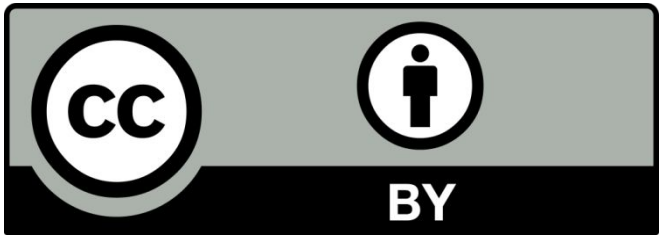
# FAIR Tools

**http://tinyurl.com/FAIRToolsPortugal**

Mark D Wilkinson, CBGP UPM-INIA, 2017
**ORCiD**:0000-0001-6960-357X



FAIR Metadata

FAIR Data

FAIR Publishing

FAIR Metrics

## FAIR Metadata

SEEK (FAIRDOM)

CEDAR

FAIR Accessor

## FAIR Data

RightField (FAIRDOM)

FAIRifyer + RDF Extension

FAIR Projection && D2RQ Mapper

## FAIR Publishing

FAIR Data Point;  smartAPI;  SADI

## FAIR Metrics

# SEEK

FAIRDOM Project


# FAIRDOM: Reproducible systems biology through FAIR asset management

Natalie Stanford, University of Manchester
