ManGO: a web portal and framework built on top of iRODS for active research data management

Paul Borgermans & Mariana Montes & Ingrid Barcena Roig
ManGo in the research data life cycle

**ACTIVE PROJECT**
- Data Collection
- Data Storage
- Metadata description
- Data process

**PUBLISHED**
- Data publishing
- Metadata publishing

**LONG TERM**
- Data archiving
- Data Preservation

- Regional / domain specific repo
- RDR (Dataverse)
- Teneo
- Cold storage

ManGo

iRODS
Data management

Data centralization
- Dispersion of data
- Duplication data
+ Accessibility to data
+ Data security

Data management
+ Standardization
+ Automation
+ Data discovery
+ Collaboration

FAIRification from the start!
ManGo Architecture

Users

Clients

ManGo Portal

iCommands
Python Client

DATA PLATFORM API

iRODS

Zone SET

Zone Biomed

Zone GHUM

Rule Engine
MD

Rule Engine
MD

Rule Engine
MD

Rule Engine
MD

Storage

6 PB (double copy)

6 PB (double copy)

Users

iRODS

Zone A

Rule Engine
MD

Rule Engine
MD

Rule Engine
MD

SSO

Users

DATA PLATFORM API

Storage

6 PB (double copy)

Users

DATA PLATFORM API

Storage

6 PB (double copy)
Serving many research projects

- Zone → projects (~60 since March 2023 launch)
- Users/Groups

Add iRODS project  Add globus share  Usage report
## Project life cycle

<table>
<thead>
<tr>
<th>Platform</th>
<th>irods (zone icts.icts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home folder</td>
<td>/icts.icts/home/variable_stars</td>
</tr>
<tr>
<td>Size</td>
<td>8.9 MB of 200.0 MB used</td>
</tr>
<tr>
<td>Inodes</td>
<td>38 of 100 inodes used (6 objects, 4 collections, 28 metadata records)</td>
</tr>
<tr>
<td>Features</td>
<td>Enforce quota, block writes if quota are exceeded</td>
</tr>
<tr>
<td></td>
<td>Enable SFTP open files for read-write</td>
</tr>
<tr>
<td></td>
<td>Enable iCommand/PRC access</td>
</tr>
<tr>
<td></td>
<td>Strict permissions, only project managers have write access to top level collection</td>
</tr>
<tr>
<td></td>
<td>Inherit permissions of the top level collection to subcollections (this implies a flat permission model, nobody is owner of any file)</td>
</tr>
<tr>
<td></td>
<td>Stored data is subject to the General Data Protection Regulation</td>
</tr>
</tbody>
</table>
**ManGo Portal features**

(Generic) Kernel

- CRUD Data / Metadata
- ICAT based search engine
- Permissions/ACL’s
- Metadata schemas
- View templates overrides
- Extension hooks
  - Signals
  - Registration (UI)

- iRODS Python Client (PRC)
- Flask (web) framework
- Jinja2 view templates
- Bootstrap 5 (GUI components)
- Vanilla JavaScript
Portal features

(Generic) Kernel
- CRUD Data / Metadata
- ICAT based search engine
- Permissions/ACL’s
- Metadata schemas
- View templates overrides
- Extension hooks
  - Signals
  - Registration (UI)

(Optional) Plugins / Extensions
- User / group management *
- Data platform UI *
- OpenSearch *
- Reporting / statistics
- Analysis
  - Metadata extraction (Tika)
  - ...

* Requires rodsadmin proxy
**ManGo**: metadata everywhere

- Collections
  - Data objects
- System
  - Automatic metadata extraction
  - Data processing pipelines
- User
  - Documentation and standardization
  - Workflow management
Analysis of embedded metadata

### Inspect with Tika metadata extractor

Analysis executed on 2023-06-13T22:04:42+00:00

<table>
<thead>
<tr>
<th>Metadata fields</th>
<th>Value</th>
<th>Selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIRMASS</td>
<td>1.2297</td>
<td></td>
</tr>
<tr>
<td>BAYERPAT</td>
<td>RGGB</td>
<td></td>
</tr>
<tr>
<td>CCD-TEMP</td>
<td>-5</td>
<td></td>
</tr>
<tr>
<td>CENTALT</td>
<td>54.35</td>
<td></td>
</tr>
<tr>
<td>CENTAZ</td>
<td>136.19</td>
<td></td>
</tr>
<tr>
<td>Content-Length</td>
<td>18104768</td>
<td></td>
</tr>
<tr>
<td>YBINNING</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>YPIXSZ</td>
<td>3.76</td>
<td></td>
</tr>
</tbody>
</table>

Add selected metadata items as regular metadata

<table>
<thead>
<tr>
<th>System properties</th>
<th>Metadata</th>
<th>Permissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>other</td>
<td>analysis</td>
<td></td>
</tr>
<tr>
<td>AIRMASS</td>
<td>1.2297</td>
<td></td>
</tr>
<tr>
<td>BAYERPAT</td>
<td>RGGB</td>
<td></td>
</tr>
<tr>
<td>CCD-TEMP</td>
<td>-5</td>
<td></td>
</tr>
<tr>
<td>EXPTIME</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>FOCALLEN</td>
<td>2086</td>
<td></td>
</tr>
<tr>
<td>INSTRUME</td>
<td>ZWO CCD ASI533MC Pro</td>
<td></td>
</tr>
<tr>
<td>OBSERVER</td>
<td>Paul</td>
<td></td>
</tr>
<tr>
<td>ORIGIN</td>
<td>Garden</td>
<td></td>
</tr>
<tr>
<td>PIERSIDE</td>
<td>WEST</td>
<td></td>
</tr>
</tbody>
</table>
ManGo portal: metadata schemas

