Comparing three institutional software repository platforms

Anusha Ranganathan
Hrafn Malmquist
Steven Eardley

us@cottagelabs.com
anusha@cottagelabs.com
hrafn@cottagelabs.com
steve@cottagelabs.com

cottages labs
Three repository software platforms

<table>
<thead>
<tr>
<th>Platform</th>
<th>Website:</th>
<th>Code:</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSpace 8</td>
<td><a href="https://or24-dspace.cottagelabs.com/">https://or24-dspace.cottagelabs.com/</a></td>
<td><a href="https://github.com/cottagelabs/or24-dspace-angular">https://github.com/cottagelabs/or24-dspace-angular</a></td>
</tr>
<tr>
<td>Invenio v12</td>
<td><a href="https://or24-invenio.cottagelabs.com">https://or24-invenio.cottagelabs.com</a></td>
<td><a href="https://github.com/cottagelabs/or2024-invenio">https://github.com/cottagelabs/or2024-invenio</a></td>
</tr>
<tr>
<td>Hyrax 5</td>
<td><a href="https://or24-hyrax.cottagelabs.com">https://or24-hyrax.cottagelabs.com</a></td>
<td><a href="https://github.com/cottagelabs/or2024-hyrax">https://github.com/cottagelabs/or2024-hyrax</a></td>
</tr>
</tbody>
</table>
InvenioRDM V12

- Packaged version of utilities powering Zenodo; from CERN + numerous partners. Zenodo ported in 2023
- Python & Flask web application with Semantic UI & React
On a fresh VM, install CLI tool to bootstrap an instance

```
invenio-cli init rdm
```

Minor changes to the included docker-compose; replaced self-signed SSL certificate with LetsEncrypt.

```
invenio-cli containers setup --no-demo-data
invenio-cli containers build
```

Customisation via override files: sass, jinja template

```
invenio-cli containers start
```

Deposit via admin user + webform
Cottage Labs: make the most of your data

- Backed by CERN
- Developer-friendly tech stack
- Extensible metadata schema & controlled vocabularies
- Very scalable and performant
- The Invenio community
InvenioRDM Challenges

- A young platform; some features hotly anticipated
- Featureset heavily influenced by Zenodo’s needs
- Something of a maze of interconnected packages can make development tricky
Hyrax
“Hyrax is a Ruby on Rails Engine built by the Samvera community. Hyrax provides a foundation for creating many different digital repository applications.”  (doc) 

Hyrax is not a web application.
Developing the OR24 Hyrax based application

- Reused existing docker-compose scripts
- Created a new Hyrax v5 application
- Created two new worktypes
  - Monograph (copied from Koppie)
  - Dataset (to showcase a nested metadata example)
- Commissioned a new machine and deployed the application
- Manually submitted the 15 records
● Develop any number of work types with
  ○ custom metadata model
  ○ custom workflow and administrative sets (admin sets - curated collections)

● Flexible adapters for metadata and storage with Valkyrie

● Custom workflows

● Integration with other services

● The Samvera community
Cottage Labs: make the most of your data

Hyrax Challenges

- It’s written in Ruby ;) and not a traditional RoR application
- It’s got a long dependency chain of gems
- Need to strike a balance between feature richness and code complexity
COMPARING THREE INSTITUTIONAL REPOSITORY SOFTWARE PLATFORMS AT OR2024

DSpace 8

OR2024 Demo Instance by Cottage Labs
- Turn key institutional repository software
- Using the main branches for version 8
Developing the OR24 DSpace

- Commissioned a new machine and deployed the application
- Forked the DSpace Angular repository and the DSpace backend
- Created a community and a collection
  - Manually submitted the 15 records as an administrator
- Applied basic custom theming and set up certificates
- Ran filter media process to generate thumbnails and extract and index text
Cottage Labs: make the most of your data

- Highly customisable
- Modern architecture
- Strong community
- Integration with other services
DSpace Challenges

- Arbitrary two level container structure imposed
- Resource intensive
- Complex technologies
- No way to ingest files in different directories
<table>
<thead>
<tr>
<th></th>
<th><strong>DSpace</strong></th>
<th><strong>InvenioRDM</strong></th>
<th><strong>Hyrax</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Setup effort</strong></td>
<td>2 to 3 days to deploy a basic repository</td>
<td>~1 day to deploy a basic repository</td>
<td>3 to 4 days to deploy a basic repository</td>
</tr>
<tr>
<td><strong>Versioning</strong></td>
<td>Record level</td>
<td>File level</td>
<td>File level</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td>Java</td>
<td>Python</td>
<td>Ruby</td>
</tr>
<tr>
<td><strong>Maturity</strong></td>
<td>2001</td>
<td>2015 - Invenio3</td>
<td>2010</td>
</tr>
<tr>
<td><strong>Customizing metadata</strong></td>
<td>Administrator</td>
<td>Developer</td>
<td>Developer</td>
</tr>
<tr>
<td><strong>Configurable submission workflows</strong></td>
<td>Yes</td>
<td>No</td>
<td>Yes - greater control</td>
</tr>
<tr>
<td>Differences 2/3</td>
<td>DSpace</td>
<td>InvenioRDM</td>
<td>Hyrax</td>
</tr>
<tr>
<td>----------------</td>
<td>--------</td>
<td>------------</td>
<td>-------</td>
</tr>
<tr>
<td>PIDs</td>
<td>Handle, DOI</td>
<td>DOI</td>
<td>DOI</td>
</tr>
<tr>
<td>Communities/Collections</td>
<td>Needs both</td>
<td>Communities</td>
<td>User and administrative collections</td>
</tr>
<tr>
<td>Usage statistics</td>
<td>Yes</td>
<td>(yes)</td>
<td>(yes)</td>
</tr>
<tr>
<td>Lease</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Embargo management</td>
<td>Automatic</td>
<td>Automatic</td>
<td>Dashboard</td>
</tr>
<tr>
<td>Sharing, proxy deposit and ownership transfer</td>
<td>No</td>
<td>(yes)</td>
<td>Yes</td>
</tr>
</tbody>
</table>
## Differences 3/3

<table>
<thead>
<tr>
<th>Feature</th>
<th>DSpace</th>
<th>InvenioRDM</th>
<th>Hyrax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lease</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Embargo management</td>
<td>Automatic</td>
<td>Automatic</td>
<td>Dashboard</td>
</tr>
<tr>
<td>Sharing, proxy deposit and ownership transfer</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Meta tags for discoverability</td>
<td>Google scholar, Signposting</td>
<td>No</td>
<td>Google scholar, Schema.org</td>
</tr>
<tr>
<td>Multilingual support</td>
<td>Yes</td>
<td>(Yes)</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Factors for choosing

- The object versioning model, types of metadata model and workflows you need to represent
- The skill set available to support the repository
- Community involvement