@nataliestanford
@fairdom_eu

**F**indable

**A**ccessible

**I**nteroperable

**R**eusable

**D**ata

**O**perating procedures

**M**odels

# Core Funders

NWO
Netherlands Organisation for Scientific Research

BBSRC
bioscience for the future

ptJ
Projektträger Jülich
Forschungszentrum Jülich

SystemsX.ch
The Swiss Initiative in Systems Biology

Bundesministerium
für Bildung
und Forschung

# FAIRDOM Usage




de.NBI
German Network for
Bioinformatics Infrastructure



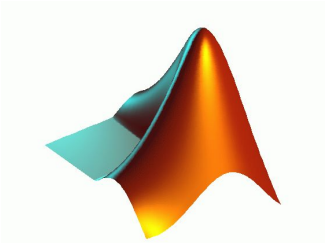





**Independent researchers**

# Researchers generate, record, store and share data in many formats.

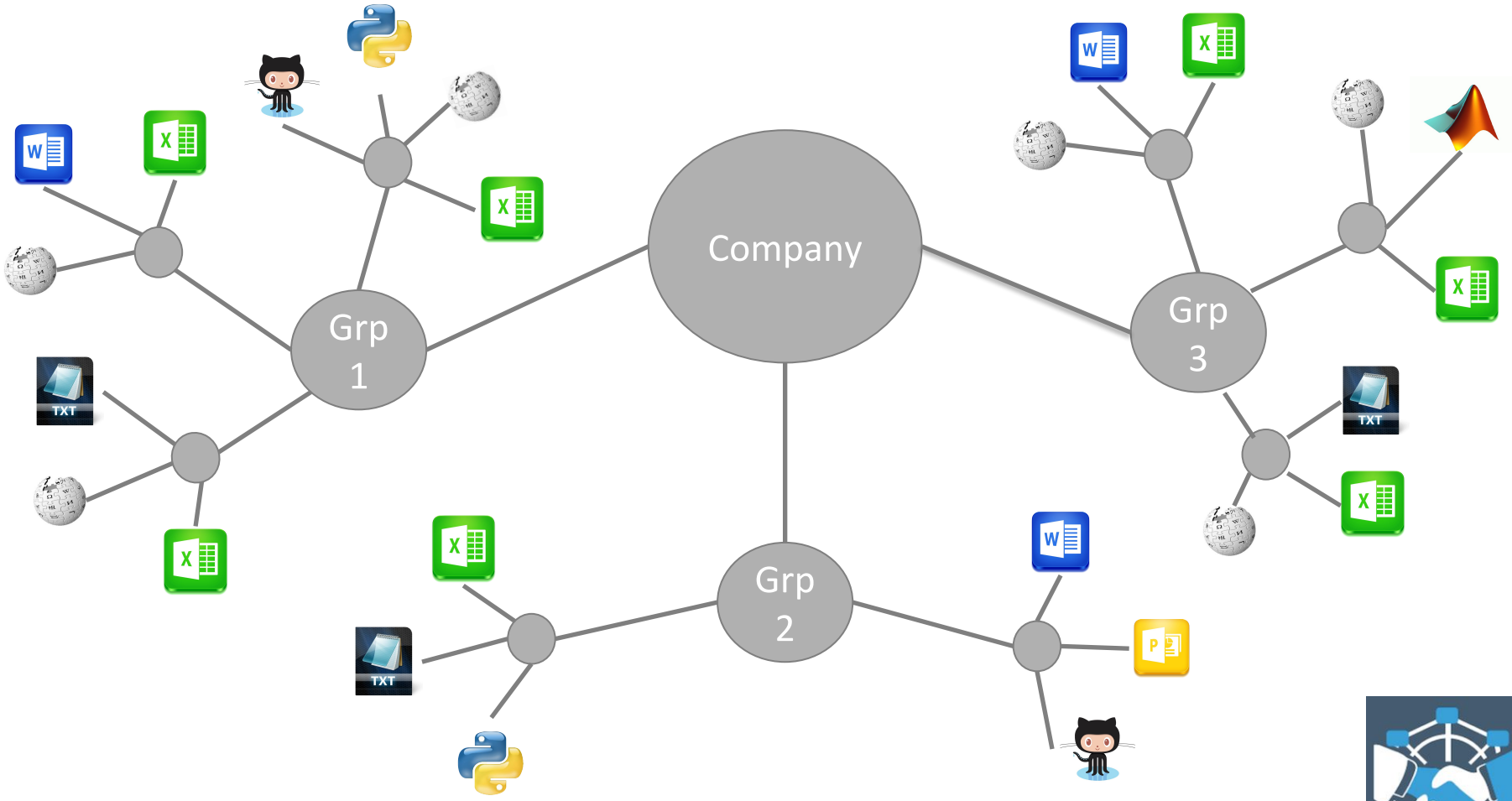




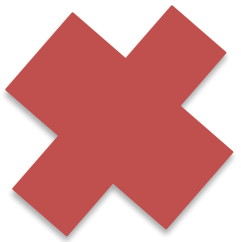


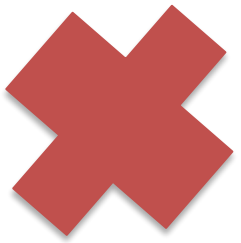



# Each researcher has their own preferences.

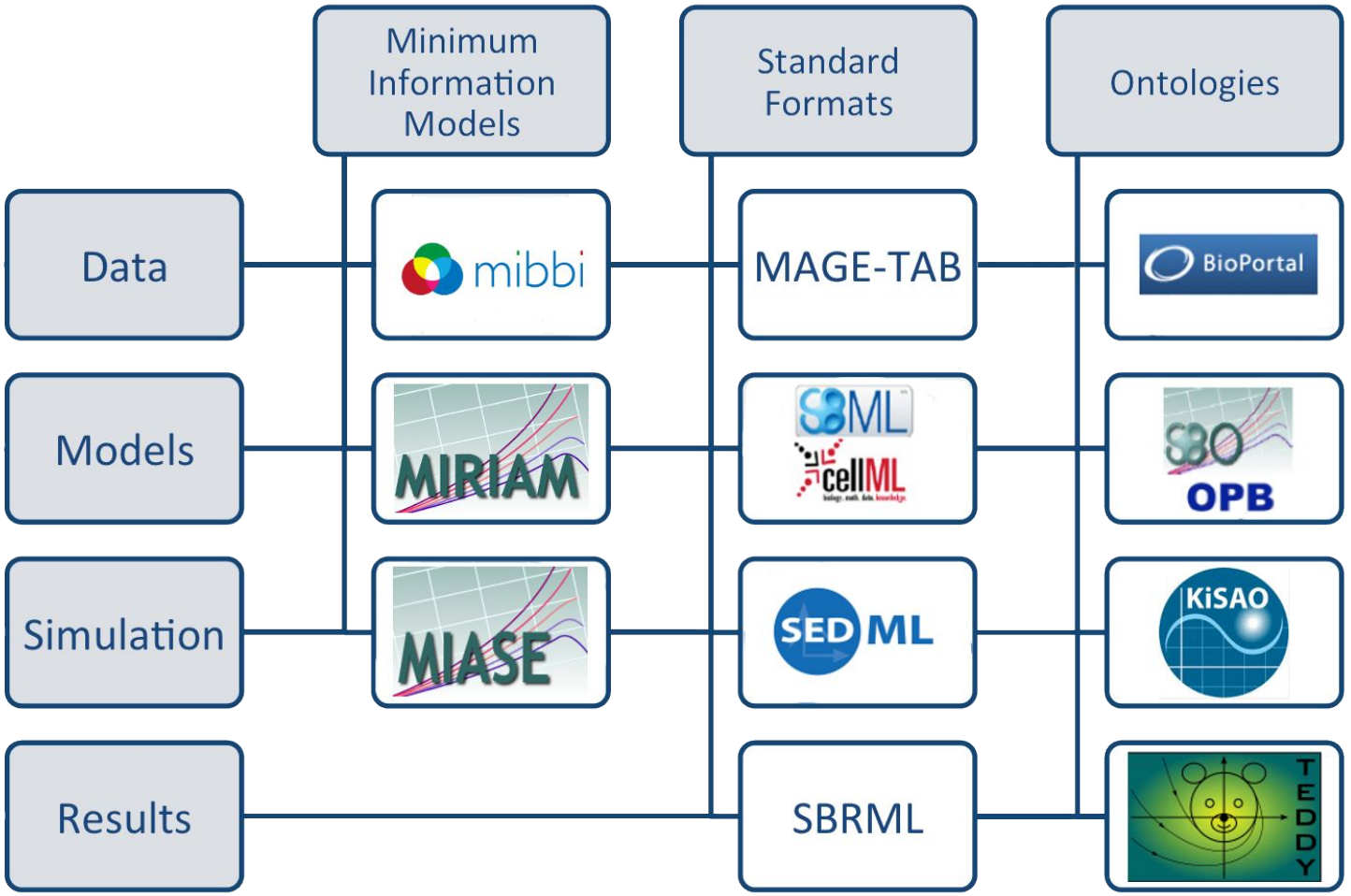


# This type of data management does not adhere to FAIR principles.

**INTEROPERABLE** 

**REUSABLE** 

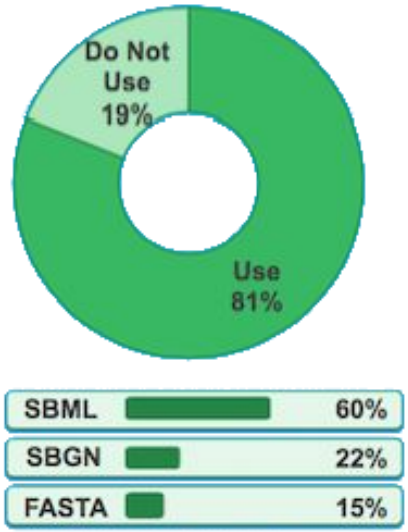
# Standards are available that help to improve understanding and exchange.



