Introduction

Sarah Sweeney, Digital Repository Manager at Northeastern University Libraries in Boston

I'm going to talk to you today about how we are using WordPress to publish our digital repository materials.

Who We Are

A bit of background before I start.

I am part of Northeastern University Libraries' Digital Scholarship Group, also known as DSG. We are an applied research group committed to helping faculty, staff, and student researchers through all stages of building a digital project.

The DSG was founded in 2014 as a response to movement at Northeastern to support the university's growing digital scholarship communities in the humanities and beyond.

The DSG supports and educates researchers at all levels on new digital scholarship techniques, including text analysis and encoding, GIS, data visualization, data management, repository services, copyright guidance, and semantic web services.

Problem Space

We believe digital curation is an essential part of our work and in the work of the digital humanities projects we support.

However, building digital curation and sustainability into DH project development is challenging, and engaging with DH scholars and researchers as full partners in curation practices is also difficult.

We wanted to figure out how we can position curation and sustainability as a shared responsibility between our library staff and project partners, so that it's not just a library or repository-only technical problem.

Solution

In response to these questions, we designed a series of tools and workflows meant to enable the library and its partners to work together to build sustainable digital projects, which we call the:

Community Enhanced Repository for Engaged Scholarship, also known as "series"

It's an interlocking set of systems and tools designed to:
- ease the burden of sustainable development
- foster stronger partnerships between library technical staff and research partners
- support community engagement with digital materials.

CERES is the larger project to support engagement.

What I'm going to talk about today is the Wordpress publishing tool, which is used to create and publish digital project materials, organized through a WordPress theme and associated plugins

---

Design Principles

The design of this system is based on a set of basic principles that fuel our project development process.

- First is that the library has the resources to build and support systems, not individual projects.
    - Our focus is on developing tools that can grow and support a wide variety of digital projects at Northeastern, and are built in such a way that the DSG can support them in the long term.
- Second is that DH research materials stored in the repository should be both curatable and usable, and the curation of these materials should be planned from the beginning.
    - Therefore researchers that want to use CERES must first build or plan a strong collection of data that meets our standards of long-term sustainability and open access principles.
- Third is that the expansion of the tool set, including
    - identifying
    - prioritizing
    - and testing new features
- Should be a shared research undertaking.

We expect the scholars we work with to understand that these tools and data standards are tied to scholarly outcomes and should be considered part of their research endeavor.

Design - Goals

We outlined four goals in support of this solution:

- Create a sustainable and shareable digital project environment for scholars.
- Build tools that could support many types of digital research materials.
- Start a program that would scale by encouraging project teams to become stewards of their own content and empowering them to customize and develop on their own.
- Increase preservation of Northeastern research by securing digital materials in our repository and web-based projects on library servers.

One of our goals here is to integrate two sets of needs that seem like they are in conflict:
- providing project teams with long-term curation of digital project materials in a repository
- and also enabling them to develop space to publish their projects on the web

Many of our researchers use popular web publishing tools such as WordPress or Omeka to store and display their materials because these tools are simple to use and easy to customize, but they are not ideal for long term file storage.

Others use the Digital Repository Service to store materials, which is ideal for long term storage, but can't be customized in support of individual project needs.

So we built an alternative tool that would satisfy the preservation and presentation needs of a project by joining repository storage with customizable web publishing tools.

■　■　■

Design - System

We met early in the design phase with representatives from a few DSG partners to discuss what they, as digital scholars, needed from a repository-integrated digital publishing service;

This early meeting was the first of many opportunities to engage with active scholars and researchers to learn about their needs.

We identified a few common needs from the group, including
- faceted searching
- visualizations
- file sharing
- downloading
- and image galleries,

which we used to inform our decisions about the initial CERES specification and project workflows.

Drupal, Omeka, and WordPress were considered as base platforms. However, none of the consulted partners or inaugural projects expressed a preference, so we selected WordPress.

WordPress is a well-known and broadly used open source blogging platform that natively supports many common web publishing features and can be customized using many different themes and plugins.

The CERES exhibit publisher, which is open source and available from these GitHub links, was developed by DSG staff by modifying the Quest WordPress theme to include commonly requested exhibit building tools and integrating it with a our custom WordPress plugin that connects it with the data source APIs.

