White Paper Report

Report ID: 107119

Application Number: HK-50015-12

Project Director: Martin Foys (mfoys@drew.edu)

Institution: Drew University

Reporting Period: 9/1/2012-10/31/2015

Report Due: 1/31/2016

Date Submitted: 2/10/2016

White Paper:

The DM Environment: From Annotation to Dissemination, 2012-2016

NEH Digital Humanities Implementation Grant

Project Directors:
Shannon Bradshaw, Drew University (former)
Martin Foys, University of Wisconsin-Madison
Dorothy Porter, Schoenberg Institute for Manuscript Studies

TABLE OF CONTENTS:

Executive Summary …… 2

Project Description …… 2

Project Timeline, Activities and Accomplishments …… 3

Future Work, Audience, and Long Term Impact ……..… 8

Appendix A
Screenshots of Current DM Environment and Projects ….. 10

Executive Summary:

Over the course of the thirty-six months of this grant period, the DM project has undergone a rather seismic set of shifts in administrative and technical staff, institutional home, technical platform, and developmental vision and goals, as well as facing a set of unforeseen technical challenges. In spite of this, the DM Project is concluding the work it achieved funded by the NEH Digital Implementation Grant with a full set of project goals realized or very near completion, with additional funding to finalize work and dissemination on current use-case projects in the next twelve months, and in a position and with new funding to continue implementation for two more years following.

Project Description:

DM at its most basic is a tool for linking media, through an environment for the study and annotation of images and texts. It is a suite of tools, enabling scholars to gather and organize the evidence necessary to support arguments based in digitized resources. DM enables users to visually mark targets of interest in manuscripts, print materials, photographs, etc. and provide commentary on these resources and the relationships among them. There are four types of resources with which DM permits the user to work: images, texts, and selections of images or texts as marked out by a user. A user may create links between any combination of resources. The most common is a link from a textual annotation to the image, text, or fragment it describes. In many projects, a single annotation will reference (e.g., for comparison) targets from several images. DM is designed to enable scholars to easily create these and other types of relationships among resources. Users can upload their own digital images and texts to create a project collection to work within, add multiple users to collaborate on an individual project, and publish a linked and annotated DM collection as a public, read-only resource with its own World Wide Web address.

An additional objective in this project is to continue to develop an understanding of scholarly work processes in order to effectively support research as it is practiced now, while opening the door for new methods of scholarship to emerge. As a result, DM developed functionality to allow users to export the linked data they create for database use outside the original DM environment and display. Data may be exported in a standard code formats such as RDF-3, XML and TTL, and are designed to be adaptable to Open Annotation Collaboration and Shared Canvas protocols.

For representative screen shots of DM's functionality and current projects, please see Appendix A, at the end of this report. Additional information on the DM project may be found on host University of Pennsylvania's Schoenberg Institute for Manuscript Studies' (SIMS) project page: http://schoenberginstitute.org/dm-tools-for-digital-annotation-and-linking/

Project Timeline, Activities and Accomplishments

Phase 1 (October 2012 – January 2014): During the first half of the grant period, the DM Project achieved the following work in the two main areas of project development – resource programming and developing a test project cohort:

- Finalizing the functionality and graphic elements for the on-line user interface. This work included refinements to the first prototype's image navigation and magnification features, graphic representation of selected targets for annotation, the implementation of a "multi-up" viewer (*Appendix* A1-A2), where any number of images in the resource can be opened and viewed for annotation or linking, the development of an architecture for authenticated users to allow access to multiple projects, and the design and implementation of a search feature for annotations and metadata (*Appendix* A9).

- The original prototype architecture and format for the linked and annotated data generated by a project was fundamentally overhauled. The RDF-3 format for the data was refined to be compliant with current common standards being developed in the Open Annotation Collaboration (OAC)[1] and the Shared Canvass Project[2] being developed at Stanford University. An export feature was built to allow users to save out all generated RDF-3 data in either XML or TTL formats for data mining in other projects. Because of issues handling the amount of RDF-3 data being generated by the larger implementation projects (see below), a new database store (4Store) was researched and implemented.

- Working in conjunction with the Stanford University's Digital Medieval Manuscript Initiatives, foundational developmental work was done to build an environment to pull images directly from the servers of several existing digital manuscript repositories (consumed by DMMI), including the Parker Library (Cambridge Univ.), Johns Hopkins (Roman de la Rose Project), Beinecke Library (Yale University) and the Walters Art Museum (*Appendix* A6).