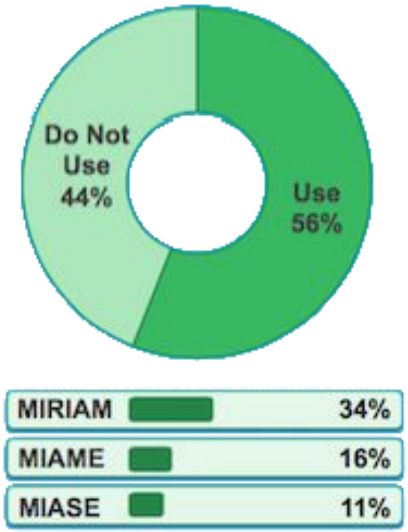

Nicolas Le Novère, Babraham Institute, UK.

# …but we know that scientists do not always use these.

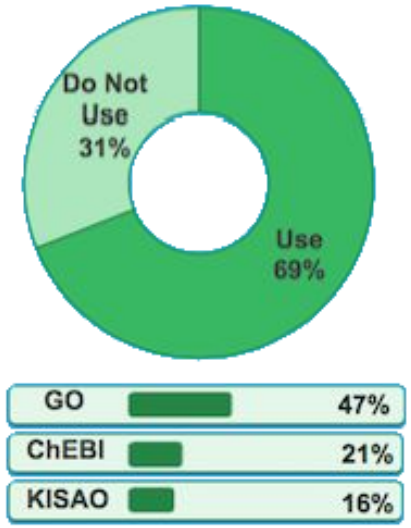
Format




Metadata




Ontologies




*top three most popular

The evolution of standards and data management practices in systems biology (2015). Stanford et al, Molecular Systems Biology, 11(12):851