Data Sources - DRS

CERES has two WordPress plugin data sources.

The primary data source is the library's Digital Repository Service (DRS), which is a homegrown Fedora / Hydra repository.

We encourage all DSG project teams to store their materials in the DRS, regardless of their participation in the CERES project.

We work with project teams to review their materials, and, if necessary, make recommendations to prepare the materials for long-term use and reuse by offering assistance in
- transferring files to curatable, open source formats
- describing materials using widely-used metadata standards
- and performing copyright and other rights evaluations

The DRS API is the connective tissue between the DRS and the CERES WordPress sites.

The DRS API is a JSON feed that passes essential information about each of the approximately 68,000 publicly available files in the DRS, including all displayable descriptive metadata, thumbnails, and download options.

To reduce noise and increase site speed, API queries for CERES sites are scoped to a particular DRS collection or custom set of DRS materials, which limits search and browse activity for the site to just the content relevant to a project.

Data Sources - DPLA

In addition to the DRS, CERES also incorporates materials from the Digital Public Library of America (DPLA), which gives site builders access to an additional 16 million digital resources from cultural heritage institutions across the United States.

Framework - General

The base implementation of a CERES wordpress site is designed to support many digital project site needs

Sites are issued to projects with a locally modified version of the WordPress Quest theme and a few prescribed plugins, including the DRS and DPLA data source API plugins and security plugins.

Sites are issued to projects with a default site framework, including a home page, about page, credits page, some with suggested elements.

It also includes standard header and footer design, as dictated by University Marketing and Communications guidelines.

Framework - File Browse

All CERES sites include Search and Browse pages, which use the DRS API to query the DRS for files.

The browse page is setup to retrieve all collections and files contained within the project's base DRS collection or set of materials, which display on the page in a grid, with facet options on the left to allow site visitors to limit the browsable items by creator, date, subject, or type.

Framework - Collections Browse

Projects with a lot of collections may also include a Collections page, which allows site visitors to browse and explore the collection hierarchy.

Framework - Search

The search page and search box are used to query the DRS API for project materials. When a user performs a keyword search in CERES, the search retrieves relevant project materials stored in the DRS, as well as WordPress pages that also contain the searched for keywords, like exhibit pages or blog posts.

The DRS results show up in the center of the page as the main results, and just like the browse feature, site visitors can limit the search results by creator, date, subject, or type.

The WordPress stored pages show up as results to the right of the main search results. Ideally these two result sets would be integrated, but we are limited to this functionality for now.

Framework - Credits

As I mentioned before some pages come with suggested elements. This is a sample of a Credits page template, which includes space for project teams to enter project staff names, declare copyright or other rights statements, and communicate how the staff would like their project to be cited.

Overall, this creates a "plug and play" experience for project teams. They don't need to set up server space, worry about data storage, or even go through complicated site design or wire frames. Project teams are encouraged to think about their site's design and customize heavily if they want to, but it's not required to move forward with data publishing, like it would be if they chose another platform.

Framework - Exhibit Tools

Project teams use the CERES plugin exhibit builder to create exhibits, pages, or blog posts that interweave DRS or DPLA materials with text to communicate the project team's work to site visitors.

- The exhibit builder uses a custom API query to request specific DRS or DPLA items based on keywords and search limits set by the exhibit designer.
- Each exhibit is customized using WordPress shortcodes, which are abridged strings of code representing exhibit settings.
- Exhibit designers know these shortcodes exist because they display on the page, but they do not build the shortcodes themselves.

Rather, the CERES plugin presents the designer with a menu of options that build the shortcode based on object and setting selections.

The user searches or browses the DRS and DPLA from this menu for their desired resources. The results are returned and thumbnails representing the files are displayed in the object selection interface so the exhibit designer can select which files to include in the exhibit and customize how they are displayed by selecting size, adding a caption, etc.

The first major CERES release included four exhibit building tools:

The single item exhibit is a simple method for inserting a single file thumbnail into a page, similar to a traditional website page layout for text with images.

The tile gallery exhibit embeds a block of file thumbnails for multiple files onto a page.

The slider gallery exhibit displays file thumbnails in a rotating slideshow.

The media playlist exhibit displays a series of audio or video files for streaming in a sequence prescribed by the exhibit designer. DRS videos can also be embedded directly in a page for streaming.