- During this development work, several implementation projects were begun or continued.

  a. Most significantly, the Virtual Mappa project[3] at the British Library was begun, under the supervision of Kimberly Kowal, a digital maps curator at the BL, with the selection and digitization of eight early English maps of the world as representative cohort for the potential of DM's features. During this work, the map cohort was uploaded to the DM environment, and project members at the British Library identified, marked and annotated some 2000

[1] http://www.openannotation.org/spec/core/
[2] http://iiif.io/model/shared-canvas/1.0/
[3] Overview and images of the work in progress provided in a series of British Library blog posts: http://britishlibrary.typepad.co.uk/magnificentmaps/2014/03/good-news-for-fans-of-medieval-maps.html; http://britishlibrary.typepad.co.uk/magnificentmaps/2014/05/tools-1.html; http://britishlibrary.typepad.co.uk/magnificentmaps/2014/05/-tools-2.html;

geographic inscriptions on all of these maps. These maps are drawn from the British Library's collection, as well as from libraries at Oxford and Cambridge Universities, and additionally include a digitized version of the Hereford Mappamundi from Hereford Cathedral, the largest medieval map of the world surviving (*Appendix* A1-A2).

b. A second project, *Insular and Anglo-Saxon Illuminated Manuscripts: An Iconographic Database*, was also by project partner Asa Mittman at California State University-Chico. by Asa Simon Mittman (California State University, Chico). Mittman is using DM to develop a new digital edition of Thomas Ohlgren's *Insular and Anglo-Saxon Illuminated Manuscripts: An Iconographic Catalogue,*[4] to expand access to, streamline the use of, and augment with digital images the original content of this invaluable print resource for medieval art historical study. Mittman and his research assistants used DM to turn the print catalogue into a digital database, and additionally through the DM environment began to link the electronic catalogue entries to relevant manuscript images already available in online digital manuscript repositories (*Appendix* A5).

c. Use-case partner the *Dictionary of Old English*[5] at the University of Toronto finished their current work using DM to annotate examples of textual cruxes found in manuscripts in the Parker collection. Through DM, these examples in Parker manuscripts where identified, targeted and annotated with reference to *DOE* entries, while DM also generated thumbnail images of the examples to be included a future release of the online *DOE*.

d. Use-case partner Lisa Fagin Davis (also the Executive Director of the Medieval Academy of America) completed her work using DM to create annotations and a searchable database for her translation, critical edition and detailed study of a fifteenth-century French world chronicle scroll (fifty feet long), just published by Brepols.[6] For the publication, the DM environment was especially adapted for display from a DVD, so a browsable, digital version of the scroll could be included with the book. Under terms of this work, in the next phase of DM development (see below), this DM version of the scroll will also be made freely available to the public through DM's new institutional home at the Schoenberg Institute for Manuscript Studies (*Appendix* A3-A4)

e. In addition, a number of smaller individual scholars and projects used developing versions of DM in their work, including the *Parker's Scribes* project at the University of Toronto and the University of Oxford, the Pompeii Quadriporticus Project (by Eric Poehler, University of Massachusetts Amherst, who used DM on iPads in the field at Pompeii to compare, mark-up, and annotate archival images Pompeii's monumental

---

[4] For overview and scope of the original resource, see Lawrence Nees' review: http://www.jstor.org/stable/2854374

[5] http://doe.utoronto.ca/

[6] http://www.brepols.net/Pages/ShowProduct.aspx?prod_id=IS-9781905375554-1

structures), while scholars at Stanford University and the University of Virginia worked with DM technical lead Shannon Bradshaw to use DM to pursue sound-based research questions focused on the work of the fourteenth-century composer Guillaume de Machaut. DM Project co-director Martin Foys also used DM to develop a digital mini edition of a short Old English text from a Parker Library manuscript, to help showcase the resource's utility for traditional textual editing (*Appendix* A8).

Phase 2 (October 2013 – September 2014):

Towards the end of this development and use-case phase, the project went through a series of key personnel and institutional changes. In May, 2013, co-director and technical lead Shannon Bradshaw took a leave from Drew University, in order to work in industry. Given his new workload, Bradshaw subsequently decided to maintain a consultative relationship only with the DM project for this period, but for all practical purposes ceased contributing any meaningful work towards the project at this time. During the next year, technical work on resource was largely maintenance-based, as Drew's developmental team consisted of a small team of students, under the direction of senior computer science student Tim Andres, who had been working on DM for the past three years. In the summer of 2013, project co-director Foys accepted a new position at King's College London, with the understanding that KCL's robust Digital Humanities program would be able to provide a new and future home for subsequent DM development and work. In 2014, due to internal issues at KCL-DH and the departure of several key people in the department, it became clear that KCL would be unable to provide a sustained institutional home for continued DM development. Consequently, Foys began investigating options for a new institutional home for the DM Project, so that the project's grant goals could continue and the remaining grant funding appropriately expended. Subsequently, in May of 2014, the Schoenberg Institute for Manuscript Studies (SIMS) at the University of Pennsylvania agreed to become the institutional home for continued development of the DM Project for at least three years, under the direct supervision of Dot Porter, who replaced Shannon Bradshaw as technical lead, with new development work slated to begin in October of 2014. In order to allow this subsequent, extended period of development work to begin under the funding of the NEH grant, a twelve month no-cost extension was requested from the NEH, and granted.

During *Phase 2*, annotation work continued on all of the projects noted above. Given the liminal and at times uncertain state of the project during this middle phase, no new projects, extension of current projects, or developmental work was undertaken, so that the projects and work already completed could be stabilized and concluded, if a new home and technical team for DM could not be found.

