INTERIM REPORT: OCTOBER 1, 2012 - JULY 31, 2013

HT-50070-12

DIGITAL HUMANITIES DATA CURATION

INSTITUTES FOR ADVANCED TOPICS IN DIGITAL HUMANITIES
NATIONAL ENDOWMENT FOR THE HUMANITIES

Project Director: **Trevor Muñoz**
                           Associate Director
                           Maryland Institute for Technology in the Humanities
                           University of Maryland

Local Project Directors: **Julia Flanders**
                               Director
                               Women Writers Projects
                               Northeastern University

                               Megan Senseney
                               Project Coordinator for Research Services
                               Center for Informatics Research in Science and Scholarship
                               Graduate School of Library and Information Science
                               University of Illinois at Urbana-Champaign

SUBMITTED JULY 26, 2013

Introduction

The Digital Humanities Data Curation (DHDC) institute series is a collaborative initiative led by the Maryland Institute for Technology in the Humanities in cooperation with the Women Writers Project, and the Center for Informatics Research in Science and Scholarship at the Graduate School of Library and Information Science (GSLIS), at the University of Illinois Urbana-Champaign. DHDC is designed to serve as an opportunity for humanities scholars with all levels of expertise—from beginners to the most advanced—to receive guidance in understanding the role of data curation in enriching humanities research projects. By the conclusion of each institute, participants will be adept in formulating solutions for existing challenges and will be able to document their data curation strategies in the form of a data curation plan and strategic risk assessment. The aggregation of resources and responses from across the institute's three events will also create a cohort of scholars focused on discipline-specific curation practices and skills.

Activities from October 1, 2012-July 31, 2013

Technical Development of the DH Curation Guide

The DH Curation Guide is a community-based curricular resource, providing an annotated, contextualized listing of key resources for understanding data curation in the humanities. The Guide was first developed at the University of Illinois as part of an IMLS-funded project dedicate to extending data curation to the humanities. For the DHDC institutes, the Guide will serve as a resource base and also as a mode of wider dissemination for the knowledge that is shared within the institute events: resources and annotations that arise through the presentations, discussion, and working groups will be logged into the Guide by the graduate student respondent. During this performance period, the project team shifted the Guide's hosting from Illinois to Maryland and also transitioned to a simpler platform that will reduce the overall time expended for programming and content management. A web page announcing the first DHDC institute was also added to http://dhcuration.org/institute/, and a screenshot of the institute page is included as Appendix A.

Curriculum Development

Due to Carole Palmer's planned sabbatical in Spring 2013, the project team began curricular development earlier than anticipated in order to benefit from Palmer's data curation expertise and prior experience in developing and running the IMLS-funded Summer Institutes in Data Curation from 2008-2011. Trevor Muñoz, Julia Flanders, and Dorothea Salo took the lead in planning a three-day curriculum, which is included as Appendix B. Each day of the workshop is comprised of presentations by session leaders, case studies, hands on activities, and group discussions. For our upcoming institute, Dr. Ted Underwood, Associate Professor of English at the University of Illinois Urbana-Champaign, has agreed to share a local case study presenting

his experiences and personal challenges addressing data curation issues related to his research, which includes running a variety of algorithmic approaches against large-scale text corpora for literary history analysis.

Planning First Institute and Participant Recruitment

Our first Digital Humanities Data Curation Institute was held at Graduate School of Library and Information Science at the University of Illinois, June 24-26, 2013. Planning for this event constituted one of the primary activities of this reporting period. All key planning activities occurred as planned and on schedule, including making local arrangements for meeting space and housing, creating an institute page on the DH Curation website, issuing a call for applications, reviewing applications, and notifying participants of their acceptance.

The project team had not, however, anticipated the overwhelming positive response to our call for applications. In total, we received 111 applications for our first institute, which is limited to 20 participants. In addition to institute applications, another 225 people signed up to receive more information about future workshops. Our first round of applicants' demographics ranged from local to international with 18 faculty members, 45 library and information science professionals, 7 alternative academics, 36 graduate students, and 5 others. With an 18% acceptance rate, we were able to be highly selective in our choice of participants, but this response indicates the need to develop methods for outreach at a larger scale. To best serve the needs of the broader DH community, we are making every effort to encourage remote participation by posting curricular materials online following the workshop.

Changes in Methods, Roles, or Work Plan

Shortly after the DHDC project was awarded grant funding, the project team revised the budget based on several factors. As stated above, we discovered that shifting the DH Curation Guide to a simpler platform would reduce our need for a programmer and allow for more streamlined content creation and management. Ultimately, we chose not to hire a programmer. In doing so we increased our funds for a graphic designer in year one and a graduate student respondent, hired through the Center for Informatics Research, in years one and two.

