White Paper Report

Report ID: 109088
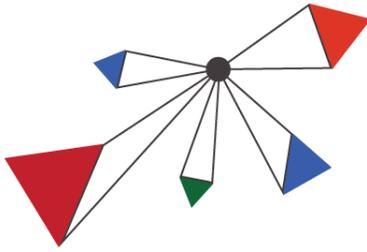
Application Number: HD-51718-13

Project Director: David Chinitz (dchinit@luc.edu)

Institution: Loyola University, Chicago

Reporting Period: 5/1/2013-8/31/2014

Report Due: 11/30/2014

Date Submitted: 11/30/2014

Metadata Schema for Modernist Networks
Level 1 Digital Humanities Start-Up Grant
(HD-51718-13)

White Paper
November 2014

Pamela L. Caughie and David E. Chinitz
Loyola University Chicago
Contact: dchinit@luc.edu

Level 1 Start-Up funding ($27,671) supported a one-day workshop leading toward the launch of Modernist Networks ("ModNets"), a federation of digital projects in the field of modernist literary and cultural studies. The workshop, which was held in Chicago on 17 August 2013, focused on the adaptation of the ARC (Advanced Research Consortium) metadata form to digital projects in modernist studies.

Background

The impetus behind the founding of ModNets arose from the perception that, although interest in digital modernist studies projects was increasing, there was little coordination between projects and no structured support for their development. The field of modernist studies has thrived and greatly expanded since the establishment of the Modernist Studies Association (MSA) in 1999. And digital scholarship in modernist studies has been flourishing, as seen not only by the growing number of digital modernist studies projects in development but by the marked increase in the number of papers and sessions with a digital focus at the annual MSA conference. Yet digital scholarship in modernist studies lags its equivalent in, for example, 19th-century studies, which for several years has had the support, the aggregating function, and the tool-development of NINES (Networked Infrastructure for Nineteenth-Century Electronic Scholarship).

Founded by Pamela Caughie and David Chinitz—both past presidents of the MSA—ModNets has the dual goals of establishing a vetting community for digital modernist scholarship and a technological infrastructure to support development of scholarly projects and access to scholarship on modernist literature and culture. ModNets aims to promote affiliated digital projects and centers; to provide editorial and technical support; to offer peer review based on content, conception, and technical design; to evolve standards and "best practices"; and to aggregate scholarly resources in the field.

In 2012, ModNets joined the Advanced Research Consortium (ARC), an overarching federation of similar organizations housed in the Initiative for Digital Humanities, Media, and Culture at Texas A&M University. The members or "nodes" of ARC include NINES (focused on the 19th century), 18thConnect (focused on the 18th), MESA (the Medieval Electronic Scholarly Alliance), and REKn (Renaissance English Knowledgebase). A meeting of the technical segment of the ModNets board in August 2012 determined that there was in fact a great deal to be gained by our joining forces with ARC, including vastly accelerated development of our aggregating infrastructure using ARC's resources and drawing on its experience with the creation of the preexisting nodes.

Metadata Issues

In order for ModNets projects to be searchable, their metadata must be consistent with the RDF metadata format used by the ARC nodes so that their resources can be categorized and searched by the COLLEX faceted search engine. ARC's metadata form originated in the scheme developed by NINES, the earliest and most mature of the nodes. However, the nodes now work together to develop metadata categories that will address the needs of scholars working in all periods. There are important differences. For example, the original ARC metadata specifications included "manuscript" as a *genre* term. But "manuscript" is not a useful classification for medievalists, for whom essentially *all* texts are manuscripts. And while to a scholar of the 19th century, a "manuscript" refers to an unpublished text, writers in the Renaissance and 18th centuries often circulated their manuscripts as a mode of publication. The metadata term "manuscript," and the category of *genre* in general, therefore needed to be rethought collaboratively, with input from all the nodes.

In the case of modernist scholarship, a key issue arises from the multiplication of *media* that came into use during the period, including film, phonography, radio, and (to a greater extent than before) photography. The historical visual and sound resources available to modernist scholars open up unique possibilities for digital projects in the field but also present challenges in terms of discovery, access, and preservation. Anticipating that digital projects in modernist studies will make considerable use of these resources, we recognized that the new media needed to be accommodated within the ARC metadata specification.

We are therefore applied for a Level I Grant to support a workshop that would bring together leaders of digital projects involving various media, ModNets leadership, and ARC representatives in order to review ARC's metadata vocabulary in the light of modernist scholarship and enhance it to describe the artifacts of modernism with sufficient clarity and richness. Two major projects, the Modernist Journals Project (MJP) and Editing Modernism in Canada (EMiC), were selected to provide representative metadata sets and use cases.

The Workshop and Its Outcomes

The one-day workshop in Chicago brought together four constituencies: (1) ModNets leadership; (2) ARC leadership; (3) project directors in digital modernism; and (4) metadata analysts.

Pamela Caughie and David Chinitz, who hosted the workshop, received assistance from their colleagues Steven E. Jones (English) and George Thiruvathukal (computer science), co-directors of Loyola's Center for Textual Studies and Digital Humanities.

The majority of the workshop participants were directors or co-directors of various digital projects in modernism: Pamela Caughie of Woolf Online; Mark Byron of the Beckett Digital Manuscript Project and the Digital Variorum Edition of Ezra Pound's *Cantos*; Tanya Clement of the Modernist Versions Project; Michael Hennessey of PennSound; Dean Irvine of Editing Modernism in Canada (EMiC); Jeffrey Drouin of the Modernist Journals Project (MJP); Laura Mandell of the Poetess Archive; Dirk Van Hulle of the Beckett Digital Manuscript Project; and Clifford Wulfman of the Blue Mountain Avant-Garde Periodicals Project. Nicholas Morris, a graduate student at SUNY-Buffalo working in the areas of digital humanities and film, attended as well, as did web developer Kristin Jensen. Also participating were Ann Hanlon and Erin Stalberg, both digital collections librarians with expertise in metadata. In addition to their aforementioned roles as project directors, Laura Mandell is the director of ARC and Tanya Clement its associate director.

