

# Preserving Antebellum Print Culture

Jason Peak

April 28, 2013

## Abstract

Poe's Magazine World, a digital humanities project centered around the literary culture of the 1840s in America, seeks to accumulate and present digital artifacts that re-contextualize this period in history for future readers, students and researchers. The project team is composed of a cross-disciplinary mix of scholars and students committed to establishing a research archive for the equally diverse audience of scholarly researchers, students and the general public. At its broadest, the scope of the project includes all American Antebellum authors and would take shape as part of a consortium of linked archives. As a first step, however, the project will focus on Poe's involvement with four periodicals, the *The Broadway Journal*, *Southern Literary Messenger*, *Burton's Gentleman's Magazine* and *Graham's Magazine*. The main tasks involved are summarized as follows:

## 1 Project Description

Understanding the project domain, background, goals and roadmap are essential to asserting the way forward. The domain of the project is hierarchical, and at its most general, it includes everything that can be defined as *Antebellum Print Culture*. Recognizing that small steps begin a great journey, the APC group have wisely decided to begin with a narrower scope, specifically, the writings of Edgar Allen Poe. As the first step, they have chosen to focus on his work as editor of periodicals. That effort is the subject of this discussion.

### 1.1 Domain

Poe is famous for his tales and poetry, indeed, every grade school child has been thrilled by *The Raven*, for instance. It is probably less well-known that in his capacity as an editor of several periodicals, he took the opportunity to offer commentary on the events and cultural themes that mark the American Antebellum period. The Poe's Magazine World project will draw attention to this aspect of the author and his perspective by creating a digital collection of four of the periodicals in which he played a key editorial role: *Burton's Gentleman's Magazine*, *The Broadway Journal*, *Graham's Magazine*, and the *Southern Literary Messenger*.

## 1.2 Background

The Antebellum Print Culture project has been underway for a few years, and in that time, there has been much discussion and converging general agreement over exactly what the content of the repository should be and how to best present it. In the last year, the project has focused its efforts around presenting a proof-of-concept prototype exploring three Poe tales. From a technical perspective, this has been achieved through adoption and customization of the Omeka web publishing platform. More will be said about Omeka later on, but for now, suffice it to say that the Omeka prototyping effort has exposed important requirements of the project, specifically, needs for:

1. a user-friendly management interface to the repository
2. flexible content presentation options
3. fine-grained metadata schemes
4. semantic relationships between entities, especially at the sub-document level

## 1.3 Goals

### 1.3.1 Immediate

Considering only the years in which he was actively employed by these editorials, the corpus contains just less than 4,000 pages. The primary deliverable for the next stage of this project will be a permanent, web-accessible repository of a scan for each page and fully searchable text.

**Scan** Each of the 4,000 pages contained in the corpus in scope will be scanned to high resolution TIFF

**Extract** OCR and minimal human processing will result in level 2 TEI[1] markup for each page

**Preserve** These artifacts must be preserved in a trusted system

As a result of these three tasks, the initial phase will establish a trusted digital repository (TDR) suitable as a research resource. Facilities for research will include, at a minimum, document retrieval and full-text search. With these basic foundation blocks firmly laid, other services can be built that further enhance the value of this resource for research including scholarly annotation, use of RDF for making semantic inferences, and a range of client web services that expose the archive to other repositories and applications.

### 1.3.2 Roadmap

Future stages of the project will enrich the text with annotations made by domain experts, reciprocal linkages with other repositories, and semantic identification of concepts at the word and phrase level.

## 2 Requirements

**Archival Preservation** The APC project requires trusted archival preservation of its primary artifacts and derivative scholarship and interactions.

**Exposure** Artifacts and their metadata must be exposed to client applications including search engines, browsers, and visualization applications

**Granular markup** Text documents will be encoded in the TEI at the various levels enumerated in [1]

## 3 Early Investigations

Early investigations into software solutions for this project have centered around the popular Omeka publishing platform. We have used Omeka as a holistic solution to the multiple requirements: archival storage, ingest, access. Omeka is a web platform written in PHP and based on the industry standard Zend framework. The application relies on a database, in our case, MySQL, to maintain state, application configuration, user access control, record content, record metadata, and RDF-like inter-record relationship data.

### 3.1 Archival Storage

Using Omeka as an archive has a number of advantages:

**open source, open standards** The full stack of technologies underlying Omeka as archive are open source and based on open standards. While the platform can be deployed on the Windows OS, it is more commonly hosted on servers running open source linux. While our institutional infrastructure has decreed that we use Red Hat linux, a commercialized distribution, we could have just as easily chosen Ubuntu or any of the myriad free distribution. MySQL is a well-known, widely-used database made available by Oracle.

## 4 Archives Software

A number of software packages exist for managing digital collections, and more are currently under development. Notable among these are ContentDM, Fedora, D-Space

## 5 Formats and Standards

### 5.1 File Formats

**Image** TIFF image format will be required

**Text** [all non0-image files will be XML]

## 5.2 XML Formats

TEI

METS

MODS

DC

## 5.3 Standards

DRAMBORA

# 6 Information Structure

## 6.1 METS Records

METS is a structural description and container format that can wrap other formats including TEI, RDF, MODS, EAD, etc. It enjoys widespread use and the endorsement of leading archival institutions, including the Library of Congress. As a structural description format, it is an ideal choice for storing records of multipage documents [citation]. Fedora uses METS.

# 7 Questions

1. What examples of RDF + Fedora exist?
2. describe the mapping from DC -> METS; is this automated in Fedora
- 3.

# References

- [1] TEI SIG on Libraries. Best practices for TEI in libraries. <http://www.tei-c.org/SIG/Libraries/teiinlibraries/main-driver.html>, October 2011.