iRODS Python/PRC based portal and tools for active data support in research contexts

Paul Borgermans
KU Leuven ICTS RDM: the team

Paul Borgermans
Mustafa Dikmen
Jef Scheepers
Ingrid Barcena

Peter Verraedt
Filip Sevenants

Tom Vanmierlo
Bart Vanneste

User Support
User tools
Policies
Documentation

iRODS
middleware

Storage
Goals

- Modular, web-based platform and optional extra tools
- Easy to use for researchers (and us)
- Metadata ❤️
- Generic core
- Customizable for many use cases
- Easy to adapt (by 3rd parties, developers)
Features (end user oriented)

• Base
  • CRUD collections / data objects / metadata
  • Search (direct ICAT/DB based)
  • Basic rights management
  • User/Group information
  • Metadata schemas/templates
• Optional (domain specific)
  • Data object preview/thumbnails
  • Data object inspection (content)
Technical components

- iRODS Python Client (PRC)
- Flask (web) framework
- Bootstrap 5 (GUI components)
- Vanilla JavaScript (mostly)
- OpenSearch (Solr?)
- Apache Tika (text and metadata extraction)
- Various Python modules as required

As much as possible main-stream and popular components
Feature rundown
Collections: basic view

Dashboard
Collections
Advanced search
Your groups
Trash
Metadata templates

Add collection
Choose file
No file chosen
Upload

Content
Metadata
Permissions

Name
Owner
Created
Modified
Size

- assignments
- ingrid
- irods-portal-python
- meetings
- microscopy-lab
- schema-only-metadata
- testlab
## Data Objects: system properties

### M31-cropped-2020-09-18.jpg

<table>
<thead>
<tr>
<th>System properties</th>
<th>Metadata</th>
<th>Permissions</th>
<th>Preview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Owner</td>
<td>vsc33436</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Created</td>
<td>2021-08-25 14:49:20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified</td>
<td>2021-08-25 14:49:20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>1.6 MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal id</td>
<td>13312681</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Status</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replica number</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Objects: basic metadata view and edits

**M31-cropped-2020-09-18.jpg**

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>irods::access_time</td>
<td>1639149110</td>
<td></td>
</tr>
<tr>
<td>ku.mime_type</td>
<td>image/jpeg</td>
<td></td>
</tr>
<tr>
<td>comment</td>
<td>Earlier attempt for Andromeda using a Nikon 180mm ED</td>
<td></td>
</tr>
</tbody>
</table>
## Data Objects: permissions view (and edit)

- **kuleuvenTier1Pilot** > **home** > **vsc33436** > **M31-cropped-2020-09-18.jpg**

### M31-cropped-2020-09-18.jpg

<table>
<thead>
<tr>
<th>User / Group</th>
<th>Access level</th>
<th>Zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>vsc33436</td>
<td>own</td>
<td>kuleuvenTier1Pilot</td>
</tr>
</tbody>
</table>
Data Objects: preview (whitelisted)

M31-cropped-2020-09-18.jpg

System properties  Metadata  Permissions  Preview

![Galaxy Image](M31-cropped-2020-09-18.jpg)
Search

- Somewhat advanced ICAT based search
  - Collection names
  - Data Object names
  - Create/modified dates
  - Metadata
  - Collection/subtree filter
Search: ICAT based search

- Dashboard
- Collections
- Advanced search
- Your groups
- Trash
- Metadata templates

Type: Data object
Collection (subtree): /kuleuven_tier1_pilot/home/vsc33436

Comparison: Contains
Name: M31

Created
Comparison: Before
Date: dd/mm/yyyy

Modified
Comparison: Before
Date: dd/mm/yyyy

Metadata
Attribute name: ku.mime_type
Attribute value: application/jpeg

Search
Search: result

<table>
<thead>
<tr>
<th>Type</th>
<th>Path</th>
<th>Name</th>
<th>Size</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>data_object</td>
<td>/kuleuven_tier1_pilot/home/vsc33436</td>
<td>M31-cropped-2020-09-18.jpg</td>
<td>1.6 MB</td>
<td>vsc33436</td>
</tr>
</tbody>
</table>

Displaying 1 - 1 items in total 1

Searching took 0.516 seconds
Metadata schemas and templates

- Manual or semi-manual editing of metadata
- Minimize user errors
- Compulsory metadata
- Project/domain based requirements
  - Preparing for transition to archived/published data
Schema editing