Phase 3 (October 2014 – September 2015):

The first months of *Phase 3* were largely administrative in nature, effecting the porting of all data and software from Drew University and SIMS, securing the transfer of the

subaward of NEH grant monies to SIMS for continuing work, and the hiring of a technical team to begin the next phase of development work. To continue technical work on the DM resource, co-director Dot Porter at SIMS contracted with Performant Software, a commercial software company that SIMS had used successfully for digital Humanities work in the past. Work began in November of 2014, but progress was slow, owing to a contracted developer who was unable to solve several issues required to move forward. In January, this developer was replaced by the senior developer at Performant, and progress accelerated almost immediately. Performant was able to implement the following functions/features in the DM project, realizing most of the outstanding technical goals for the project, including:

- a read-only user mode allowing public access to DM projects
- a web embed feature allowing public access to DM projects on the WWW from other WWW pages
- user enabled upload of images to the DM environment,
- administrative architecture for users, and the ability to add new and multiple users to project accounts (*Appendix* A7)
- the ability for users to create new projects, and a number of performance refinements.
- a full backup system, so users can formally save data and images contained in a project and restore it to the DM environment in the case of data corruption or other emergency.

Moving institutional hosts for DM also meant porting older projects over, and the upgrading and debugging of three current projects from Drew University in the new network environment at SIMS: the British Library's *Virtual Mappa* project, Asa Mittman's *Insular and Anglo-Saxon Illuminated Manuscripts: An Iconographic Database*, Martin Foys' *Old English Bell Tokens* project. All of these projects are now on the SIMS servers, in the latest iteration of the DM environment, and will be able to be made publically available once the final version of DM can finished and debugged (see *Phase 4*, below). In addition, a new project, Marie Turner's *Medieval Scrolls*, was begun as a test case in interoperability, with its XML data from a pre-existing project successfully ported into SIMS new DM environment for continued work.

At the British Library, a Mellon-funded digital Humanities project, *Pelagios 3.0*,[7] was able to use exported DM data from the *Virtual Mappa Project* to contribute to their ongoing initiative to create a data-driven, online gazetteer of the ancient, medieval and early modern world.

Unfortunately the last month of Performant's grant-funded work through SIMS, serious bugs in the software manifested, which, as data in a DM project increased, increased the likelihood of fatal and un-reversible corruption of the project's data. Multiple weeks of work were lost trying to find the source of this programming flaw, which had not presented until the latest stages of the project's development, and additionally compromised and set back the work of several of the ongoing projects. Ultimately, and despite working for a number of weeks *pro bono*, Performant

[7] http://pelagios-project.blogspot.com/p/about-pelagios.html

developers were unable to identify and fix the source of the fatal data corruption, and consequently the grant period ended without the ability to finalize and publish both the DM resource for free, public use, nor the major and minor projects which had continued through this development phase.

At the end of this phase, co-directors Foys (now at the University of Wisconsin-Madison) and Porter (SIMS) held a series of meetings with Performant to figure out the necessary steps to rebuild the backend of the database architecture so that DM can be finished, as the majority of At the end of this phase, an additional $50,000 funding and material support was secured from University of Wisconsin-Madison and SIMS to allow work necessary to the completion of current project goals.

Phase 4 (February 2016 – June 2016; *outside the schedule and funding of the NEH Digital Implementation grant).*

To allow current, fatal data corruption issues in the DM Project environment to be corrected, ongoing projects to be published and made publically available, and the DM resource itself to be released for public use, dissemination and adaptation, SIMS and Performant have developed a technical strategy and workplan to replace the corrupting backend data structure with a more stable architecture, while keeping all of the front end interface and functionality. This work will commence in February-March 2016, and be finished in four weeks, with testing and porting of the data of current projects into the new data environment, and be funded by support from SIMS and a new internal grant for DM development from the University of Wisconsin-Madison (a mandate of a new position that Foys accepted last year).

After the DM resource has been stabilized, the current projects will receive final edits and be ready for public release, hosted on SIMS servers. The code base for the resource will also be open source, and uploaded with all related materials to the projects GitHub site.

Future Work, Audience, and Long Term Impact

Future of DM at SIMS, public release and dissemination

Immediately after the bugs have been fixed, SIMS will update the DM source code available through the University of Pennsylvania Library's GitHub account, which is currently in the 2012 version provided by Tim Andres.[8] Accompanying the code will be complete documentation on how to install and use the software. SIMS will also install a working copy of DM and make it available as a sandbox for users who are interested in experimenting with it before they install it themselves. In the future SIMS may make available a more permanent DM for hosting projects, but that will depend on the interest of the community.

[8] https://github.com/upenn-libraries/DM

SIMS will continue to host Marie Turner's edition of 15th-century genealogical rolls as a SIMS publication, as well as supporting ongoing projects such as *Virtual Mappa*, *La Chronique Anonyme Universelle* and others currently in development, and we fully expect to use DM for SIMS projects in the future. Dot Porter and Will Noel (Director of SIMS) will teach a class with Rare Book School on "The Medieval Manuscript in the 21st Century," and they look forward to incorporating DM as an option for the final course project. To further encourage the use of DM in the medieval studies community, SIMS will host DM workshops at the International Congress on Medieval Studies and other subject conferences.

Finally, SIMS will continue to seek out opportunities to continue to develop DM. SIMS is particularly keen to have DM compatible with the International Image Interoperability Framework (IIIF), a developing standard protocol for enabling access to digital images which is being adopted quickly by the world's medieval manuscript-holding institutions, and which Penn will be adopting itself over the next couple of years.[9] This work will allow DM to activate existing, but currently latent architecture for opening, adding and annotating manuscript pages directly from digital repositories (*Appendix* A6).

