government publications for over 100 years, and collecting digital versions for less than 20 years. We're sure there will be changes in technology and new formats in the years to come. We're confident that State Library staff will rise to the occasion and continue to provide permanent public access for the citizens of Oregon.

References

Hulshof-Schmidt, R. (2006). State Library launches virtual documents repository. *OLA Quarterly, 12*(1), 7. DOI: https://doi.org/10.7710/1093-7374.1103

Iltis, D. & Wann J. (1998). Oregon documents programs at the State Library. *OLA Quarterly, 4*(2), 11–14. DOI: https://doi.org/10.7710/1093-7374.1482

Wann, J. (2006). History of the Oregon Documents Depository Program. Oregon State Library. Retrieved from https://digital.osl.state.or.us/islandora/object/osl:2970

Useful Sources
Oregon Government Publications Digital Collection:
https://digital.osl.state.or.us/islandora/object/osl%3Adocs

State Library of Oregon Digital Highlights
https://digital.osl.state.or.us/islandora/object/osl:digital_highlights

Oregon Documents Depository Program website:
https://www.oregon.gov/Library/collections/Pages/State-Government-Publications.aspx

Access to Oregon Government Publications infoguide:
https://libguides.osl.state.or.us/OrDocsRefRoom