My metadata templates

- another-schema.json [Edit] [View] [Delete]
- project.json [Edit] [View] [Delete]
- ui-schema.json [Edit] [View] [Delete]
- rdr.json [Edit] [View] [Delete]
- Project.json [Edit] [View] [Delete]
- datacite.json [Edit] [View] [Delete]
- bio-incubator.json [Edit] [View] [Delete]
- template-name.json [Edit] [View] [Delete]

Create new template

Load JSON
Schema editing

Load JSON from PC  Save template  Cancel

Project

Add element

Project name

☑️ Required  T  ↓  Delete

Add element

Description

☑️ Required  T  ↓  Delete

Add element
Schema editing

What form element would you like to add?

Text options: regular text, number, date, email, time or url
Title
example placeholder

Title (Needs at least 5 options, otherwise use radio)
Open this select menu

Title
✓ Option 1
☐ Option 2
☐ ...

Title (Can have at most 4 options, otherwise use select)
Schema based metadata editing

**microscopy-lab**

Type: collection

Metadata schema: **Project**

- **Project name**: Main optical microscope types
- **Description**: Laboratory assignments for bachelor students including polarized light microscopy
- **Area**: Biology

[Save] [Cancel]
Schema based metadata: name spaces

$ imeta ls -C .

AVUs defined for collection /kuleuven_tier1_pilot/home/vsc33436/microscopy-lab:
attribute: irods::collection::type
value: default
units:
----
attribute: ku.project.name
value: Main optical microscope types
units:
----
attribute: ku.project.description
value: Laboratory assignments for bachelor students including polarising setups
units:
----
attribute: ku.project.area
value: Biology
units:
Schema based metadata editing (collections)
Schema based metadata results

microscopy-lab

Add collection

Choose file No file chosen

Upload

Content Metadata Permissions

other project

Project name Main optical microscope types
Description Laboratory assignments for bachelor students including polarising setups
Area Biology

Edit
Data Object inspection (optional)

- Domain specific
- Metadata “discovery” and extraction into iRODS
- Tool and/or component for analysis chains
  - Enhanced indexing and searching
  - Processing workflows
Data Object inspection example: Apache Tika
Data Object inspection: Apache Tika results

Metadata fields

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component 1</td>
<td>Y component: Quantization table 0, Sampling factors 2 horiz/2 vert</td>
</tr>
<tr>
<td>Component 2</td>
<td>Cb component: Quantization table 1, Sampling factors 1 horiz/1 vert</td>
</tr>
<tr>
<td>Component 3</td>
<td>Cr component: Quantization table 1, Sampling factors 1 horiz/1 vert</td>
</tr>
</tbody>
</table>

Compression Type: Baseline
Content-Type: image/jpeg
Data Precision: 8 bits
Exif IFD0:Software: Google
File Modified Date: Mon Jul 14 23:02:10 +0000 2022
File Name: apache-tika-7512063351860139944.bmp
File Size: 104560 bytes
ICC:Blue Colorant: (0.1571, 0.0666, 0.7841)
ICC:Blue TRC: para (0x70617261): 32 bytes
ICC:CMM Type: appl
ICC:Chromatic Adaptation: sRGB (0x766f726b): 44 bytes
ICC:Class: Display Device
ICC:Color space: RGB
ICC:Device manufacturer: APPL
ICC:Green Colorant: (0.292, 0.592, 0.0419)
ICC:Green TRC: para (0x70617261): 32 bytes
ICC:Media White Point: (0.3142, 1.0, 0.8249)
ICC:Primary Platform: Apple Computer, Inc.
ICC:Profile Connection Space: XYZ
ICC:Profile Copyright: 1 enUS(Copyright Apple Inc., 2022)
Around the corner

- Advanced search module
  - Based on OpenSearch (Solr may be another option)
  - Expanded indexing and searching
  - Collection metadata inheritance for data objects
  - Additional metadata through extraction/analysis of data objects
  - Scaling for huge datasets
- Permission management
Around the corner or a bit further

- Minimal “project management/admin” module (POC)
- Dataset and metadata packaging
  - Research Object Crate
  - Tar, with metadata in json schema
  - ...
- Integration with Globus
Questions?