Customizing DSpace: Developing Private Communities & Collections for Future Repository Integrations
Susan Parham
Open Repositories 2022

In this presentation, we’ll describe the collaboration between Georgia Tech Library and the DSpace service provider Atmire to deploy code customization in preparation for future repository integrations.

This work arose from efforts at Georgia Tech to bridge the traditionally separate worlds of the institutional repository and the archives. Our goal is to overcome barriers that lead to siloed content, workflows, and systems.

Through strategic planning, we developed our digital curation roadmap, with the long-term ambition of employing a loosely integrated ecosystem using DSpace, ArchivesSpace, and Archivematica.

We are not the first institution to embrace a version of this particular ecosystem. We are indebted to our peers at the University of Michigan Bentley Historical Library whose Mellon-funded ArchivesSpace-Archivematica-DSpace Workflow Integration project (2014-2016) led the way in this area. The University of Edinburgh has also looked at this integration.

With the creation of this roadmap, we’ve come to a shared organizational understanding of our curation infrastructure as an ecosystem of components rather than a single, monolithic product - we utilize a constellation of community-based and in many cases, vendor-supported tools. (While the ecosystem concept is not new to this crowd, it was important for us to come to this shared understanding as an organization)

We engaged Atmire to help us realize our first-year goals for the multi-stage project, starting with code customization to provide greater granularity over access permissions, allowing us to merge two locally hosted DSpace repositories.

These two repositories were long-standing, locally hosted DSpace instances used for quite different purposes: our open access institutional repository and a dark archive of both unprocessed born-digital materials and preservation copies of digitized files.

Because DSpace is inherently designed for providing open access, we required code customization that would allow greater access control over the unprocessed or otherwise restricted archival materials that we planned to migrate.

We wanted the ability to easily create communities and collections that are not only hidden from all potential discovery points, but also restricted by direct link.

The code customization allows us to create a resource policy that flags a community or collection as private and only viewable by a designated group. These flagged collections are hidden from general users in the DSpace discovery interface, OAI, etc. and via direct link.

In addition, the customization supports the cascading of collection-level permission changes down through all collection items with a single click.
Georgia Tech has a long history of hosting and (in the past) customizing DSpace locally and wanted to maintain a level of local administrative control, while rebuilding our technical institutional knowledge of DSpace for sustainability efforts.

To address this dual purpose of accomplishing the tasks at hand while building our own local knowledge base, we collaborated with Atmire to develop a phased deployment process using multiple instances of DSpace 6.3:

- An Atmire development instance for demos and code change review
- A Georgia Tech stage instance to test deployment methods and
- Our production repository, all on locally managed servers.

We utilized both virtual and physical machines running RHEL 7, with the physical host running multiple instances of DSpace that shared a large asset store.

With this multi-stage deployment strategy, our local GT technologist Eli Patterson deployed code from the tested and approved Atmire instance to our local stage repository. Once these changes were approved on stage, he then deployed those changes to production. Mark Diggory from Atmire provided consultation during the process.

What were the benefits of this approach?

We have completed the initial phase of the project and are one step closer to our vision of a curation ecosystem that supports both the IR and archival digital collections.

Having a single DSpace repository helps us bridge these worlds as we prepare for a major repository re-conceptualization with DSpace 7.

By partnering with Atmire, we’ve made use of their deep expertise and position within the DSpace community – we don’t have to try to reinvent the wheel locally. The partnership greatly enhanced our timeline and established norms for future collaboration. In addition, because Atmire is a certified DSpace partner and committer, we feel this work helps support broader DSpace sustainability.

This deployment strategy provided a learning opportunity for Georgia Tech technologists and repository managers alike. The staged approach forced the creation of more explicit documentation on our part and led to a better understanding of our technical environment. This strategy also made it easier for us to clear the project with Institute cybersecurity.

What challenges did we face?

The multi-stage deployment strategy added to the project timeline and having to sync various versions of the repository database & asset store introduced complexity and the potential for error. In addition, working with Atmire put us in the situation of trying to ramp up local technology skills rapidly – our new technologist had to learn DSpace and set up our local environment for vendor use simultaneously. By nature, the pacing of work between Atmire and Georgia Tech was at times at odds. As a company with a dedicated team of experts, Atmire could move much more quickly than Georgia Tech – even the state procurement process for this work was difficult and delayed project start. Repository managers had to approve multiple rounds of testing and customer service was impacted when he had to freeze repository deposits to sync asset stores during the deployment process. And of course, we now have the cost of maintaining this customization going forward.
That said, we’re on our way for a major re-conceptualization of the Georgia Tech digital repository using DSpace 7 ...

... and are ready to begin work with Atmire to develop crosswalks between ArchivesSpace and DSpace.