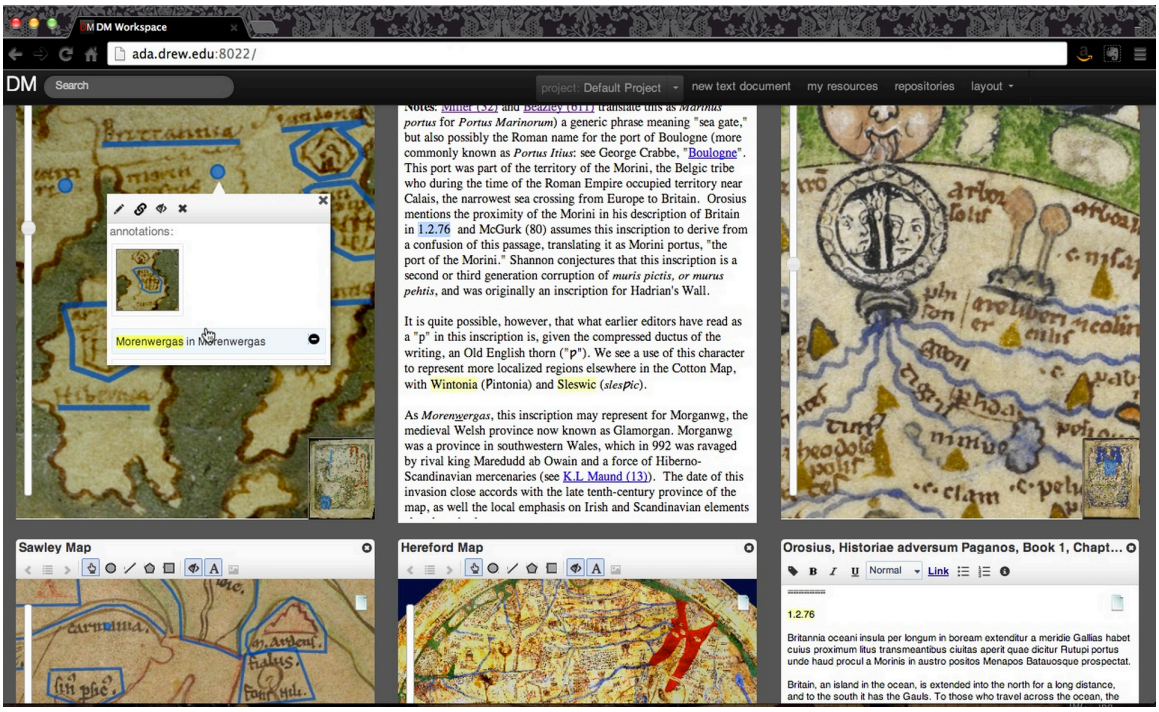
Additional DM work at University of Wisconsin-Madison

At the University of Wisconsin-Madison, implementation of the DM Project will also continue once *Phase 4* work is complete. On the strength of the work completed during the NEH grant period, co-director Foys has continuing internal funds to implement DM-related projects at UW-Madison. Part of this funding is planned to be used to update the already implemented feature of browsing and editing existing federated manifests of digital manuscript collections stored at Stanford University; this feature is no longer active because of changes to the protocol by which the federated manifests are accessed, and the need to integrate IIIF compatibility (see above). Part of this funding will also be used to finish the only unrealized development goal from our current NEH grant work, namely, to provide the ability for DM users to pull into a project any digital image publically available from a persistent Internet address for annotation and linking within project data. Additionally, the DM-Project has now also partnered with UW-Madison's Center for Digital and Print Culture, and has been awarded funding for a two-year Council on Library and Institutional Research (CLIR) post-doctoral fellowship,[10] where work will involve planning and implementing long-term data sustainability plans for DM projects, and creating/curating new, UW-Madison-specific content using DM, in addition to continuing work on expanding the British Library's *Virtual Mappa Project* and other DM work.