# Tooling can help to reduce the barriers to implementing standards.

COPASI

RightField



libSBML

CellDesigner.org

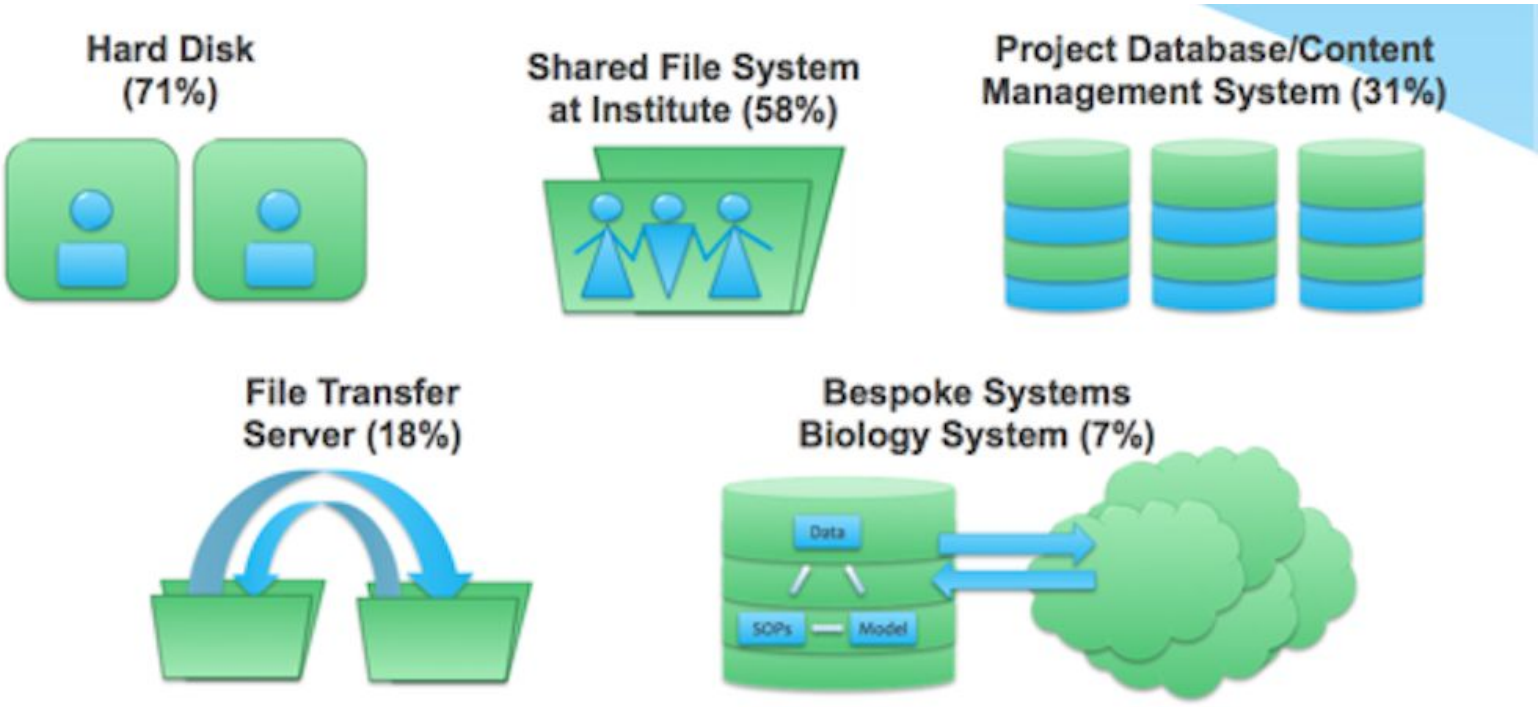
cellML

# But how about storage and citing?

**FINDABLE ??**

**ACCESSIBLE ??**

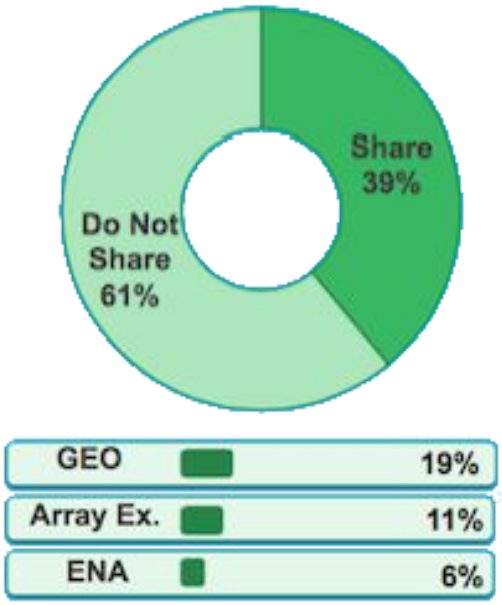
# Researchers tend to store their data on their own hard disk




The evolution of standards and data management practices in systems biology (2015). Stanford et al, Molecular Systems Biology, 11(12):851

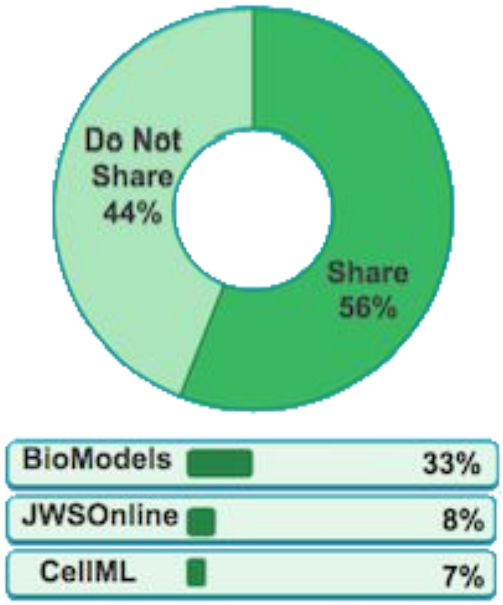
# Many researchers do not share their data in open repositories.

