



11<sup>th</sup> International Conference on Open Repositories  
Trinity College, Dublin, Ireland  
13th - 16th June, 2016

## VTechData: An Institutional Data Repository

Zhiwu Xie, Julie Speer, Yinlin Chen, Tingting Jiang, Collin Brittle, and Paul Mather  
University Libraries, Virginia Tech  
[zhiwuxie, jspeer, ylchen, virjtt03, rotated8, pmather]@vt.edu

### Session Type (select one)

- Panel
- Presentation

### Abstract

We introduce VTechData, a Sufia/Fedora based institutional repository specifically implemented to meet the needs of research data management at Virginia Tech. Despite the rapid maturity of Hydra and Fedora code bases, the gaps between the released packages and a launched production-level service are still many and far from trivial. In this presentation we describe the strategy and efforts through which these gaps were filled and lessons learned in the process of creating our first Hydra/Sufia-based repository.

### Conference Themes

Select the conference theme(s) your proposal best addresses:

- Supporting Open Scholarship, Open Data, and Open Science
- Repositories and Cultural Heritage
- Repositories of high volume and/or complex data and collections
- Managing Research Data, Software, and Workflows
- Integrating with the Wider Web and External Systems
- Exploring Metrics, Assessment, and Impact
- Managing Rights
- Developing and Training Staff

### Keywords

Hydra, Sufia, Fedora, institutional repository, data management.

### Audience

Repository managers, developers, librarians.



11<sup>th</sup> International Conference on Open Repositories  
Trinity College, Dublin, Ireland  
13th - 16th June, 2016

## Background

In 2011, Virginia Tech Libraries began offering research data management services to university researchers in the form of data consulting services: assisting researchers in creating data management plans to comply with funding agency requirements and understanding and applying data management best practices. Four years later, we applied our knowledge gained through consulting and campus research environmental assessment efforts, to begin designing and developing an institutional research data repository for long term preservation and access. Research data have specific characteristics and requirements that call for different treatments and strategies for preservation and dissemination. Presenters will share their experience building a data repository based on Sufia and Fedora.

## Presentation content

We first describe our strategy to deploy developer and IT resources towards implementing and deploying VTechData<sup>1</sup>, Virginia Tech's institutional data repository based on Hydra/Sufia and Fedora 4. This includes:

- Position the data repository in the long-term digital library strategy and road map. Why do we develop a separate data repository and how does it relate to the existing IR and other library systems?
- Assess the status of community developments. What were the statuses of various Hydra/Fedora community developments when we started, how did they progress, and how did these changes impact our strategy and implementations?
- Phase the launch to balance short-term needs and long-term aspirations. What are the phases, why, and how did these phases interleave with other technology developments in the library?
- Manage the expectations from senior management, researchers/end users, and library colleagues. What have we promised to deliver and how did we reach the consensus?
- Prioritize local software developments. How did we manage and maintain many different library systems and also focus resources on new developments?
- Skill preparation through contributing to relevant community projects. What skills are required, what are the main skill gaps, and how did we bridge these gaps?

---

<sup>1</sup> <http://data.lib.vt.edu>



11<sup>th</sup> International Conference on Open Repositories  
Trinity College, Dublin, Ireland  
13th - 16th June, 2016

- Separate IT resources for new developments from those for production-level services, Why and how did we move the new developments to AWS?
- Consensus building among all stakeholders on specific goals and deliverable. How did we set goals, requirements, and development cycles.

We then describe the specific development efforts in this local adaptation of Sufia/Fedora, including:

- Usability evaluation and dashboard redesign;
- CAS integration, allowing users to sign in with their university credentials as well as name authority;
- ORCID integration for linking to researcher profiles;
- EZID integration for creating DOIs;
- Embedding DataCite search, CrossRef search, and VT People search in web forms to reduce manual input when adding content;
- Deployment and DevOps improvements from bash scripts to vagrant to Ansible playbooks.

The future work of the repository includes:

- More flexible metadata schema through PCDM;
- Integration with our institution's mass storage facility;
- Interface with our institution's High Performance Computing resources;
- Expand the data repository to include High Throughput Computing resources, e.g., Science DMZ, 10G data kiosks.

## **Conclusion**

Implementing a research data repository should be more than simply re-purposing a textual document repository. We share the experience of the first phase of Virginia Tech's institutional data repository and offer the lessons learned.