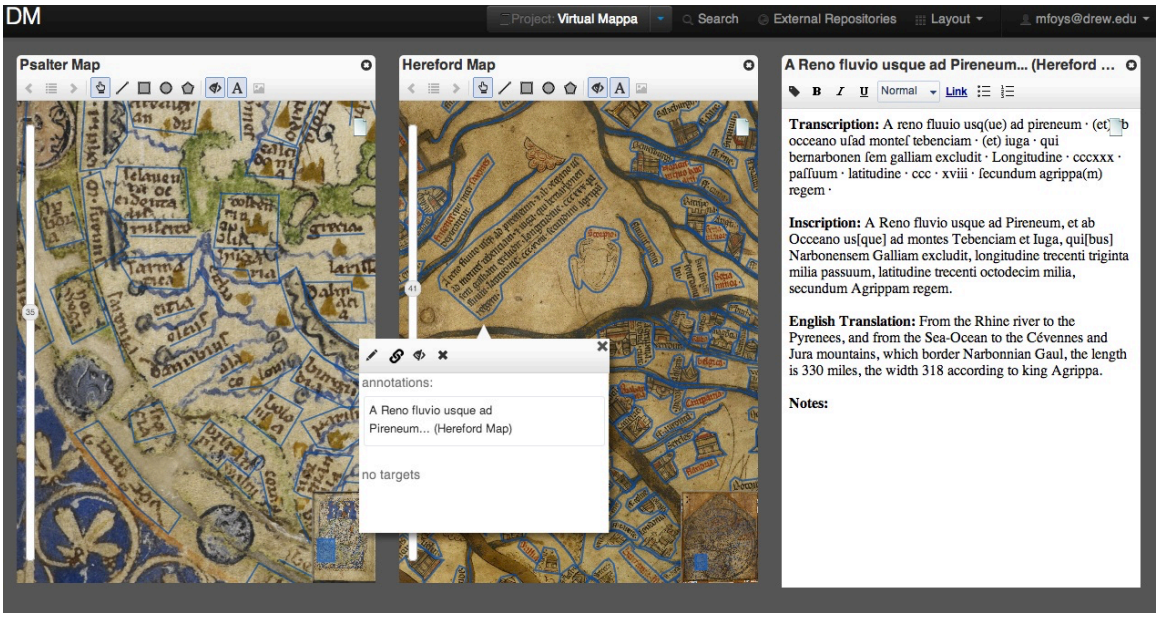
[9] http://iiif.io/
[10] http://www.clir.org/fellowships/postdoc/applicants/university-of-wisconsin-madison

Appendix A:
Screenshots of Current DM Environment and Projects

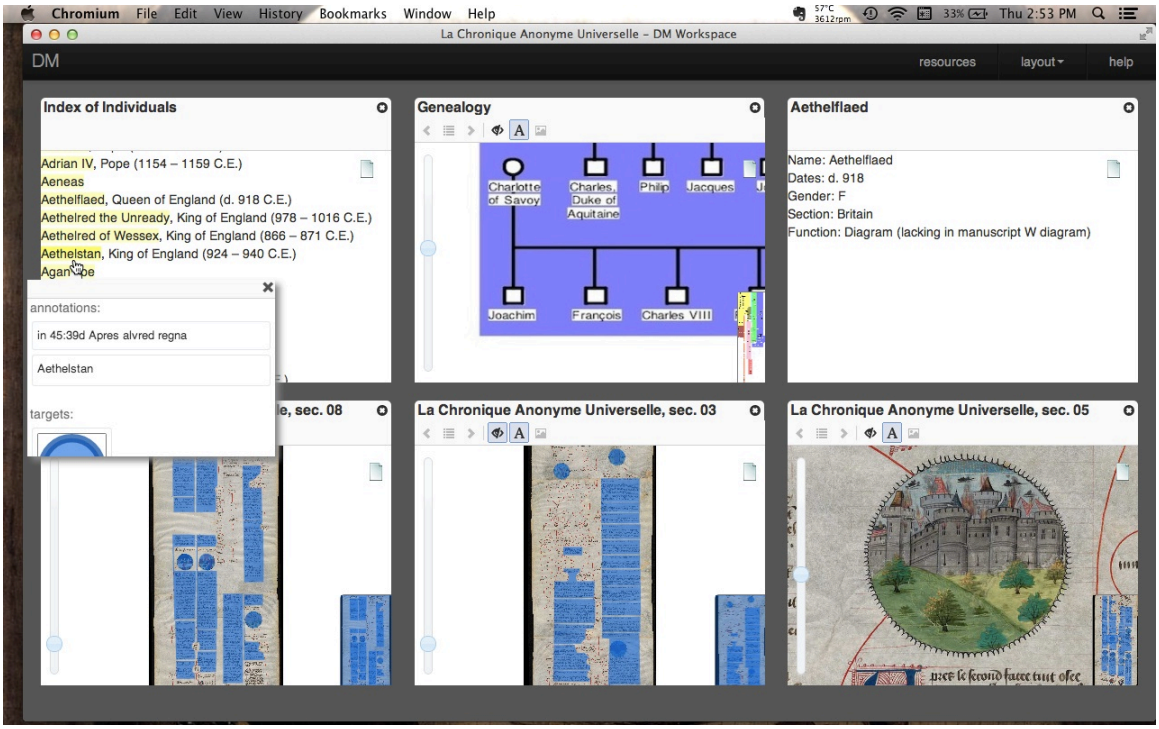
A1: Multi-up viewer showing several maps, texts and annotations from the British Library's *Virtual Mappa Project*

