

The Chesapeake Project: Preserving “Born Digital” Documents

by Dee Dee Dockendorf

HAVE YOU EVER RETURNED to a website looking for a document that you know was there last week, just to find the document has been replaced or taken down?

More and more, documents that are published on the Web are disappearing with no back-up or archived copies. The Virginia State Law Library, Georgetown University Law Library, and Maryland State Law Library are participating in a digital preservation program that hopes to solve this problem by harvesting documents that originated in digital form and might not be published in print. The Chesapeake Project, as it is called, was developed by the libraries to address permanency of online materials — a subject of great importance to the legal research community.

In the past, government reports, nonprofit studies, and other materials not commercially published were distributed to libraries across the country for cataloging and maintenance. As the Internet has become an easier and more affordable way to disseminate information, organizations have ceased to print and distribute materials and instead have created “born-digital” materials.

These materials have no hard-copy version — they were created digitally and published digitally. This in turn created a new obstacle for libraries. These materials are important and need to be preserved, but how? Are we to trust website managers to keep items on a website forever? Once that document is taken down, can it ever be accessed again? The Chesapeake Project aims to answer some of these questions by identifying, cataloging, and archiving born digital documents.

Each of the member libraries created its own collection development policy that states specifically what types of documents it is responsible for archiving.

The Virginia State Law Library’s statement is:

The digital archive collection of the Virginia State Law Library consists of all publications issued by the Supreme Court of Virginia, such as annual reports, special studies, handbooks, directives, etc.

The Law Library is also committed to collecting all publications issued by the Judicial Council of Virginia and the range of administrative divisions, commissions, and task forces operating within Virginia’s judicial branch of government. The Library will seek and retrain copyright permission for those materials not in the public domain.
— <http://www.legalinfoarchive.org/>

As an example, the Supreme Court of Virginia’s website contains Adult Drug Treatment Court Standards under the Drug Treatment Court section. The current version of the document reads “Adopted September 23, 2006 (Revised 10/07).” Unfortunately, the original version of this document was replaced when the revision was released. Since both of these documents were born digital there is no hard copy of the original standards in any library in Virginia. The Chesapeake Project has harvested both the original document and the revised document, so both documents will be forever accessible through the archive.

The Chesapeake Project participants have been collecting statistics and data about the project for the past two years, and early results have proven the effort to be very effective. A recently conducted two-year project evaluation showed 14.3 percent of the items harvested are

no longer available on the Web, up from 8 percent after the first year. We fully expect this number to continue to rise as the project continues. During this time, the project’s digital archive was populated with more than 4,300 digital items representing nearly 1,900 Web-published titles, the vast majority of which have no print counterpart.

If you are interested in viewing the materials collected by the Chesapeake Project, visit <http://www.legalinfoarchive.org/>. From this page, users can browse each institution’s collection or search single or multiple collections. To browse a particular collection, select Browse from the main page. Then use the drop-down box in the upper left corner to select each institution’s collection. To search, select the Search link from the main page and use the search fields to compose a search. The default is to search across all libraries in the project, but the user can easily change the selected libraries to one institution.

We will continue to add materials to the archive, and we are currently looking to recruit other law libraries to the project, hoping to inspire a “nationwide preservation program.” While the Chesapeake Project cannot save every born-digital document, we hope to create a robust archive of materials and to provide researchers permanent access to these critical documents.