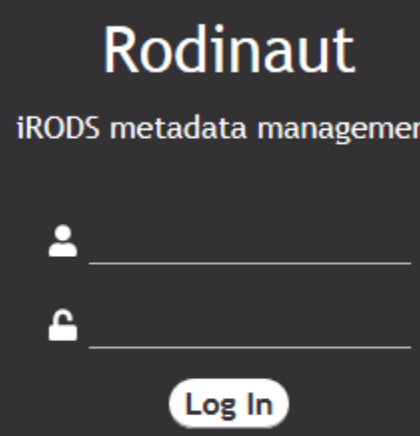
Welcome to …



Othmar Weber // Bayer Business Services // iRODS UGM 2019 Utrecht

iRODS @ Bayer

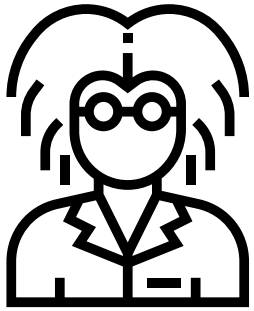
- When?
  - Introduced in 2014
- What?
  - Scientific data (images, genome, …)
- Who?
  - Data scientists
  - Data manager
  - Data loader
- How much?
  - Three department installations
  - ~160 TB

iRODS

@

BAYER

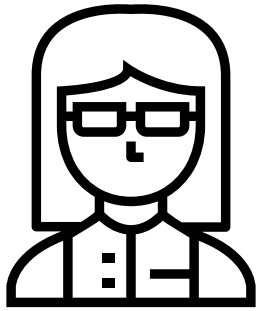
iRODS usage in genetic research



Data Scientist

I am mainly interested in my research project. I need the data for my work. When I am done I have to archive the raw and result data into iRODS and tag it properly accourding our rules.

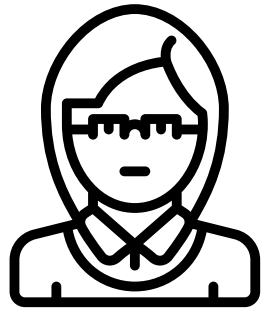
I don‘t like bureaucracy and rules!



Data Owner

As a data owner I need to make sure that the access to the data is limited to the absolute minimum. If a patient retracts her/his consent I need to identify her/his data and make sure that it gets deleted.

For me data is handled like an asset.