We also budgeted funds to ensure that the project coordinator and the graduate student respondent are able to attend the second and third institutes at the University of Maryland and Brown University. Increasing the time commitment of the graduate student respondent will prove invaluable in maintaining an ongoing connection with institute participants through the DH Curation Guide. The student's presence at the institutes will also allow us to more quickly and efficiently integrate materials coming out of the institute with the DH Curation Guide. Finally, we decreased Carole Palmer's efforts in year one due to her sabbatical leave in Spring 2013. This allowed us to budget support for an extra alternative academic participant at any one DHDC institute in the series.
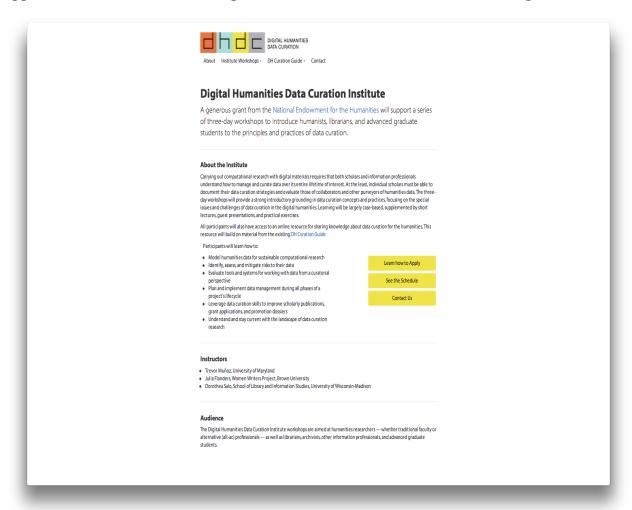
During this performance period, we hired Alison Fenlon as our graduate student respondent. Fenlon is pursuing a master's degree at the Graduate School of Library and Information Science at the University of Illinois, and she currently works as a graduate assistant at the Center for Informatics Research in Science and Scholarship. As an undergraduate student at the University of Denver, she studied English, German, and Creative Writing.

Next Steps

In the upcoming performance period (August 1, 2013-September 30, 2013), the project team will:

- Post edited curricular materials from first DHDC Institute to DH Curation Guide
- Maintain ongoing relationships for the cohort from the first DHDC Institute, which includes soliciting updates and facilitating ongoing virtual discussion
- Complete interim evaluation survey of participants from first DHDC Institute
- Plan second DHDC Institute at the University of Maryland
- Recruit the second round of DHDC participants

Appendix A: Screenshot of the Digital Humanities Data Curation IATDH Page

To view the webpage in its entirety, please visit http://dhcuration.org/institute/.

Appendix B: Curriculum for First DHDC Institute

Day One

9:00-9:15

Welcome

9:15-9:30

Carole Palmer, Digital Humanities Data Curation in the Data Curation Education Program (DCEP) context

9:30-10:00

Introduction to Humanities Data Curation, Part 1

10:00-10:15

Break

10:15-11:00

Introduction to Humanities Data Curation, Part 2

Exercise Discussion of pre-workshop readings

11:00-12:30

Participant Introductions

Exercise Data Profiles

12:30-2:00

Lunch

2:00-3:00

Social Dimensions of Data Curation

3:00-5:00

Data Management and Data Management Planning

`Exercise` Data Management Planning Tools

Day Two

9:00-10:00

Nature of data and digital objects

10:00-10:15

Break

10:15-11:30

Tour of Data Curation Systems and Platforms

11:30-12:30

Data Curation Case Study: Ted Underwood, Associate Professor of English, University of Illinois

12:30-2:00

Lunch

2:00-3:30

`Exercise` Case Study Problem Solving and Discussion

3:30-3:45

Break

3:45-5:00

The Role of Collections

`Exercise` Collection Evaluation

Day Three

9:00-11:00

Data Curation Practice

Exercise Curating Data from New York Public Library's "What's On the Menu"

11:00-11:15

Break

11:15-12:30

Data and the Law

12:30-2:00

Lunch

2:00-3:30

Risk assessment & Mitigation, Sustainability

3:30-4:00

What to do next and how to stay current

4:00-5:00

Wrap up and final questions

Slide sets and other curricular materials can be found on the institute wiki space:
https://github.com/digital-humanities-data-curation/dhdc-workshop/wiki