

# Persisting complex, hierarchical repository content in an S3 object store

Anusha Ranganathan  
Partner, Cottage Labs  
Email: [anusha@cottagelabs.com](mailto:anusha@cottagelabs.com)

Paul Walk  
Director & Founder, Antleaf  
Email: [paul@paulwalk.net](mailto:paul@paulwalk.net)



cottage labs



# Context

- 4 year development project with Ruhr University Bochum (Germany)
- Develop a Research Data Management System - "ReSeeD"
- Based on:
  - Samvera Hyrax (currently v3, moving to v5 this year)
  - Two large, private S3-based data object stores for storing the research data
- Developed for general use at the university, but also tailored development for particular research groups with particular metadata models (e.g. CRC 1280)
  - Samvera Hyrax is flexible enough to cater for this

# Requirements

Three key requirements have, taken together, led to some interesting challenges:

1. RDMS must support multiple, complex data/metadata models
2. RDMS must preserve data privacy within each research group, until it is ready for publishing
3. RDMS must persist all data & metadata in RUB's S3 object storage facility

# Demonstration

# Users and their role

|                            |   |
|----------------------------|---|
| user2@hyrax                | CRC manager for <a href="#">CRC 1280</a>                              |
| user3 to user7             | Members of <a href="#">CRC 1280</a>                                   |
| user3@hyrax                | CRC group manager for <a href="#">A12</a> in <a href="#">CRC 1280</a> |
| user4@hyrax<br>user5@hyrax | CRC group members of <a href="#">A12</a> in <a href="#">CRC 1280</a>  |
|                            |   |

# Challenges

1. Accommodating data files from a prescribed, hierarchical filesystem in a web-based repository
2. How to provide meaningful search interfaces to users when supporting different metadata models for different "work types"
3. Ensure tiered rights are coherently carried across all functionality available in Hyrax and presenting it to the user in an intuitive fashion
4. Persisting complex and heterogeneous data and metadata in RUB's S3 object store and allowing for edits, and updates

All of the above is increasingly challenging when the data is scaled up to non-web-browser-friendly content sizes