Our goals for the one-day workshop were to produce

- A demonstrable working set of RDF documents derived from MJP and EMiC metadata that can be indexed and searched via COLLEX, the open-source aggregator for digital projects used by the ARC nodes; and
- A draft recommendation that details changes to the existing ARC vocabularies necessary to describe modernist resources.

The workshop's morning hours were devoted to presentations and discussions of the ARC metadata form and of metadata samples from MJP and EMiC. In the afternoon we divided into two breakout groups, each delegated the task of mapping either the MJP or EMiC samples to the ARC scheme. By engaging directly with this task of conversion, the groups were compelled to grapple with the details of the different metadata schemes, bringing to light the elements that mapped straightforwardly from one to the other and those that did not. We then reconvened to share our results and to plan out the "next steps" whose necessity had emerged from this hands-on work.

The workshop led to the realization that, for the most part, the flexibility and deliberate leanness of the ARC metadata form made the mapping process manageable for modernist digital projects, especially those, like MJP, that already had well-structured metadata. This was true thanks in part to ARC's ongoing program of metadata reform, in which ModNets personnel had already begun to participate. With the requirements of ModNets in mind, these modifications had included significant expansions in the options available for the *genre* and *discipline* tags, as well as in the list of media or formats available for the *type* tag. Participants in the ModNets metadata

workshop were pleased to discover that these changes had successfully anticipated many needs of digital projects in modernist studies. One of the most positive outcomes of the workshop was that the several project managers present left not only excited about the possibilities for the dissemination of their work through participation in ModNets but encouraged that the process of metadata mapping would not be as arduous as they had feared.

That said, the participants in the workshop also recognized the need for some follow-up work. In particular, it would be important to create RDF metadata samples for non-textual objects, particularly sound objects and film objects. An additional next step would be the creation of an XSLT for either MJP or EMiC that could be tested by ARC. Several participants volunteered to carry out these tasks in the months following the meeting.

These experiments resulted in a proposal to ARC for several extensions to its metadata scheme. The proposed extensions were adopted at its meeting of 24–26 Apr. 2014:

- Added to *genre:* Advertisement, Animation, Chronology, Documentary, Essay, Interview
- Added to *discipline:* Dance, Fine Arts, Sound Studies
- Added to *role:* Broadcaster, Cinematographer, Conductor, Director, Former owner, Interviewer, Interviewee, Owner, Producer, Production Company

Inevitably, as projects prepare themselves for ingestion by ModNets, we will learn about additional requirements that cannot yet be foreseen.

The metadata schema work accomplished by the workshop was a necessary prerequisite for this central service provided by ModNets. With this work completed, the mounting of ModNets is continuing to move forward toward a public launch. The metadata workshop also functioned to strengthen ties between the ModNets leadership team and key partners whose metadata will provide testing material for ModNets development and a searchable metadata core upon launch.

Additional Funded Work and Next Steps

Unspent money from the workshop was used, with NEH permission, to hire a student assistant to begin implementing the standards we refined at the 2013 workshop by actually ingesting metadata. As a result, the metadata for one project, WoolfOnline, has already been ingested successfully, the metadata for MJP is now on the verge of ingestion, and the metadata for Princeton's Blue Mountain Project is expected to be ready for ingestion by January.

ModNets will launch for public use in spring 2015 with these projects included in its searchable database, and with Editing Modernism in Canada on the way. At this writing we are in the process of hiring a project manager and are recruiting additional projects.

Appendix 1

Itemized Workshop Schedule

Saturday, August 17th

8:30 am – 9:00 am	Breakfast
9:00 am – 9:15 am	Welcome (Pamela Caughie and David Chinitz)
9:15 am – 9:30 am	Workshop Goals (Clifford Wulfman)
9:30 am – 10:30 am	Presentation of ARC metadata form (Laura Mandell)
10:30 am – 10:45 am	Break
10:45 am – 11:15 am	Presentation of MJP metadata samples (Jeff Drouin)
11:15 am – 11:45 am	Presentation of EMiC metadata samples (Dean Irvine)
11:45 am – 12:00 pm	Work plan for mapping (Clifford Wulfman)
12:00 pm – 1:00 pm	Lunch
1:00 pm – 3:30 pm	Working groups on mapping MJP and EMiC samples
3:30 pm – 3:45 pm	Break
3:45 pm – 4:30 pm	Discussion of mapping results and revisions to ARC schema
4:30 pm – 5:00 pm	Planning of next steps

Appendix 2

List of Participants

Appendix 3

Breakout Groups

MJP group	EMiC group
Jeff Drouin (MJP manager)	Dean Irvine (EMiC manager)
Cliff Wulfman (project manager)	Tanya Clement (project manager)
Ann Hanlon (metadata librarian)	Erin Stalberg (metadata librarian)
Laura Mandell (COLLEX expert)	Kristin Jensen (COLLEX expert)
Michael Hennessey (project manager)	Nicholas Morris (project manager)
Mark Byron (project manager)	Dirk van Hulle (project manager)
Pamela L. Caughie (ModNets leader)	David E. Chinitz (ModNets leader)

Appendix 4

ModNets Search Page (Prelaunch Staging Site)