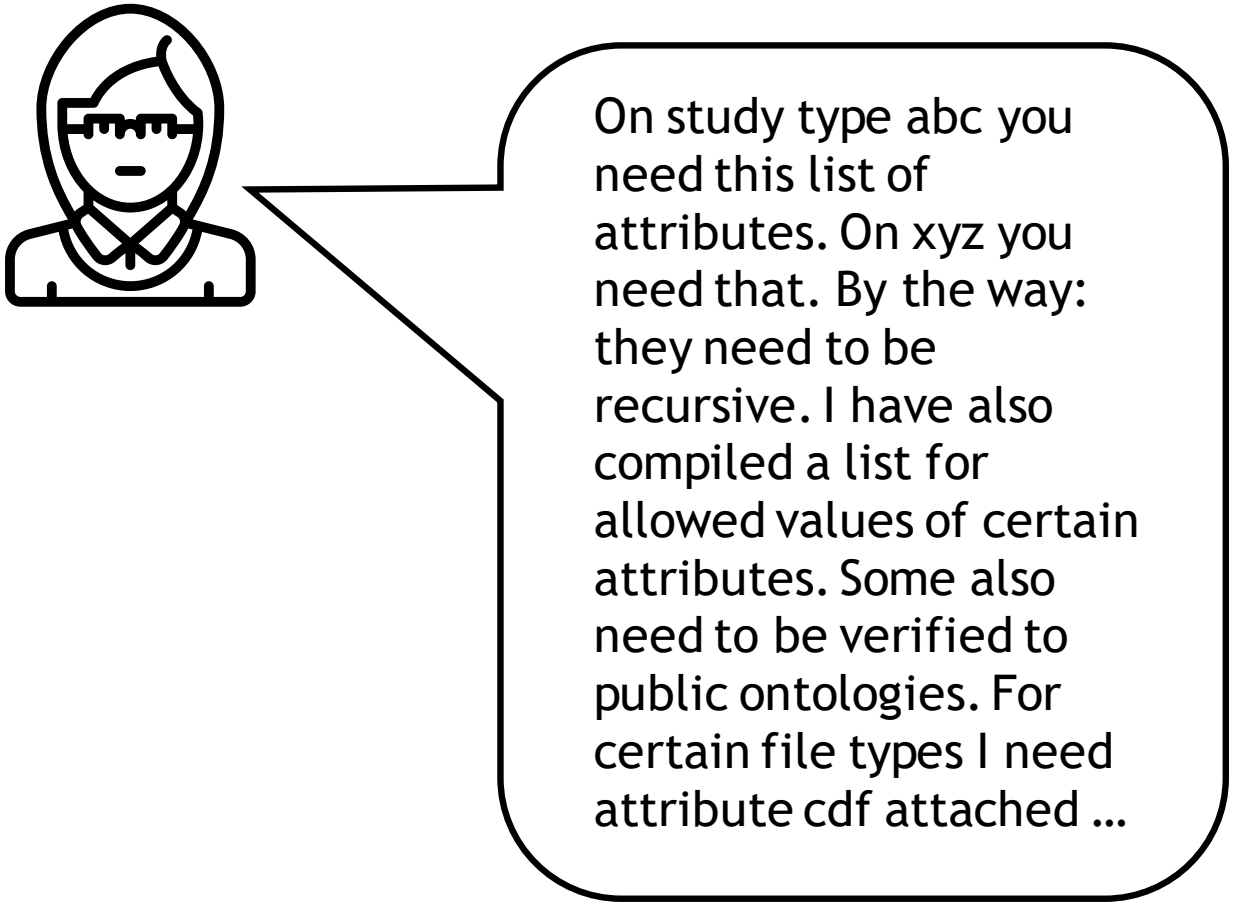


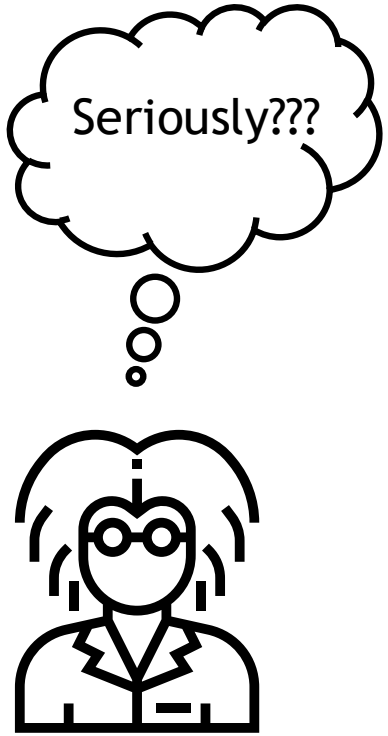
Data Manager

To make sure that our data is well organized I have compiled a list of metadata attributes that need to be attached to each item. The rules differ from data type and research type.

Rules are needed to make data FAIR!

Following the rules …





Finding: Creating rules is easier than complying to them!

1st approach: iRODS Metatool


```
<!-- ************************************************************************** -->
<!-- Rules which apply to all bioinformatics studies -->
<!-- ************************************************************************** -->

<rule_group>
  <selectors>
    <irodspath pattern="/bhc.*Zone/bhcbio/.*studies.*/.+" />
  </selectors>
  <rules>
    <require_one_attribute><name>data_classification</name></require_one_attribute>
    <require_one_attribute><name>study_owner</name></require_one_attribute>
    <require_one_attribute><name>study_owner_cwid</name></require_one_attribute>
    <require_one_attribute><name>contact_person</name></require_one_attribute>
    <require_one_attribute><name>contact_person_cwid</name></require_one_attribute>
  </rules>
</rule_group>

<!-- Only match study directory -->
<rule_group>
  <selectors>
    <irodspath pattern="/bhc.*Zone/bhcbio/(research_studies_secret|research_studies|clinical
  </selectors>
  <rules>
    <require_one_attribute><name>retention_class</name></require_one_attribute>
  </rules>
```

Rules are defined in XML file

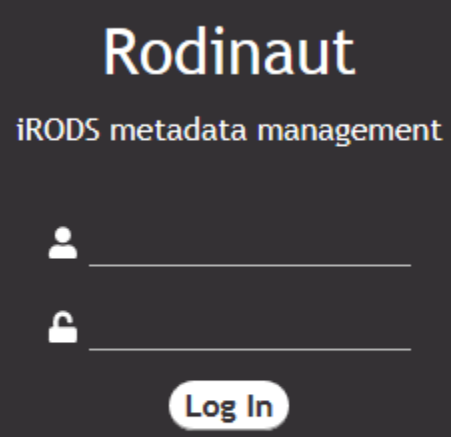
	A	B	C	D	E	F	G
1	MainPath	- absolute collection path, e.g. /bhcbioZone/bhcbio/research_studies/abc -					
2	**Path**	**Recursive**	**Attribute**	**Value**	**Unit**	***Message***	***Attribute Definition***
3	C- .	x	study_type	phase II			type of the study, i.e. research, p
4	C- .	x	bayer_study_id	abc			study ID within Bayer
5	C- .	x	contact_person	John Doe			acting as backup for the data ow
6	C- .	x	contact_person_cwid	JD			acting as backup for the data ow
7	C- .	x	cross_study_analysis_	allowed			statement by data owner, if data
8	C- .	x	data_classification	Internal			bayer data classification
9	C- .	x	external_reference	study123			study reference, e.g. URL to pub
10	C- .	x	health_status	fit and well			health status of the samples; us
11	C- .	x	indication	Malignant tumor of l			detailed indication for this study -
12	C- .	x	indication_permitted	Malignant tumor of l			research area, in which data car
13	C- .	x	species	Rattus norvegicus			species, i.e. human
14	C- .	x	study_description	Demo Study			brief description of the study (i.e.
15	C- .	x	study_owner	Claire Grube			owner of the data, usually princip
16	C- .	x	study_owner_cwid	CG			owner of the data, usually princip
17	C- .	x	bayer_compound_no				e.g. BAY number
18	C- .	x	other_restrictions				Please provide all potential restri
19	C- .	x	retention_class	mid_term			Time interval before the data ow
20							

Users can use Excel to create input for the metatool


```
~]$ #irods-metatool -i LargeAnimalLab.tsv.txt  -m OVERWRITE_OR_APPEND -p /cardtestZone/labs/III.10/2018/phmnt-01228
~]$
~]$
~]$
~]$
~]$
```

Metadata can be attached via the metatool which checks the rules

Next iteration …

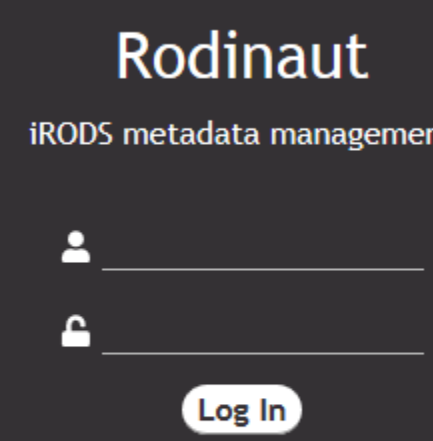


For Scientists who need to ensure compliance with data security/privacy, and find information in iRODS, Rodinaut is a web application that enables viewing and managing metadata. Unlike the existing command-line tool, our product is self-explaining, easy and fast to use and improves user experience with iRODS.

iRODS Metadata mangement made easy…

- Intuitive & visual metadata management
  - Controlled vocabulary via template definitions
  - Supporting lists and ontology lookups
- Using standard iRODS features
  - access rights
  - No change of iRODS code
- Concept of metadata inheritance
  - Asynchronous with locking
- Based on Open Source technology
  - Docker
  - ReactJS (Frontend)
  - Java Spring Boot (Backend) + Jargon

Live (?) Demo …



Navigating the tree...

Rodinaut

Log Out

bhctestZone / testcases / Rodinaut / OW-irods-metatool-backend

- clinical_studies
- coll
- collection_no_permissions
- file
- special study
- study-attribute-settings
- study-invalid-metadata
- study-no-metdata
- study-without-inheritance-at-top-level
- study1

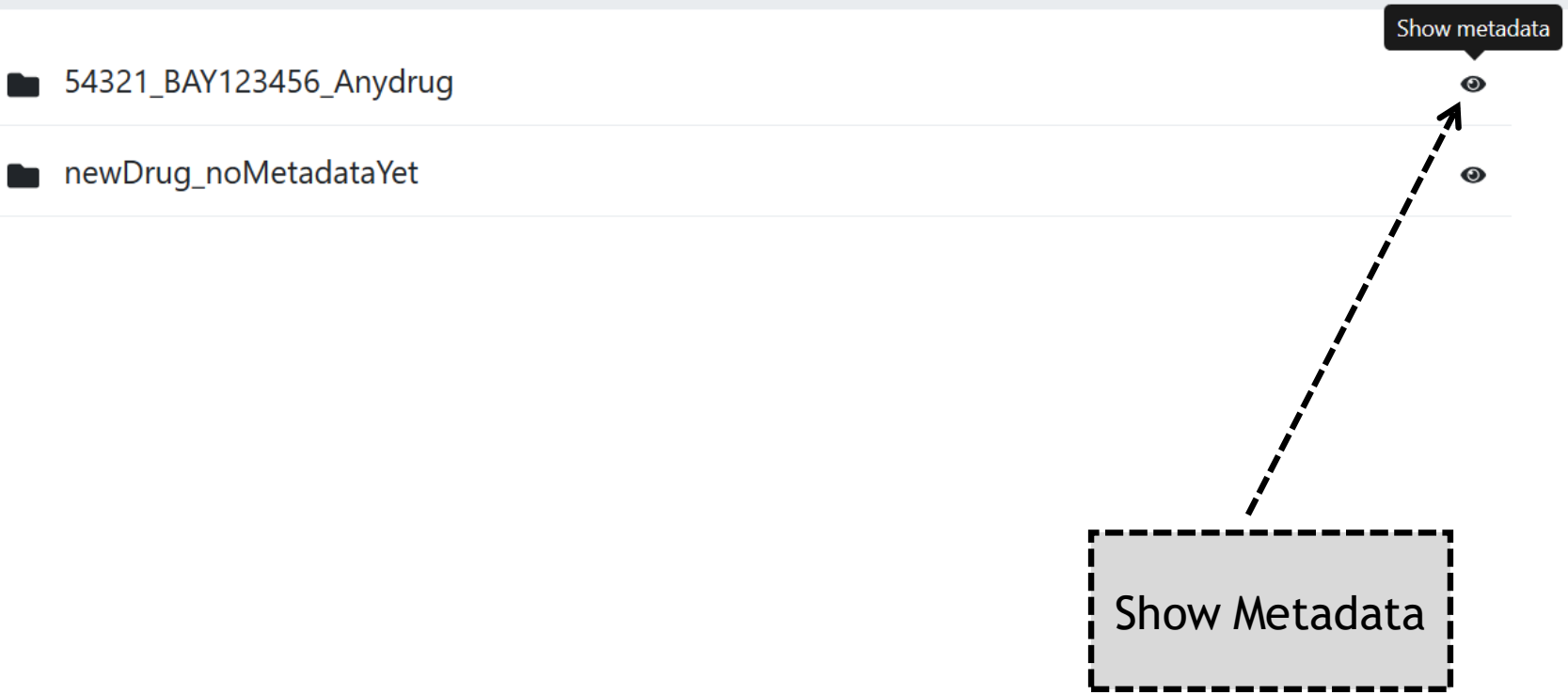
Navigation via breadcrumbs

Get to the Metadata

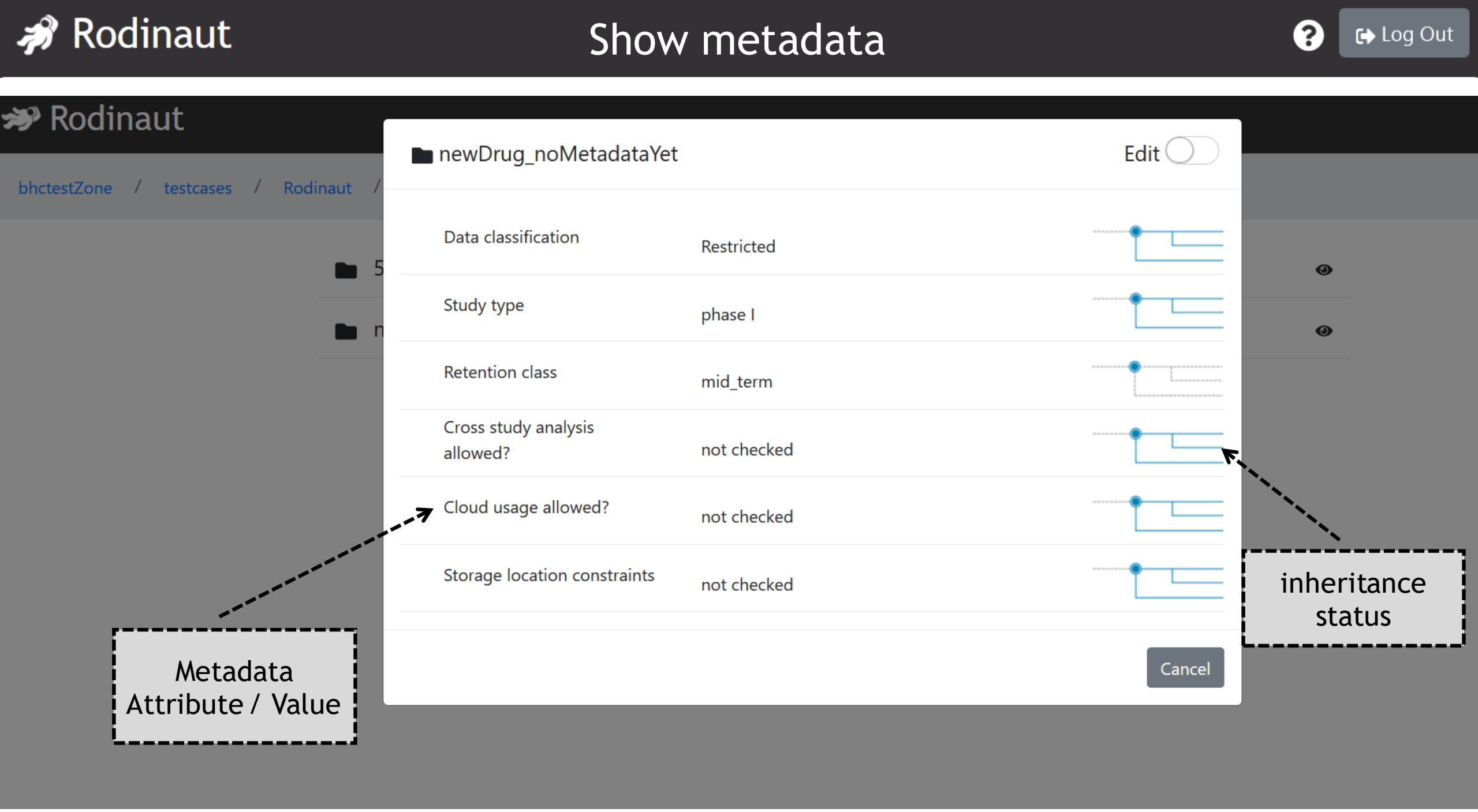
Rodinaut

Log Out

bhctestZone / testcases / Rodinaut / OW-irods-metatool-backend / clinical_studies



Show metadata



Editing Metadata

Rodinaut

Log Out

Rodinaut

bhctestZone / testcases / Rodinaut

newDrug_noMetadataYet

Edit

* Bayer Study ID: 12345

* Data classification: Restricted

* BAY number: any text (multiple values in multiple lines)

Study type: phase I

* Study Owner: any te

* Study Owner CWID: any text

* Contact Person: any text

* Contact Person CWID: any text

Required attribute based on tree

Predefined list

Not required attributes can be removed

Editing Metadata

Rodinaut

Log Out

Rodinaut

bhctestZone / testcases / Rodinaut

* Contact Person: My Contact

* Contact Person CWID: adasdf

Invalid length for metadata attribute value, must not be longer than 5 characters

Attribute specific validation

* Retention class: long_term

Ontology verification

* Indication: bla

SNOMED CT Browser

Metadata value "bla" is not found as a preferred term or synonym in ontology SNOMED CT (terms below "Clinical finding (finding)").

* Cross study analysis allowed?: no

* Cloud usage allowed?: not checked

* Storage location constraints: none

Add additional attributes

Add Optional Attribute...

Check metadata of subdir

Rodinaut

Log Out

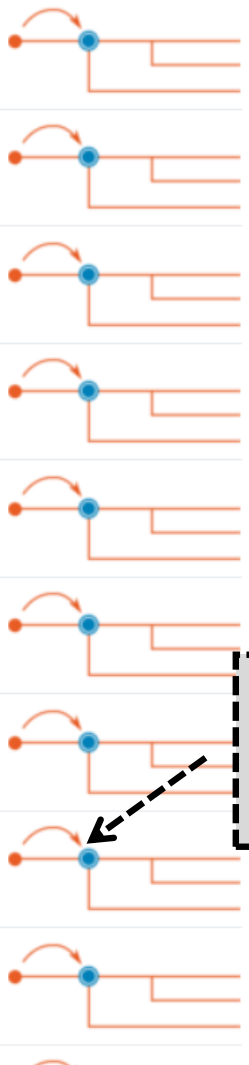
Rodinaut

bhctestZone / testcases / Rodinaut

subdir

Edit

Bayer Study ID	12345
Data classification	Restricted
BAY number	anytext
Study type	phase I
Study Owner	John Doe
Study Owner CWID	bla
Contact Person	My Contact
Contact Person CWID	foo
Indication	Fit and well
Cross study analysis	



Inherited from parent

Break inheritance

Rodinaut

Log Out

Rodinaut

bhctestZone / testcases / Rodinaut

subdir

Edit

* Bayer Study ID: 12345

* Data classification: Restricted

* BAY number: anytext

Study type: phase I

* Study Owner: John Doe

* Study Owner CWID: bla

* Contact Person: My Contact

* Contact Person CWID: foo

Inherited attributes are not editable

inherit value from parent

overwrite inherited value

Change inheritance to allow overwrite

Break inheritance

Rodinaut

Log Out

Rodinaut

bhctestZone / testcases / Rodinaut

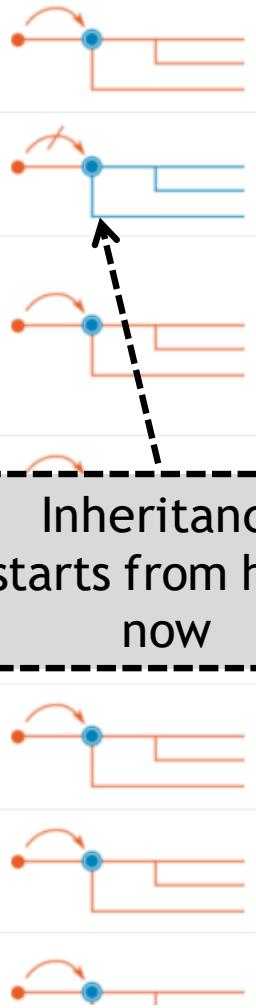
subdir

Edit

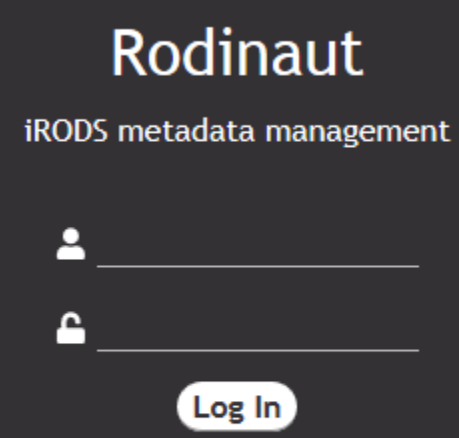
* Bayer Study ID	12345
* Data classification	Restricted
* BAY number	anytext
Study type	phase I
* Study Owner	John Doe
* Study Owner CWID	bla
* Contact Person	My Contact
* Contact Person CWID	foo

Attribute now changable

Inheritance starts from here now

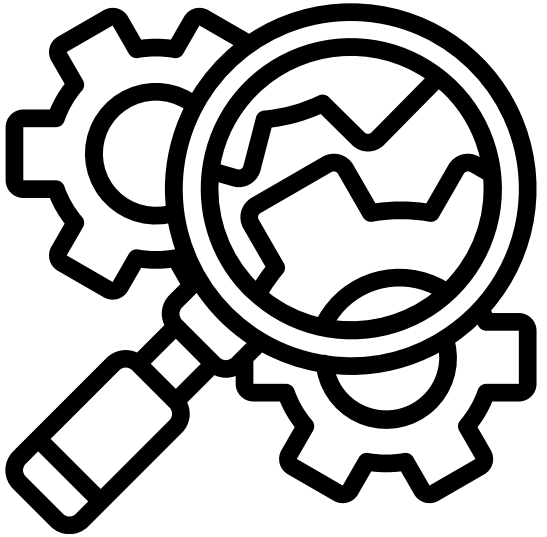


Demo mode off …



Behind the scenes

- Custom attribute to realize inheritance (pass_on_attribute)
  - Sets the root of an inheritance
  - Contains the name of the attribute that should be inherited
  - Can be set multiple times for different attributes
  - Setting on a child level means breaking the inheritance

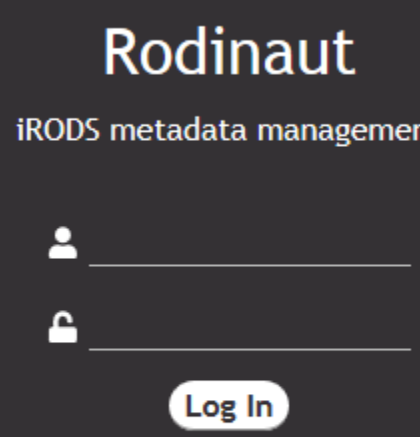


- Background job to recursively apply the attributes to a tree
  - Needed to avoid blocking the UI while large trees are processed
  - Locks the tree where inheritance is performed (just for Rodinaut)
  - Runs as service with the rights of the original user

Work in progress …

- Metadata Search
- File Creation, Upload & Download
- Session handling bug requires daily restart
- Strong separation of user and system rights
- Better documentation of rules.xml, create XSD
- Better abstraction of infrastructure services (auth provider, ontology definition)

Logout completed …



The team:

Henrik Seidel

Ekaterina Nevedomskaya

Thomas Leyer

Marc Schwering

Othmar Weber

Carsten Jahn

Media sources

- p3,4,5: Icons taken from flaticon.com, Author: Eucalyp, freepik, freepik
- p8: Graphics taken from freepik.com, Author freepik
- p19: Icon taken from flaticon.com, Author: Smashicons
- p20: Icon taken from Wikimedia, Public Domain
- p23: Icon taken from flaticon.com, Author: Elegant Themes