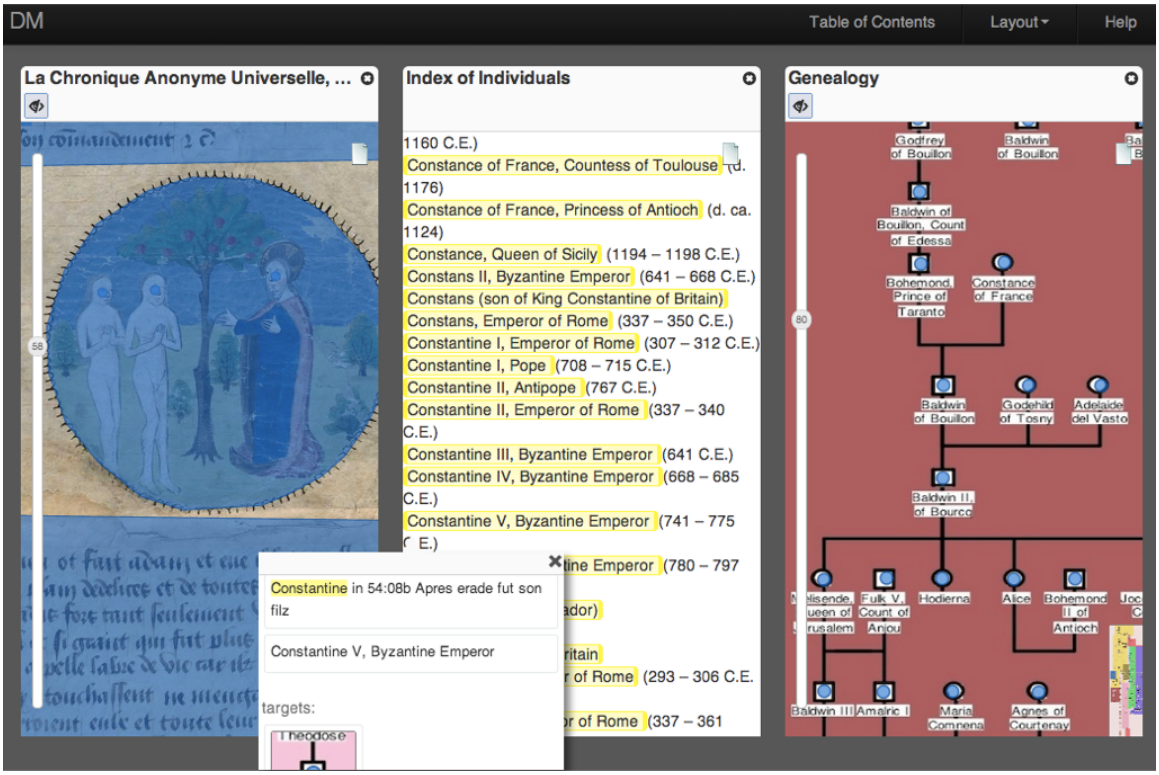


A2: *Virtual Mappa Project,* showing annotation of the large format *Hereford Mappamundi* (middle and right) and smaller Psalter Map:

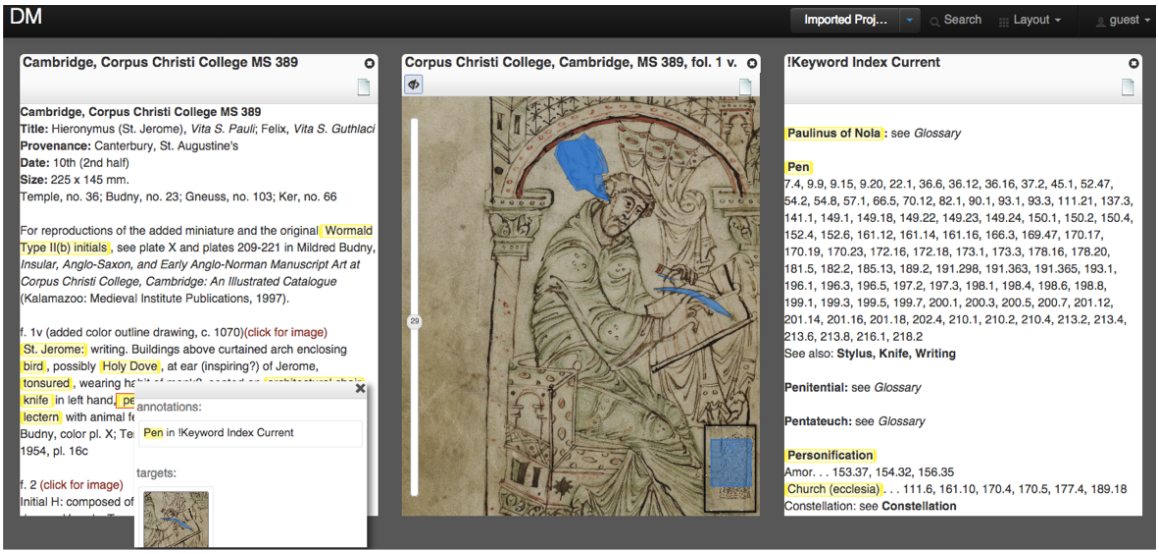


A3 & A4: Finished DM version of Lisa Fagin Davis's edition of *La Chronique Anonyme Universelle*, displaying from DVD:

