Supporting Digital Humanities Research: The Collaborative Approach

- Greenberg*, Hamlin°, Mylonas & Yott*
- *UVM, °Wheaton, ‘Brown

Copyright Hope Greenberg, Scott Hamlin, Elli Mylonas, Patrick Yott 2008. This work is the intellectual property of the authors. Permission is granted for this material to be shared for non-commercial, educational purposes, provided that this copyright statement appears on the reproduced materials and notice is given that the copying is by permission of the authors. To disseminate otherwise or to republish requires written permission from the authors.
Supporting Digital Humanities Research: The Collaborative Approach

This session is being blogged. Please check in on it and add your comments.


In this session we are going to present to you the experiences of 3 very different institutions as they take on the task of fostering and supporting digital research in the humanities. Before we each present a snapshot of our organizations and our work to you, I’d like to touch on some terms and characteristics which we feel define the projects we work on and the relationships that they engender.
Projects

A Digital Research Project in the Humanities:

‣ Engages with a research topic, whether practiced by faculty or students
‣ Can be motivated and carried out by faculty and by students
‣ Is informed by standards and practices recognized in the growing DH and DL communities
‣ Is self conscious in its use of the digital medium and digital rhetoric

A great deal research in digital humanities is formulated as projects, and that’s one area we will focus on today. In order to identify not only what constitutes a good project, but what is a dh research project at all, we are going to be using these criteria:

The project is one where the participants are engaging in research, they can be students or faculty in a class or not, but it’s not something that is primarily used for instruction. “Research is something we can all do”

A DH project are aware of current and emerging standards for information and is applying best practices to how it develops its data and its software.

A DH project is concerned with moving information and interactions to the digital realm, but is often also aware of the differences that working in a new medium makes to the way in which research is conducted. New participants may emerge, new questions may be asked because capabilities change and because the “remediation” of information makes us look for contexts and questions that we weren’t aware of earlier.
Roles

‣ Content Expert: faculty member/student researcher/archivist/community member
‣ Data Creator/Data Encoder
‣ Cataloger/Metadata Creator
‣ Technology Bridge Person: digital humanities staff/instructional technologist/digital librarian
‣ Technical Expert: programmer, software architect
‣ UI and Graphic Designer (accessibility, too!)
‣ Data Infrastructure Provider: digital repository
‣ Digital Infrastructure Provider: systems staff

As in traditional large research projects, there are activities in DH projects that require several types of expertise. Although in reality many of these roles may be provided by a single person, it is useful to understand the different contributions, and even the different moments at which they are necessary.
Characteristics

Digital humanities research projects have sometimes contradictory goals:

- Flexibility
- Individualized presentation
- Digital humanities standards and best practices

- Sustainability
- Standardized presentation
- Digital library standards and best practices

In this slide juxtaposes some of the desiderata and values of the two areas in which a great deal of DH work takes place, the DH support area and the DL.

DH projects, at the point where they interact with a faculty member, and try to instantiate their vision in digital form, are mostly focused on flexibility, branding and the best practices that come to them from other DH projects.

DH projects when they are regarded as library resources, are tend to prioritize longevity and standardization. They need to be incorporated into larger sets of collections, and be available for re-use. They are informed by DL standards.
Digital libraries

- Digital projects need resources that are high quality, reliably available and well cataloged
- Libraries preserve and disseminate digital resources as they do traditional materials.
- Digital projects should build on digital library assets, and drive their development

Early work in Digital Humanities (then called Humanities Computing…) was instigated by scholars and disciplines. A relatively new player is the emerging digital library. The role of the DL is very important, and the collaborations that are happening are a great development for the stability and support of DH projects.

Because; DH projects should be able to rely on high quality resources that are easy to find and are easy to use and re-use.