Data



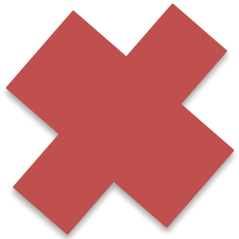

*top three most popular

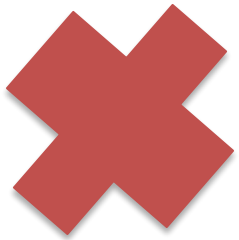
Models

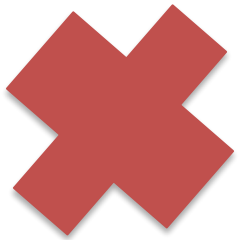



The evolution of standards and data management practices in systems biology (2015). Stanford et al, Molecular Systems Biology, 11(12):851

# This type of data management does not adhere to FAIR principles.

**FINDABLE** 

**ACCESSIBLE** 

**REUSABLE** 

# The FAIRDOM Platform


## Front end: Science Commons

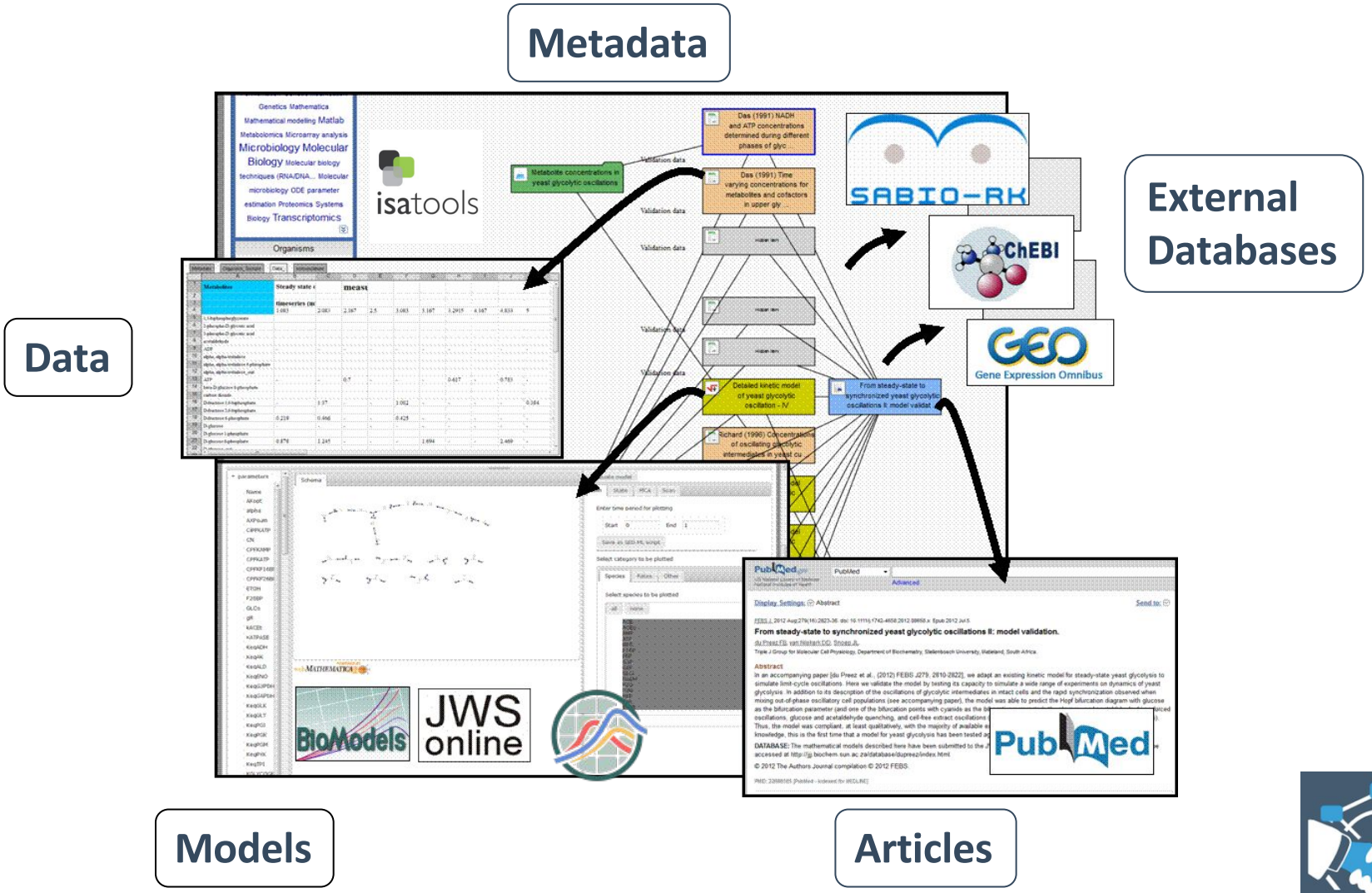
Web-based Cataloguing and Rich web interface for describing, finding, linking and promoting ongoing research and outcomes. Small files, aggregates across data archives.


