

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



User Support
User tools
Policies
Documentation



Peter
Verraedt



Filip
Sevenants

iRODS
middleware



Tom
Vanmierlo

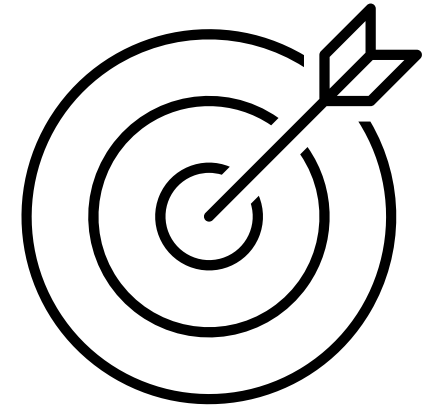


Bart
Vanneste

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

[kuleuven.tier1.pilot](#) » [home](#) » vsc33436







vsc33436

[Add collection](#) [Upload](#)

[Content](#) [Metadata](#) [Permissions](#)

Name	Owner	Created	Modified	Size
assignments				
ingrid				
irods-portal-python				
meetings				
microscopy-lab				
schema-only-metadata				
test-rob				

Data Objects: system properties

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

[kuleuven_tier1_pilot](#) > [home](#) > [vsc33436](#) > M31-cropped-2020-09-18.jpg

M31-cropped-2020-09-18.jpg



System properties

[Metadata](#)

[Permissions](#)


[Preview](#)

Owner	vsc33436
Created	2021-08-25 14:49:20
Modified	2021-08-25 14:49:20
Size	1.6 MB
Internal id	13312681
Status	None
Replica number	0



Data Objects: basic metadata view and edits

[kuleuven_tier1_pilot](#) > [home](#) > [vsc33436](#) > M31-cropped-2020-09-18.jpg

M31-cropped-2020-09-18.jpg



[System properties](#) [Metadata](#) [Permissions](#) [Preview](#)

Name	Value	Unit
irods::access_time	1639149110	
ku.mime_type	image/jpeg	
comment	Earlier attempt for Andromeda using a Nikon 180mm ED	 

[Add metadata](#)

Data Objects: permissions view (and edit)

[kuleuven_tier1_pilot](#) > [home](#) > [vsc33436](#) > M31-cropped-2020-09-18.jpg

M31-cropped-2020-09-18.jpg



[System properties](#)

[Metadata](#)

Permissions

[Preview](#)

User / Group

Access level

Zone

vsc33436

own

kuleuven_tier1_pilot

Data Objects: preview (whitelisted)

[kuleuven tier1_pilot](#) > [home](#) > [vsc33436](#) > M31-cropped-2020-09-18.jpg

M31-cropped-2020-09-18.jpg



[System properties](#)

[Metadata](#)

[Permissions](#)


[Preview](#)





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	Attribute value	Attribute unit
ku.mime_type	application/jpeg	

Search

Search: result

Show/hide filters

Results

displaying **1** - **1** items in total **1**







Type	Path	Name	Size	Owner
data_object	/kuleuven_tier1_pilot/home/vsc33436	M31-cropped-2020-09-18.jpg	1.6 MB	vsc33436

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

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

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



Delete

Add element

Description

Required



Delete

Add element

Schema editing

What form element would you like to add?

Text input

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

Title

example placeholder

Select

Title (Needs at least 5 options, otherwise use radio)

Open this select menu



Checkboxes

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 *

Description *

Area ▾

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)

[kuleuven tier1_pilot](#) » [home](#) » [vsc33436](#) » microscopy-lab

microscopy-lab

[Add collection](#) No file chosen [Upload](#)

[Content](#) [Metadata](#) [Permissions](#)

[other](#) project

irods::collection::type default

[Add metadata](#)

Edit/add metadata via schema

another-schema ▼

[Edit](#)

Schema based metadata results

[kuleuven_tier1_pilot](#) » [home](#) » [vsc33436](#) » microscopy-lab

microscopy-lab

[Add collection](#) 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

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

[kuleuven_tier1_pilot](#) > [home](#) > [vsc33436](#) > [schema-only-metadata](#) > test3-text-ocr.jpg

test3-text-ocr.jpg

System properties Metadata Permissions **Preview**



Analysis

Perform OCR on the image to extract text Skip cache [Ask Tika](#)

Data Object inspection: Apache Tika results

test3-text-ocr.jpg



Analysis via Tika

org.apache.tika.parser.DefaultParser

org.apache.tika.parser.ocr.TesseractOCRParser

org.apache.tika.parser.jpeg.JpegParser

Text content for full text indexing

```
EU Digital COVID
Certificate
EU Digitaal COVID-Certificaat
Certificat Numérique Européen com Ga ass *
Digitales EU-COVID-Zertifikat: '
```

Metadata fields

Component 1	Y component: Quantization table 0, Sampling factors 2 horiz/2 vert
Component 2	Cb component: Quantization table 1, Sampling factors 1 horiz/1 vert
Component 3	Cr component: Quantization table 1, Sampling factors 1 horiz/1 vert
Compression Type	Baseline
Content-Type	image/jpeg
Data Precision	8 bits
Exif IFD0:Software	Google
File Modified Date	Mon Jul 04 23:02:10 +00:00 2022
File Name	apache-tika-7312863351860139944.tmp
File Size	104580 bytes
ICC:Blue Colorant	(0.1571, 0.0666, 0.7841)
ICC:Blue TRC	para (0x70617261): 32 bytes
ICC:CMM Type	appl
ICC:Chromatic Adaptation	sf32 (0x73663332): 44 bytes
ICC:Class	Display Device
ICC:Color space	RGB
ICC:Device manufacturer	APPL
ICC:Green Colorant	(0.292, 0.6922, 0.0419)
ICC:Green TRC	para (0x70617261): 32 bytes
ICC:Media White Point	(0.9642, 1, 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?