That’s what libraries do already with “analog materials”
Digital publications

- Significant as process and as product
- Are a way to transfer (digital) scholarly skills to undergraduate and graduate students
- Digital projects become scholarly product
- They become assets to be cataloged, preserved and disseminated
- They themselves become an object of scholarship
Digital Humanities @ Wheaton College

Scott Hamlin
Director of Technology for Research and Instruction
shamlin@wheatonma.edu
Wheaton College Overview

- Norton, MA
- Small private liberal arts undergraduate college
- >1500 students
- student faculty ratio: 11.7: 1
- 3% of classes have over 50 students
- LIS division
Digital Humanities Projects @ Wheaton

Support
- Research and Instruction Department
- Liaison model

Genesis of Digital Humanities Projects
- Awards and Grants
- Faculty Initiated
- Liaison Initiated
Character of Digital Humanities Scholarship @ Wheaton

- Research and Pedagogy linked
  - Example from the Sciences
  - Examples from the Humanities
  - Recent example where research comes first
- Using large, complex technologies in small ways
  - GIS
  - TEI
- Technology as a way to depth
- Projects are Collaborative:
  - Faculty/Archivist/Librarians/Technologists
Character, Part 2

- Level of support is high
- Challenge: Hardly ever polished or complete
- Advantages
  - Collaborations
  - Ability to carry out a project quickly
  - Faculty willingness to experiment with technology for teaching and research
- Process over Product
Book of Misfortunes and Shipwrecks:
A Text Encoding Project (TEI XML)
[Map of the world], Vesconte Maggiolo. Neapoly [Naples]: 1511
Overview of the Process

- Choosing the text (Faculty and sometimes Archivist)
- Project Planning (Faculty, IT, and sometimes Archivist)
- Transcription (Students)
- Structural Tagging (Students)
- Content Encoding (Students)
- Analytical Encoding and Glossaries (Students)
- XML-> HTML Transformation (IT and Students)
Determinado t&etilde;go &damp;spir; reduzir en e&longs;te vltimo libro algunos ca&longs;os &damp;spir; infortunios &iota; naufragios &iota; co&longs;as acaecidas en la mar:
Some of the entities used

<table>
<thead>
<tr>
<th>p with a tıde below</th>
<th>p</th>
<th>&amp;tilda;</th>
</tr>
</thead>
<tbody>
<tr>
<td>q tıde</td>
<td>q</td>
<td>&amp;tilda;</td>
</tr>
<tr>
<td>r tıde</td>
<td>r</td>
<td>&amp;tilda;</td>
</tr>
<tr>
<td>c cedilla</td>
<td>c</td>
<td>¸</td>
</tr>
<tr>
<td>iota (to indicate &quot;and&quot; in Spanish)</td>
<td>i</td>
<td>ι</td>
</tr>
<tr>
<td>small letter long s</td>
<td>s</td>
<td>˛</td>
</tr>
<tr>
<td>spiritus lenis d</td>
<td>d</td>
<td>ψ</td>
</tr>
<tr>
<td>spiritus lenis v</td>
<td>v</td>
<td>υ</td>
</tr>
<tr>
<td>p with low back hook</td>
<td>p</td>
<td>&amp;thickskip;</td>
</tr>
</tbody>
</table>
Encoded Text
Transformed text

Gonzalo Fernández de Oviedo
Libro último delos infortunios y
naufragios.

*Published*:

*Published as part of the *Missives General de las Indias*, Sevilla, En la imprenta de Juan Chandelger, 1556.*

*Permission has been given to Dawson/LaSalle to publish the text in electronic form by The John Carter Brown Library, Providence, RI.*

*Available for academic purposes, not for commercial distribution, all rights reserved.*

Supporting Digital Humanities Research
Student Response

“I learned not only a new way to represent data electronically, but it allowed me to do a very very close reading of each of the words, phrases, characters, definitions, etc.”

“[This work] is more interesting, allows a closer reading, and an online representation of your work, which is very rewarding.”

“I usually stink at computers, and this project showed me that I can do something really good with them and come up with a great final product.”

“I feel like its a great idea to build on this work we've done. It would be interesting to have more research done on the various people and places that we did not research.”
Student Reactions (continued)

“It was an interesting new way to analyze a text. I love learning languages and reading, but sometimes the lit classes are too dry. This class allowed us to better analyze the text linguistically and contextually.”

“I feel like I better understand the Spanish language. And not just in comprehension, but also the inner workings. For example, I used to put a tilde over a 'n' just because, but now I know what their purpose is. I also feel like I wasn't wasting my time with my work. Projects that can actually be used for something or by someone make me feel like there is a point to doing them.”

“I actually like working with TEI because it is a very mechanical process. It gets accomplished in steps so it is not so intimidating, although it is confusing at first.”
My talk today is going to be about how digital humanities projects have evolved at UVM, a historical tale, or perhaps…
Digital Humanities at UVM:

A History of Failure

A cautionary tale. Unless we define failure as Edison did…
“I have not failed. I've just found 10,000 ways that haven't worked.”

- Thomas Alva Edison
The University of Vermont (http://www.uvm.edu) is a mid-size university of approximately 12,000 undergraduate, graduate and medical students, across 7 colleges and over 100 programs. The core IT organization is Enterprise Technical Services, but many colleges and departments have IT groups of their own. Generally, at present the central IT organization and the satellite IT groups support things like administrative systems, the network, the computer store, the helpline, etc. In other words, faculty, staff, and student computing needs from the hardware level up through the application level.

I am ¼ of Academic Computing Services which was recently moved out of the main IT department and combined with the Center for Teaching and Learning and Media Development to form the Learning Resources Group. In our case, our mission, our credo if you will, has not changed…
Academic Computing Services

“We will explore
we will recommend
we will encourage you
we will teach you how to do it

but we can’t do it for you.”

Although not always possible or desirable we try to stick to this credo. Partly because our size makes any other option impossible, but mostly because we feel that the best way to understand IT’s impact on scholarship is to experience it. But what does this mean in actuality?
Yes we teach and support applications, but these are always changing. So when faculty come to us with questions, ideas, or pedagogical problems that IT might solve, we work with them to find a solution. And sometimes we work to find a solution to a problem that has not yet been expressed.
But what about digital humanities?

And that’s where the digital humanities story starts. In the early 1990s, about the same time we were installing our first web servers, I discovered the Text Encoding Initiative. Wonderful. It seemed obvious that this was something with true scholarly potential. That deep involvement with texts through markup combined with the delivery potential of the web, would add a fascinating dimension to humanities scholarship. So…
The Idea

We can build and support an infrastructure for humanities faculty and students to create the digital resources they need to advance their teaching, learning, and scholarship goals.
So, our first experiments began with some research: what were emerging standards, what technologies, what applications, etc. and some marketing. I found that a colleague in Special Collections was also interested in the application of SGML to humanities texts.
Together we planned a series of lectures and small events to inform and to gauge interest. Attendance was.....disappointing.
For our second experiment we decided to build some model collections. I’ll show you a couple of those.
Library Special Collections and Academic Computing, received a grant from Inso Corporation to install 2 DynaWeb servers to house electronic text collections. DynaWeb was an SGML-based corporate document publishing system.
The library proceeded to digitize and publish their Finding Aids using the Encoded Archival Description DTD.
For my sample project I chose to digitize several issues of Godey’s Lady’s Book, one of mid-19th century America’s most popular magazines. It is a frequently used primary resource for History, Literature or any scholars interested in the culture of this time period. So, great resource. Each page was scanned and OCRd, with the resulting text being marked up according to the Text Encoding Initiatives guidelines. The quality of the OCRd text was too poor to be displayed as a transcription, but it was just good enough to provide reasonable search results, upon which time we would display the actual page image. This model, fondly known as “dirty OCR,” was used by several humanities computing projects. This project gave us a sense of how much time would be involved in creating this type of resource.
Meanwhile. The library had received a grant from the Woodstock Foundation to digitize correspondence from its George Perkins Marsh collection. Author of Man and Nature, published in 1864, was one of the first to recognize and describe in detail the significance of human action in transforming the natural world.” He has been called the father of the environmental movement. This project also used the Text Encoding Initiative DTD but, as the TEI was built to be extensible, they used a customized version specific to letters/correspondence. In terms of our experiments, we learned the importance of help. This project hired student workers and also received many hours of help from some dedicated volunteers, including a professor emeritus who was interested in Marsh.
Our 4th project came from Nancy Gallagher, a woman who had just completed an MA in History at UVM and had published her book “Breeding Better Vermonters; The Eugenics Project in the Green Mountain State.” As part of the research for that book she had scanned hundreds of documents from repositories across the state. She had received a grant to develop this topic into an interdisciplinary educational resource to facilitate use of primary historical sources. But there was enough interest in the community for something broader than that. Most of the grant went to student help, which was an excellent investment.
Experiment 3:  

