Open Repositories 2014   Authors: Robin Ruggaber, the University of Virginia
Conference Paper, Main Track   Rick Johnson, the University of Notre Dame

Hydramata: Building a Nimble Solution with Hydra to Transcend the Institutional Repository

Abstract

We are all facing rapidly changing repository demands that call for maximizing flexibility and interoperability among a heterogeneous network of technologies, varied workflows and wide array of evolving formats that are pushing the limits of our Institutional Repository technology. We need a repository to support research data, large image collections, articles, digital exhibits, video, and data visualizations. We live in an environment of decreased resources, increased complexity in how scholarly work is produced, stored and disseminated. Join us for an update on a project undertaken by six Hydra institutions with varied legacy repository solutions in order to build a flexible, sustainable framework to meet the diverse needs of the institutions. We will share how the technology, the resourcing strategy and the project methodologies are meeting prioritized needs while reducing risks and increasing efficiency. If you are a small, medium, or large institution, a repository manager, developer, librarian, digital library technologist, director, or anyone else struggling with similar challenges, then this presentation is for you.

The Challenge

Put simply, we can no longer meet the challenges we face for repository services alone or afford providing boutique tailored solutions to solve individual problems. Our needs for research data, image collections, open access content, theses and dissertations, video, and other large collections were not being solved by current solutions and are pushing the limits of our definitions of an Institutional Repository. As the complexity of content, how it relates and how it is accessed increases while resources in the best case remain static or in the more common case are trending downward, the need to maximize efficiencies is critical.

We are six academic institutions[1] who each had open source that were limited, consortial solutions that were dissolving or built solutions using some commonly developed components and were ready to take the next step whether to upgrade or expand our services. However each legacy solution was offered or built as a stand alone application with a minimal or tailored approach to serve a specific need - institution repositories for highly tailored workflows, repositories of specific formats or single disciplines, grant funded applications for narrowly focused content or for disseminating content to a specific audience. We had largely sourced these efforts in house and due to decreasing budgets and the rich talent of our staff had often been the target of having our staff raided by private industry.

---

[1] Indiana University, Northwestern University, the University of Cincinnati, the University of Notre Dame, the University of Virginia, Data Curation Experts

To our advantage is the rising demand for our services and the number of opportunities to offer additional services. Mandates for how research data is managed and persisted, support for interdisciplinary scholarship, preservation of content that is rich in format and large in size or quantity and the expansion of digital scholarship is largely driving this demand. This requires a focus not only on preservation, but also plugging into network systems like DOI, and ORCID. Therefore, our solutions also need to be conduits to and from the larger ecosystem of scholarly information.

Shared Goals

Our shared solutions had to offer the ability to divide and conquer. By this we mean that solutions had to be delivered in a way that they could be leveraged for multiple purposes, work in concert in a robust way and be easy to maintain and evolve. The emerging landscape would require more attention to solving name ambiguity, object ambiguity, adherence to emerging standards to support interoperability, large data and expanded features for legacy content such as ETDs. We will share the specific common needs across our institutions which largely drove us to not reinvent the wheel alone. We will also share how the way we work resulted in quick delivery of functions and features that could then be tested, validated and leveraged by early adopters while still building out functions to meet the next major revision of legacy services.

Resourcing Strategy - Realizing Efficiencies through Collaboration

All of us are part of the Hydra Community which is based on the core belief that the sum is greater than its parts. For this effort, we realized that by combining the talent of our various institutions we could achieve solutions that met our individual needs and our shared goals. Combining talent also benefited individuals in that they are no longer alone in support of an application but have a network of colleagues with diverse expertise and perspectives. New staff ramp up more quickly by working with more experienced staff and both have realized benefits. So how does this impact sustainability? We will share examples of how progress on shared goals continues when staff are diverted to other projects or leave our employment. We are actually decreasing the need for local cross training and succession planning due to leveraging combined resources.