In addition to small improvements and bug fixes, the development for CERES' second major release focused on three new features:

The map exhibit embeds a map onto a page, on which all or some project materials with geographic metadata can be plotted.

Similar to the map exhibit, the timeline exhibit embeds a chronological timeline into a page, on which all or some project materials with date metadata can be plotted.

The image viewer embeds a high-powered IIIF image viewing tool into a page.

Workflow

Accompanying this tool infrastructure is a set of practices, policies, and roles that we believe are equally essential in establishing a culture of partnership and shared responsibility.

All DSG projects, including those that utilize the CERES tools, are asked to:
- submit a project proposal
- participate in an intake interview
- write a data management plan
- and participate in annual check-in meetings

All four activities are formally tracked and documented to produce a project history and ensure continuity over time

Both the project proposal and the intake interview serve to teach us more about the project work, including the history of the project, the size of the project staff, and project goals.

The intake interview is an opportunity to discuss the scope of our involvement in the project and set expectations, particularly around labor and development timelines.

Project teams requesting a CERES site must also have a clear project lead and a plan for designing the site, either with their own or hired labor.

We also have new projects write a data management plan by asking the team to answer a series of questions about their data and digital materials.

This prompts project teams to think deeply about their project materials in ways they may not have otherwise considered and gives us an opportunity to learn about how they are managing their research materials so we can make informed recommendations to improve their data management practices.

The annual check-in meeting is an opportunity to update basic information about a project, like new deadlines or staff changes. These meetings allow us to discuss any new or revised project goals and help inform our own internal decisions about infrastructure or other development efforts.

These four project activities facilitate a shared understanding of project details and needs, which helps to ensure smooth implementation of all the CERES workflow stages, from content review to site launch.

It does not guarantee that the project will run smoothly or linearly (workflow stages are often revisited or revised with new project information), but it does prevent some of the more common project pitfalls, like surprise deadlines.

CERES Projects

More than twenty CERES sites have been built since the project launched in 2015.

Faculty, staff, and students with faculty sponsors may request a new CERES site at any time throughout the year. Project teams requesting additional feature or tool development must submit a proposal during the two-month Call for Proposals period.

The Call for Proposals process is an opportunity for us to promote the service widely to the Northeastern community and to learn about what tools are needed to properly exhibit research at Northeastern.

Obviously I don't have time to look at all these projects, but you can visit the URL here to browse the publicly available sites yourself. I'm going to talk about three of the projects that we think exemplify the CERES project.

The Early Black Boston Digital Almanac

In 2016, a professor in the department of English designed the curriculum for an undergraduate experiential education course around creating an almanac detailing the lives of African Americans in the city of Boston from the 17th century to today.

In her proposal, the professor suggested a collaboration with the DSG to use CERES as a framework to create and publish an almanac of digital exhibits describing early African American life in Boston and New England.

This framework would allow students to demonstrate their research using a mix of traditional essays and web writing as they explored the experiences of African American lives throughout Boston's history.

DSG staff were involved in many elements of the course and we were given the opportunity to insert lessons on exhibit design into the syllabus, as well as specific training on CERES itself. DSG staff were present for many sessions to introduce students to best practices for website and exhibit design, and to provide guidance on building exhibits using CERES, selecting appropriate materials and resources, and writing for the web.

We also held hands-on sessions designed to give students the opportunity to ask questions about using CERES to build their exhibits.

We used this course collaboration to learn how CERES could be used as a framework for digital humanities projects in a pedagogical setting.

One of the long-lasting outcomes of this collaboration was a rough sketch of a reproducible plan for using CERES in the classroom.

We have built on this preliminary curriculum to develop a set of resources for using CERES as a pedagogical tool, including a suggested course schedule and activities, lecture topics, sample assignment prompts, and helpful resources for students and faculty.

A Proud Past: Boston-Bouvé College, 1913-1981

Northeastern University Archives and Special Collections has a decade-long history of designing and building online exhibits to showcase its digital collections.

As more and more archival collections were digitized and stored in the DRS, Archives staff expressed a need to unify exhibit building workflows, upgrade legacy exhibits, and simplify maintenance for all online exhibits.

During the CERES pilot phase, Archives submitted a proposal to test CERES as a tool for streamlining their exhibit building process, which lead to the development of A Proud Past: Boston-Bouvé College, 1913-1981.

The exhibit, which uses digitized archival photos and documents to illustrate the history of the school from its founding in 1913 to its merger with Northeastern University and its eventual rebranding to become the Boston-Bouvé College of Health Sciences, was successfully reconstructed using CERES and quickly inspired Archives staff to replace several of the legacy exhibits with new CERES sites.

Since the 2015 pilot, Archives has created 10 sites, which is a significant portion of the total number of CERES sites hosted by the DSG and that number continues to grow. We expect Archives to create more than twenty exhibit sites over the next few years.

The quick adoption of CERES by Archives firmly tested the scalability of CERES technology and workflow with satisfying results, as infrastructure, staffing, and maintenance was not adversely affected by the increase in use.

This proved the original conceit of CERES, that exhibit design control can be placed in the hands of the subject expert while the infrastructure can be firmly controlled by the technical experts, creating a mutually beneficial partnership of shared responsibility.

The Catskills Institute

The Catskills Institute is an organization that was founded to promote research and education on the significance of the Catskill Mountains for Jewish-American life run by Phil Brown, a Professor of Sociology and Health Sciences at Northeastern.

On behalf of the institute, Brown has collected thousands of images, postcards, programs, and other artifacts representing Jewish-American life in the Catskills Mountains.

The institute worked with Brown University's Center for Digital Scholarship to describe and digitize more than 1,300 of the items in the institute's collection, as well as create a website to widely share the institute's work.

When Brown and The Catskills Institute became affiliated with Northeastern in 2012, the DSG offered to house the digitized materials in the DRS and modernize the research website.

The commitment needed to support the institute's work using the CERES infrastructure was significant and, as with many projects, setting expectations and allocating labor were major challenges.

The DSG and the Catskills Institute formed a relationship before the CERES infrastructure was developed, therefore it was difficult to accurately estimate the time it would take to fully realize the modernized Catskills Institute website.

Another challenge was the shared labor responsibilities for this project, as neither Brown nor the DSG had additional staff to dedicate to designing and building the new Catskills Institute CERES site.

The DSG relied on interns from Simmons College and a Northeastern work-study student to revise the cataloging, design the site, and build site pages and exhibits.

This project helped us to refine our project workflows and reinforced the value of developing a partnership of shared responsibility, rather than simply providing CERES as a service.

Outcomes

The three principles I mentioned earlier have served as important motivating and scoping devices for the strategic design of CERES.

By placing the digital repository at the foundation of our support for digital scholarship, the library and DSG have expressed a commitment to long-term curation of that scholarship, which empowers scholars while also setting clear parameters for our support.

The key outcomes from the project so far have been both practical and cultural. In practical terms, after two full development cycles, we have begun to see concrete payoffs from the early investment in repository infrastructure.

The design of the CERES WordPress interface allows project teams to focus on their scholarly work rather than a detailed technical implementation, which broadens opportunities for substantive participation.

The use of a shared platform also means that all skills acquired on one CERES project are transferable to other CERES projects, making it possible to develop a common set of training materials and a strong pool of highly-trained student collaborators.

Next Steps

Looking forward, the next phase for CERES will focus on improvement and refinement.

We will engage with current and future CERES exhibit designers to gather feedback about the design and usability of the tools, which we will use to prioritize and develop enhancements.

We also will start planning our next major set of requested features, including exhibit tools for displaying annotations, transcriptions, and translations.

A non-trivial dependency for these features will be the ability to store and display the annotation, transcription, and translation data, and we are actively developing a repository system to support that activity.

We will also continue to research strategies for technical preservation of the sites themselves, including the text, dynamic resources, and site experience.

Ultimately, CERES increases the visibility and security of digital scholarship projects and lowers the barrier for digital object storage and website design.

Using CERES ensures that many goals are achieved
- project teams are able to achieve their goal of contextualizing, publishing, and distributing their work openly on the web to a broad audience
- the DSG achieves its goal of connecting researchers to digital tools and services that improve their workflow and advance their research
- and the library achieves its goal of securing Northeastern's intellectual output by storing materials in our digital repository and securing web resources produced by researchers.