“Grow” student expertise

The response to our model projects was wonderful. Many users, many inquiries. At least from people outside the university. Our faculty... well. So we planned the next experiment. We had had such good results with student help. We thought that if we could develop a cadre of experienced students, we could then match to faculty projects, either pay them or give them credit. I also wanted an opportunity to teach about digital humanities. Neither History nor English were interested in such a course at that time, but the Computer Science department allowed me to offer CS005: Introduction to Applied Humanities Computing, as an undergraduate course. The course was a combination of the theoretical and the practical. We studied topics in humanities computing and also the practical aspects of digitization, scanning, OCR, etc. I taught them xml in the form of the TEI, and xslt to transform our marked up text for web publication.
The text that published was from Special Collections. It was an unbound edition of Hiland Hall’s “History of Vermont” published in 1868. It had been unbound and then each page had been remounted on individual pages. This had been done at the behest of Hall’s great-grandson. Hall had been a collector of autographs and documents related to Vermont History. His great-grandson continued to acquire documents, and envisioned including one such document for each page of the History. What he was creating was a sort of paper hypertext. At any rate the students loved working with real historical documents and building something that could be used by others. There was just one little problem. Because the course could not be used by lower level students to fulfill History or English requirements, most of the students who took it were seniors. Our cadre of experienced students... graduated.
Experiment 4:
Combine forces

UVM tends to be a highly distributed environment. Ideas and projects tend to happen in isolation, or to use an icon of the Vermont landscape...
Silos. Projects and resources developed in one area don’t necessarily connect to similar resources developed in others. This isn’t necessarily a bad thing, except, of course
it makes project longevity a more precarious proposition
Interest in online collections was growing, so we decided to bring together the
various groups and people on campus who wanted to further efforts in this
area.
We made several discoveries: our several constituencies really did have differing needs. In fact, though we had been thinking in terms of “humanities computing” several of the projects were not actually in the humanities. The types of needs fell, roughly, into:
- image collections for use among several faculty in conjunction with courses, whether in classroom or for assignments. Challenges: need for secure access due to copyright restrictions
- exhibitions which combined searchable image collections with narrative frameworks, for class use, but also for the larger community.
- electronic texts, both for display and for study.
Most members of our group were looking for ways to build image collections, so we started there. We had already installed dSpace for experiments with an Institutional Repository but it also provided an easy interface for our faculty to build image collections. We added ContentDM for its image handling capabilities and easy interface, then began to compare the two. BTW we are a mixed Mac/Windows environment, and ContentDM was and is a Windows-only product, which was why we did not go completely in that direction, although we later wrote some scripts to batch upload jobs from Excel so that Mac-using Art faculty could use the product.
The six to seven faculty who rotate through teaching the several sections of Intro to American History built the History 11/12 collective where each member could add the images they needed for their sections.
Several Art History courses built a similar collection of art images. Many of their images were under copyright restriction which necessitated a secure environment which ContentDM handled easily.
The Landscape Change Program was funded externally and was built from the ground up. The project itself compares historical images of Vermont with contemporary images. In many cases the contemporary images are taken by students in the related course but the project also accepts images taken by community members. You can search by town, etc.
These experiments led us to conclude that there was indeed room for digitization efforts at UVM, and that broader institutional support might now be possible. We needed a more formal approach.
i.e., get more $$
So, after successful application and granting of earmark funds, we hired our first Digital Initiatives Librarian, Winona Salesky, in 2006. The current emphasis (based on the main source of funding) is on the congressional papers of Senators Jeffords and Leahy. Another large project is to digitize the thousands of photographs of Vermont in the McAllister collection. BTW this system is built using eXist, the XML database. Winona has also created a system using xForms whereby library cataloguers can handle the creation and application of the metadata for these items so that task is not left up to the students who are scanning and transcribing.
We have learned…

ideas/vision: good
Institutional buy-in (support or money) is necessary
We will continue to build collaborations and are mindful that many faculty are quite IT-skilled, while many IT people have a deep interest in the content that is being created.
One size need not fit all: we do not have to build one major system that will do it all. Working with multiple systems is still a viable approach, not quite the support nightmare one might envision.
In terms of Academic Computing and faculty projects, the CDI relies on a Collections Development Committee, so the initial faculty-built collections will likely use materials drawn from Special Collections. However, we will probably run a parallel server based on the systems Winona is developing, for faculty-built collections that are drawn from elsewhere, especially those that may not have the metadata needs of the “official” projects (for example, faculty photograph collections used for classes).

We still need to work on getting more faculty to think like producers of digital projects instead of just consumers

We will continue to experiment, to find new methodologies for adaptation, to encourage and react to our faculty needs. We will continue to do research...

And of course, our definition of research has been and may continue to be...

History methods class
“If we knew what we were doing it wouldn’t be called research.”
University of Vermont Center for Digital Initiatives:
http://cdi.uvm.edu

Questions:
hope.greenberg@uvm.edu

Winona Salesky’s blog on the process of building the CDI:
http://thedil.wordpress.com/
Digital Humanities at Brown

The semi-permeable membrane...

Elli Mylonas, Scholarly Technology Group
Patrick Yott, Center for Digital Initiatives
In Brief

• “University-College”
• 680 faculty
• 5800 undergraduates and 2200 graduates
• Support for digital activities
  • CIS-Teaching and Learning Services (6)
  • CIS-Scholarly Technology Group (3)
  • Library-Center for Digital Initiatives (5.5)
1964-Development of Brown Corpus (Linguistics)
1968-Development of Fress Hypertext System (CS)
1983-Institute for Research in Information and Scholarship (IRIS) develops seminal Intermedia hypertext system
Late 1980s-> George Landow and later Bob Coover teach hypertext fiction and digital writing.
1994-The Scholarly Technology Group (STG) is formed
Supporting Digital Humanities Research
• 3 Staff
• 4-5 undergraduates, 3-4 graduate students
• Attract projects through a “grants” program
• Serve as seed projects for external funding
• NEH/other grants
• Close collaboration with the Women Writers Project
• Close collaboration with the library’s Center for Digital Initiatives
Monastic Archaeology Project

Exemplary Project:
http://dev.stg.brown.edu:8080/exist/monarch/index.html
Florentine Gazetteer based on 1585 axonometric map.

Exemplary project (http://www.stg.brown.edu/projects/florentine_gazetteer/)
Fortune Hunting: An art project that plays with the idea of the archive

Art Project: http://www.fortunehunting.org
• 5.5 staff (+ affiliated staff throughout library)
• Digitization of library’s special collections
• Support for faculty projects
• Development of campus-wide digital repository
• Development of dl software and tools
• Close collaboration with STG
Digitizing Special Collections
Digitized Materials in the classroom
Faculty projects
This selection of works from the extraordinary collection of Vincent J. Buonanno '66, focuses on views and maps of the Eternal City from the 16th-18th centuries. In addition, festival prints and architectural treatises from the collection of the John Hay Library are included.
This is the STG view of the same project. Note that it allows access to the book as objects, but also has a number of other items that allow different forms of access. It’s also more “branded”
For example, we want to link the digitized books in the library repository, as well as images from other collections of related materials (at Uchicago, for ex.) to an encoded inventory of the contents of a contemporary print shop.
This is the same page that we saw from the CDI site. This is also different in that it contains metadata about the page on the right.
I went to another page with richer metadata in order to show you the extent of it. These pages are the heart of the collaboration between the CDI and STG on this faculty project, and the locus of the semi-permeable membrane. We had planned on building the project on top of the library repository, and drawing images and metadata directly from it. We knew that it was very important to the faculty member, who studies books and also engravings that she wanted to be able to interact with information at a level that was more detailed than the kind of cataloging a librarian can afford to do, and may even want to do.

So STG took the initiative to build on the libraries metadata structures, and, working with the CDI, to enhance the METS and MODS records in order to record information about each page. The research and encoding of this are ongoing and are being done by a graduate student in Art History.

The really interesting aspect is that this project demonstrated a way that STG and CDI not only can work together to develop robust resuable information so that it can serve a particular research interest as well as the archival needs, but also to push all of our understanding of ways in which we have to encode and disseminate rich information.
Thank You!