Create metadata schemas

- Minimize user mistakes
- Allow for data validation
- Compulsory/optional metadata
- Domain specific metadata

Metadata schema manager for u0123318

New schema

Draft

Published

Archived

Title*
Input type: text

Creator
Data Objects: preview

- WhaleGalaxy-NGC4631-Luminance-session_1-St-Sat-SharpenAI-Focus (2).jpg

Options:
- System properties
- Metadata
- Permissions
- Preview
- Metadata inspection and extraction
Portal: Custom views

Template overrides for zone icts_icts

Source: common/collection_view.html.j2

```json
[
  {
    "target": "common/gallery_carousel.html.j2",
    "matches": {
      "all": {
        "metadata": {
          "match_type": "any",
          "items": [
            {
              "mg.collection_type": "gallery_carousel"
            }
          ]
        }
      }
    }
  },
  {
    "target": "common/collection_view_trash.html.j2",
    "matches": {
      "all": {
        "subtree": "/{{zone}}/trash"
      }
    }
  }
]
```

- Zone specific
- Criteria: “any”, “all”
  - Path elements
  - Exact strings
  - Special variables
- Metadata
- Yaml
## ManGo portal: ACL’s

<table>
<thead>
<tr>
<th>User / Group</th>
<th>Access level</th>
<th>Zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>variable_stars (rodsgroup)</td>
<td>own</td>
<td>icts_icts</td>
</tr>
<tr>
<td>variable_stars_observers (rodsgroup)</td>
<td>modify_object</td>
<td>ict_ict</td>
</tr>
<tr>
<td>variable_stars_planets (rodsgroup)</td>
<td>read_object</td>
<td>ict_ict</td>
</tr>
</tbody>
</table>

### Groups and Rights
- **Groups**: variable_stars
- **Rights**: Own

[Apply Button]
### ManGo portal: bulk operations

#### Select destination collection

- home
  - gbiomed_eximious_pilot001
  - public
  - test_project_portal
  - tudelft
  - u0123318
  - variable_stars
    - exo planets past saturn and jupiter
  - meetings
  - supernova and galaxies

#### Table of data objects

<table>
<thead>
<tr>
<th>Owner</th>
<th>Created</th>
<th>Modified</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>files</td>
<td>exoplanets</td>
<td>2023-03-09 18:00:13</td>
<td>2023-05-31 15:09:12</td>
</tr>
<tr>
<td>checked</td>
<td>00 12 57 lapl4 ap2376 conv_gimp.jpg</td>
<td>2023-03-31 13:46:57</td>
<td>2023-03-31 13:46:58</td>
</tr>
<tr>
<td>checked</td>
<td>m2Z.jpg</td>
<td>2023-02-10 13:56:37</td>
<td>2023-02-10 13:56:37</td>
</tr>
<tr>
<td>unchecked</td>
<td>ngs6888.jpg</td>
<td>2023-02-10 13:56:38</td>
<td>2023-02-10 13:56:38</td>
</tr>
</tbody>
</table>

- **Move**
- **Copy**
- **Delete**
**ManGo portal: group management**

- Via rodsadmin proxy account
- Enforced naming conventions

---

**Group manager for realm** *variable_stars*

**variable_starsObservers**

**Current members**
- u0123318
- u0118974

**Other members of variable_stars**
- u0137480
- u0116999
- u0089722

---

**Add group**

Group name suffix variable_stars_[suffix]

- astronomers
Admin

Open search index plugin

Launch ad hoc indexing job

Indexing thread status
Current status active

Client/server connection health
Query server: □
Ingest server: □

Index job queue (0)
No active jobs on this node

Add collection to index jobs

Change status
Restart indexing thread

Refresh client/server connections
### Overall OpenSearch and zone stats

<table>
<thead>
<tr>
<th>Collection</th>
<th>sub collections</th>
<th>data objects stats</th>
<th>size distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>/icts_icts</td>
<td>count</td>
<td>count: 520</td>
<td>0 Bytes - 10.0 kB: 992</td>
</tr>
<tr>
<td></td>
<td></td>
<td>total size: 23.9 GB</td>
<td>10.0 kB - 100.0 kB: 22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>max size: 8.2 GB</td>
<td>100.0 kB - 1.0 MB: 753</td>
</tr>
<tr>
<td></td>
<td></td>
<td>min size: 0 Bytes</td>
<td>1.0 MB - 10.0 MB: 1252</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10.0 MB - 100.0 MB: 26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100.0 MB - 1.0 GB: 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.0 GB - 10.0 GB: 4</td>
</tr>
</tbody>
</table>
Future

• Expand OpenSearch integration
• Refactor KU Leuven specifics into extensions (few left)
• Workflows

Photo by Drew Beamer on Unsplash
Thank You!

paul.borgermans@kuleuven.be
mariana.montes@kuleuven.be