

VTechData: An Institutional Data Repository

Zhiwu Xie, Julie Speer, Yinlin Chen, Tingting Jiang, Collin Brittle, and Paul Mather
University Libraries, Virginia Tech, Blacksburg, VA, USA

Background

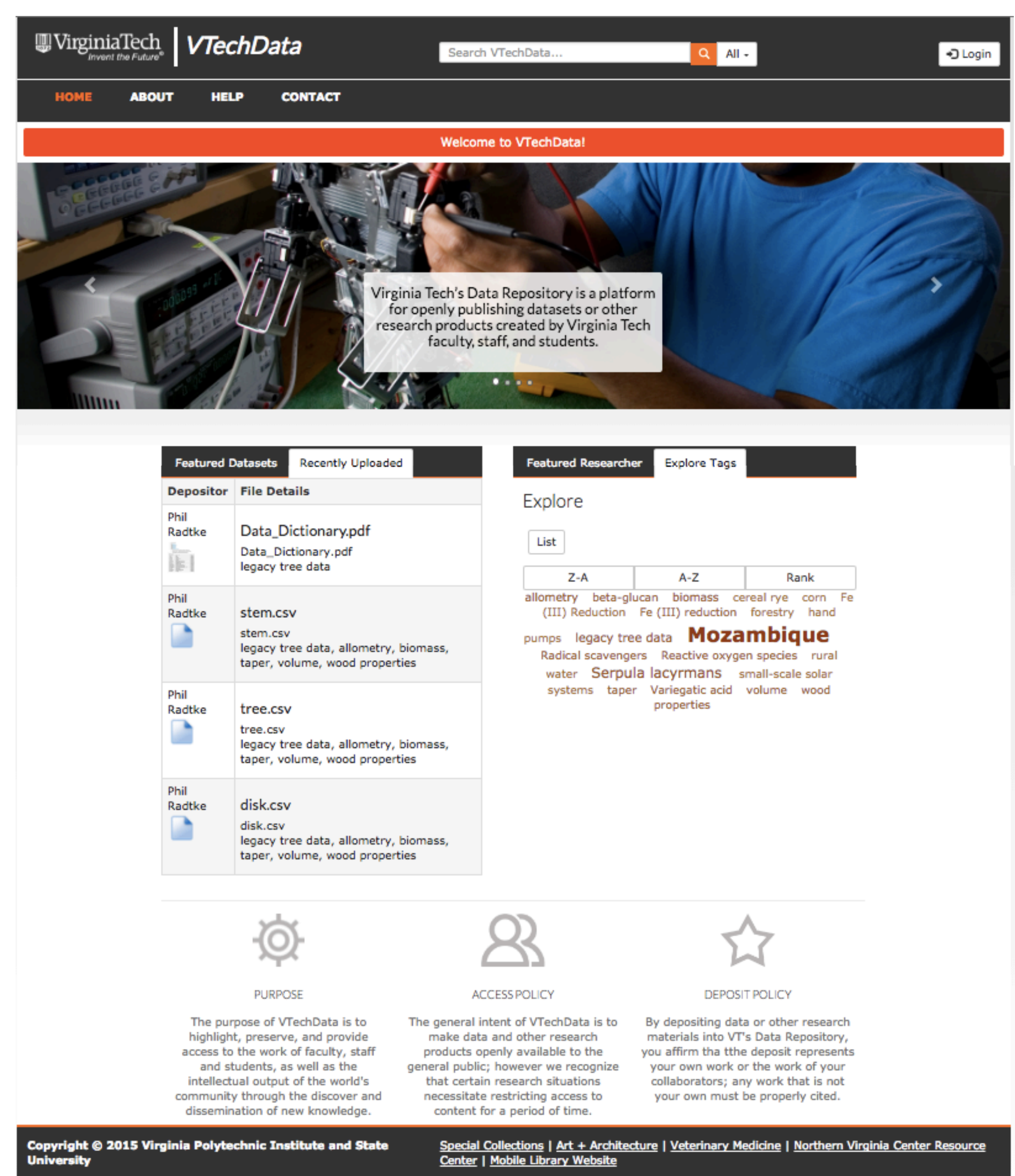
In 2011, Virginia Tech Libraries began offering research data management services to university researchers in the form of data consulting services: assisting researchers in creating data management plans to comply with funding agency requirements and understanding and applying data management best practices. 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.

VTechData Strategy

- Position the data repository in the long-term digital library strategy and road map.
- Assess the status of community developments.
- Phase the launch to balance short-term needs and long-term aspirations.
- Manage the expectations from senior management, researchers/end users, and library colleagues.
- Prioritize local software developments.
- Skill preparation through contributing to relevant community projects.
- Separate IT resources for new developments from those for production-level services.
- Consensus building among all stakeholders on specific goals and deliverable.

VTechData Development

- Based on Fedora/Sufia
- 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.



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

Timeline

- First prototype hackathon Nov 2014
- Formalize requirements and dev sprints Spring 2015
- Sneak peak demos June 2015
- Main feature demos August 2015
- Phase 1 soft launched August 2015
- Phase 1 formal launched Feb 2016
- Maintenance and tuning continues and ongoing
- Phase 2 with PCDM to be completed by 2016

Future Work

- Integration with institutional mass storage facility;
- Interface with institutional High Performance Computing resources;
- Expand the data repository to include High Throughput Computing resources, e.g., Science DMZ, 10G data kiosks.