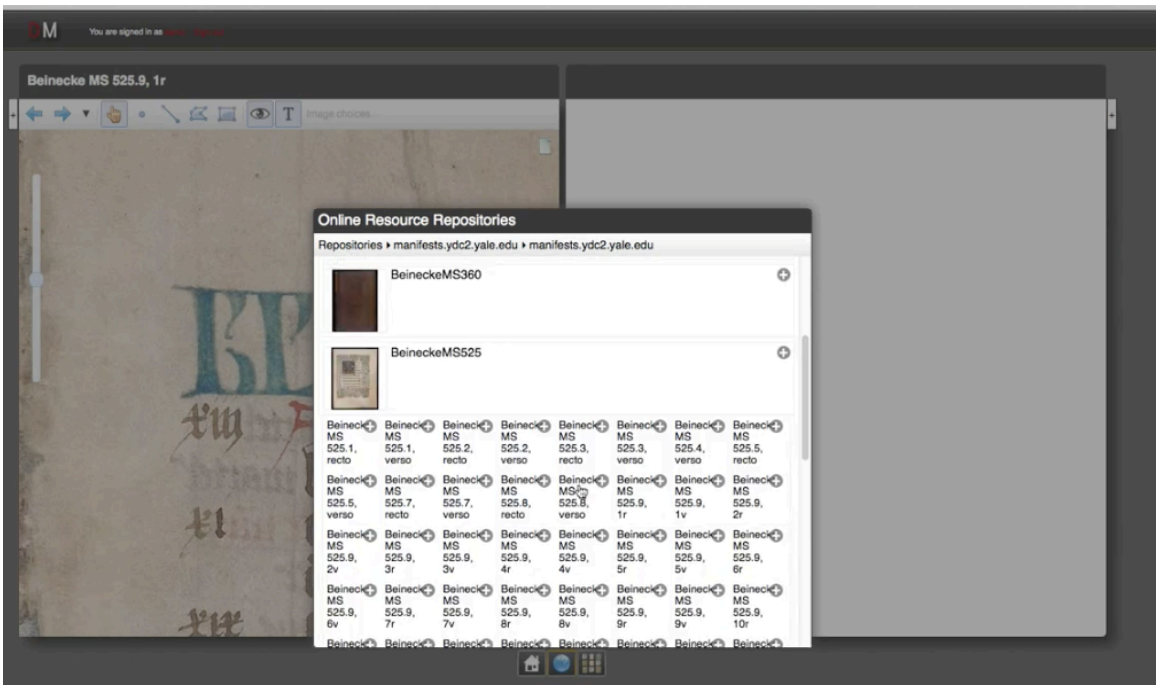




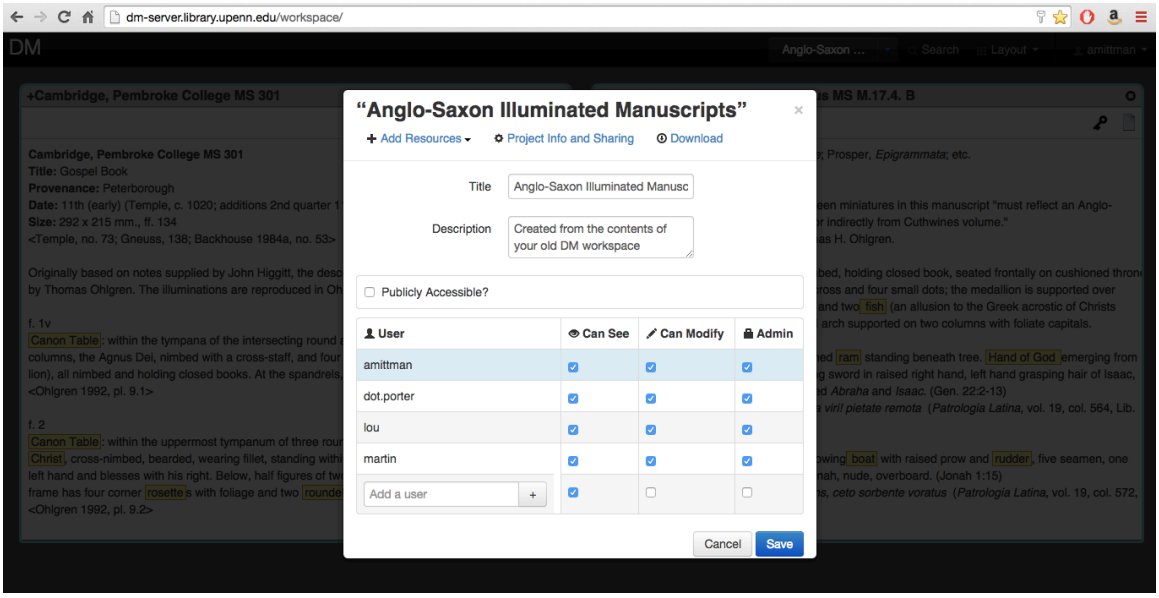
A5: DM environment for Asa Mitman's *Ohlgren's Insular and Anglo-Saxon Illuminated Manuscripts: An Iconographic Database*:



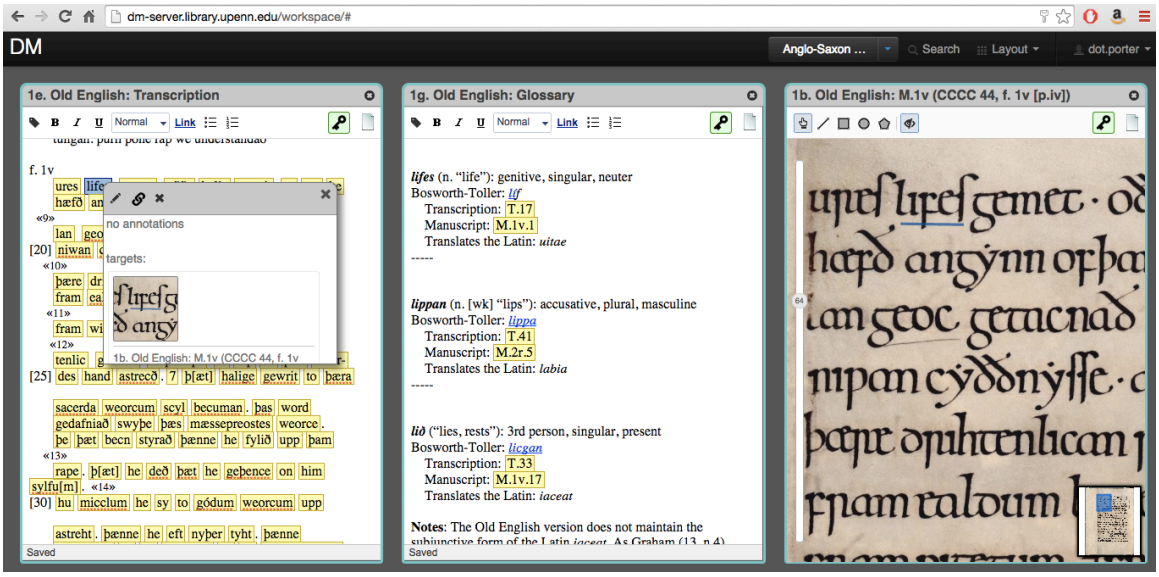
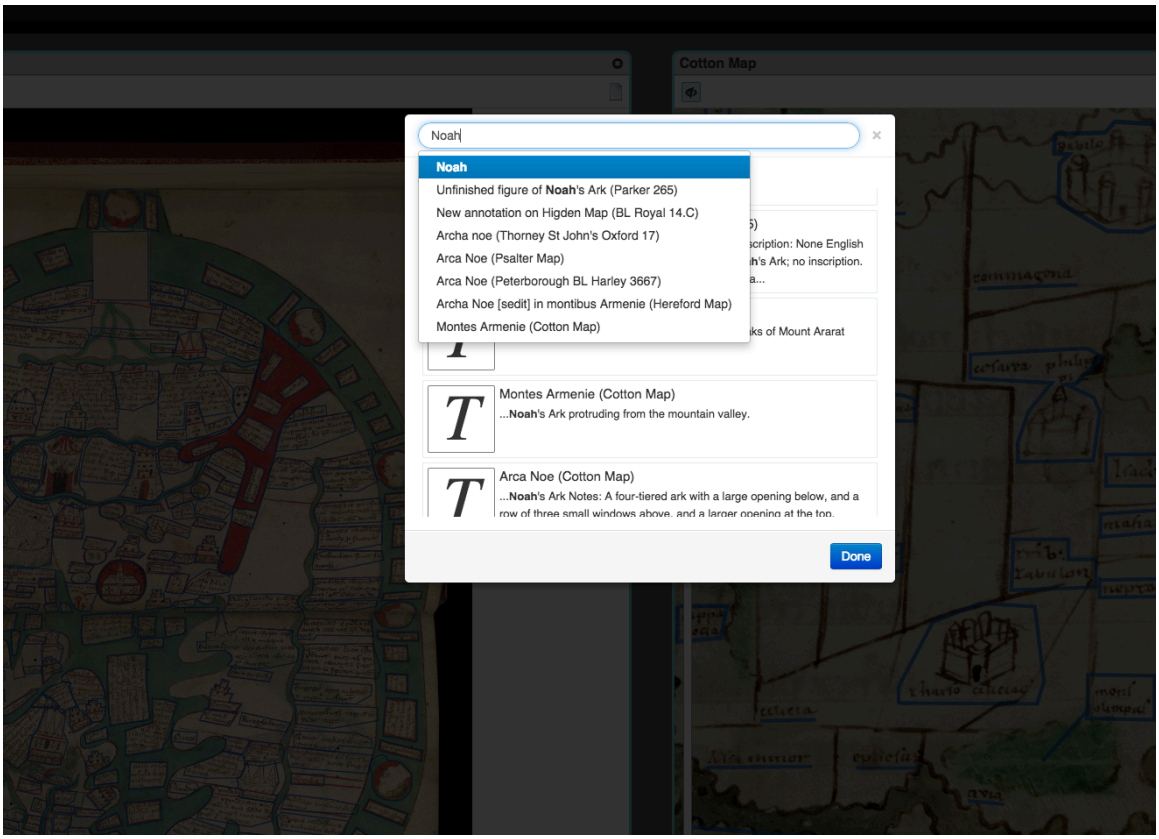
A6: DM resource for the access and annotation of manuscripts from online digital repositories, here showing the browsing and loading of manuscripts from Yale's Beinecke Library:



A7: DM environment for the *Insular and Anglo-Saxon Illuminated Manuscripts: An Iconographic Database;* showing the admin page for users and access:



A8: DM environment for the *Old English Bell-Tokens* project, showing textual editing capacities:



A9: DM search feature of annotations, with predicative returns: