Modern Language Association 2012 Annual Meeting

Session 125. What's Still Missing? What Now? What Next? Digital Archives in American Literature
http://www.mla.org/program_details?prog_id=D019A

Thursday, 5 January, 5:15–6:30 p.m., 608, WSCC

Program arranged by the Division on Late-Nineteenth- and Early-Twentieth-Century American Literature

Jeremy York Remarks:

These are my reflections on digitization over the last several years, and framing of the digitization and research landscape.

1990s - mid 2000s were about experimentation with digitization, developing standards, coming to understand the value that digitization has not only as a means of providing access to content, but of preserving it.

A number of what have come to be known as boutique projects were undertaken, focused on digitization of specific volumes or sets of volumes.

There were experiments also, particularly at Michigan and Cornell, Library of Congress, with production-level systems for digitized materials (experimentation with e-Text resources goes back to the 70s and 80s in projects like Project Gutenberg). So on one hand experimentation with digitization specs, small boutique scanning, on the other hand working to build systems capable of supporting digital preservation and access at scale (see Making of America as part of this latter).

From the mid-2000s on these have continued, but we have entered a period of large-scale digitization. The possibility was demonstrated by Google, there was money from Google, interest from libraries and other initiatives (e.g., IA), clear value to libraries and users.

Similar to boutique scanning projects, there have been targeted efforts that have arisen around certain authors, topics, texts, to use the digital materials in different kinds of research.

NINES is a project that seeks to bring many of these together, providing federated search across a variety of digital resources having to do with the long nineteenth century, including websites such as Willa Cather Archive, British Women Romantic Poets, many others, there are sites such as the Mark Twain archive, UVa was an early leader in this, many projects.

Many arise as parts of doctoral study - first gather materials, get them in a form

where they can be used for research, finally able to perform research.

Difficulties for projects have come in 1) lacking comprehensive coverage of the topic or area under investigation due to lack of digitized material; 2) different research needs leading to different ways of marking up materials and storing the underlying data (databases, XML), leading to barriers for integration, and cross-site or cross project research. There are also challenges libraries have faced in providing staff and resources to support digital humanities projects that faculty and students would like to do.

We have entered new era of collaboration around digitized materials, preserving them, providing access at scale. So will we, or should we, enter new era of collaboration around the ways we extend and make it possible to use the digital materials. Right now have distributed (traditional) model where researchers seek to gather material and build their own resources for their research (either on their own or in collaboration with libraries).

Also have librarians building resources (often with grant funding for preservation and/or access purposes and not always with input of researchers).

In a new model, as libraries collaborate to assemble a comprehensive archive of material, for instance in efforts like HathiTrust, we can imagine researchers drawing materials out of this archive for their projects to build their research corpus, but doing so in a common environment that is equipped with common tools and functionality (with capability for extension). Doing this in a common environment will on the one hand

1. Allow researchers to select, curate, and use their selected materials with less time devoted to establishing infrastructure
2. Provide greater possibilities for interoperability, long-term curation, and preservation of the materials produced in their research

And on the other hand

1. Allow libraries to offer digital humanities services at scale (more projects for more interested faculty, students, and researchers), and
2. Have the possibility to maintain the outputs of research projects for reuse and future expansion by others.

This is precisely what libraries have begun to do, for instance in the establishment of the HathiTrust Research Center.

Libraries have long been collaborating to improve services to their constituents and improve their own ability to serve research needs. Economic pressures, technological possibilities, and user demand are pushing libraries to a new level of collaboration on an unprecedented scale - collaboration with other libraries, with

faculty, students, and researchers. The drivers are there for transforming (if that is the right term - maturing or causing to mature might be better) the current model of curating and using our digital resources. I think It will take some time, some shifting in our thinking, trust that we can do more together than we can do apart, for the collaboration and proper infrastructure to develop to realize our full potential in this area. There will be a temporary gap, then, between the current smaller projects that have and are being created (projects that arise in the meantime) and projects that are "born-digital" in a new collaborative space, but I think this is what we are trending towards and should have in our minds as we develop new projects.