Technical Strategy - Laser Beam Attention to Flexibility and Reuse

We realized we could leverage both the talent and resources at our partner universities to truly make our digital platform sustainable, while not sacrificing sustainability for innovation. We are in agreement about the kind of solution we need, however each institution still has specific requirements that require accommodating many use cases at once. These include complex workflows and a variety of advanced services. Also, because content increasingly involves a mixture of formats, it is imperative that we support multiple formats coexisting within single datasets or collections. Finally, it was clear we need to support the wide spectrum of use cases of our campus faculty, students, and other researchers ranging from research data, image collections, theses and dissertations, video, and library owned collections.

In the past, the Hydra framework has allowed us to build applications quickly, but once created

it has been difficult to share our work across institutions because of divergent workflows, content models, and dependent services like image and streaming servers. As a result, to really sustain each other it became clear that we needed an architecture that was flexible, scalable, and modular to be tailored for many different institutions; provide a full working solution out of the box; and be an extensible platform for future innovations.

Within the Hydra community, we also realized long ago that it is difficult to tailor a single application to meet all the needs of every audience or use case.   This is juxtaposed with our need for a common discovery environment for our users, where content of any format can be discovered alongside any other format and allows inline interaction with materials like image viewing, streaming video, or data viewing. Therefore, it was clear we need to accommodate both a unified environment and parallel environments focused on data, video, exhibits, or other specific disciplines.

All of these needs were realized in the Hydramata architecture with individual Hydra-based automata, "Hydramata", that will both be used independently to serve a single content type or service and be assembled together to form a unified environment for management and discovery of any kind of scholarly work and collections. This unified environment, shown in Figure 1, provides an out of the box solution for our urgent use cases that can be tailored if necessary.
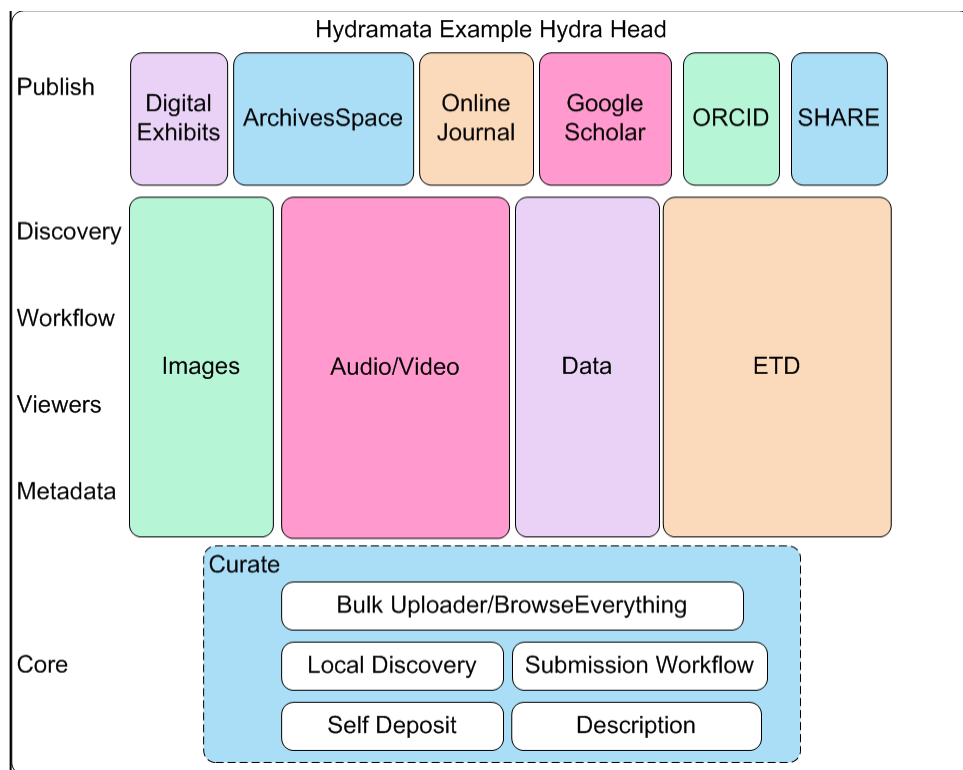

Figure 1

Each individual module, "Hydramaton", acts as a plug-in to the unified discovery and

management tool that can be swapped in and out to meet the needs of a particular institution. Examples include content focused Hydramata like Images, Data, ETDs, and external network and environment facing like Digital Exhibit publishing, ORCiD integration, SHARE, and Google Scholar. A Hydramaton also serves the dual purpose of being the baseline for a use case specific web portal that allows content to coexist in both the specialized and unified environments. Additionally, a Hydramaton can be made up of other smaller Hydramata focused on more discrete tasks like deposit, ingest, discover, view, and disseminate that can be swapped out accordingly to serve institution specific workflows or other dependent services as shown in Figure 2 below.
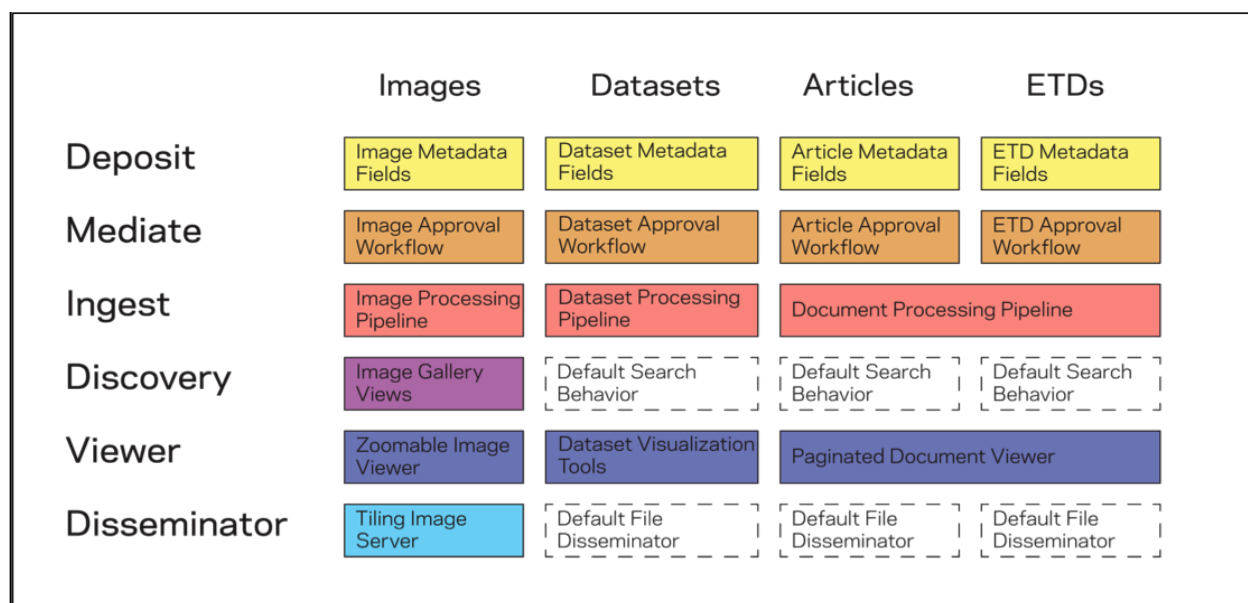

Figure 2

Ecosystem – what we have, what's coming and does it play well with others?
Our project is complex and continuing to produce results. We will demo what we have produced to date, what is on the horizon and how it relates to other technologies within our wider ecosystem.

Conclusion

 This way of working and what we are trying to achieve together is unique and evolving within our community and experiences. We aim to share both our challenges and our successes. We also want to learn what others think about our approach, have similar experiences and how that worked for them, or have interest in supporting our efforts or leveraging our solutions.