## Back end: Scaled local LIMS and analytics

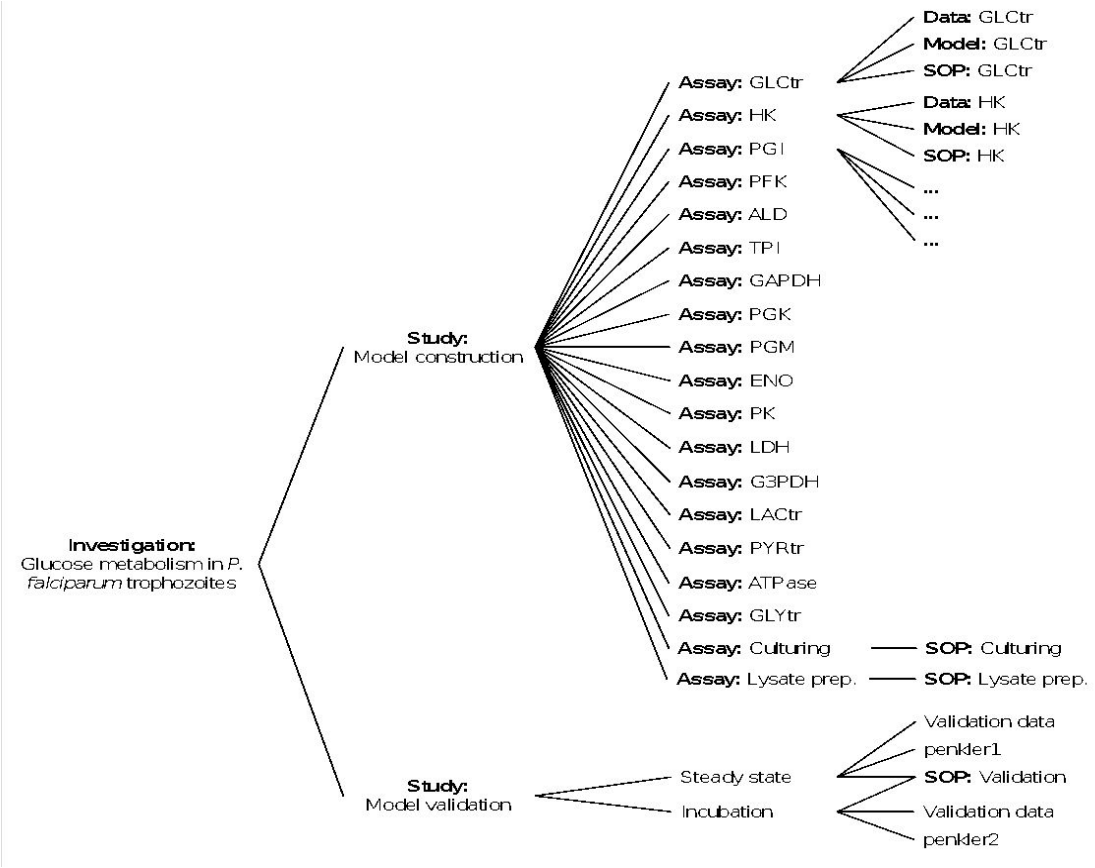
Extract, Transform and Load tooling direct from the instrumentation, data analysis pipelines. Automatic archiving. Handles large data.

# You can use SEEK as a local instance, or the FAIRDOMHub.

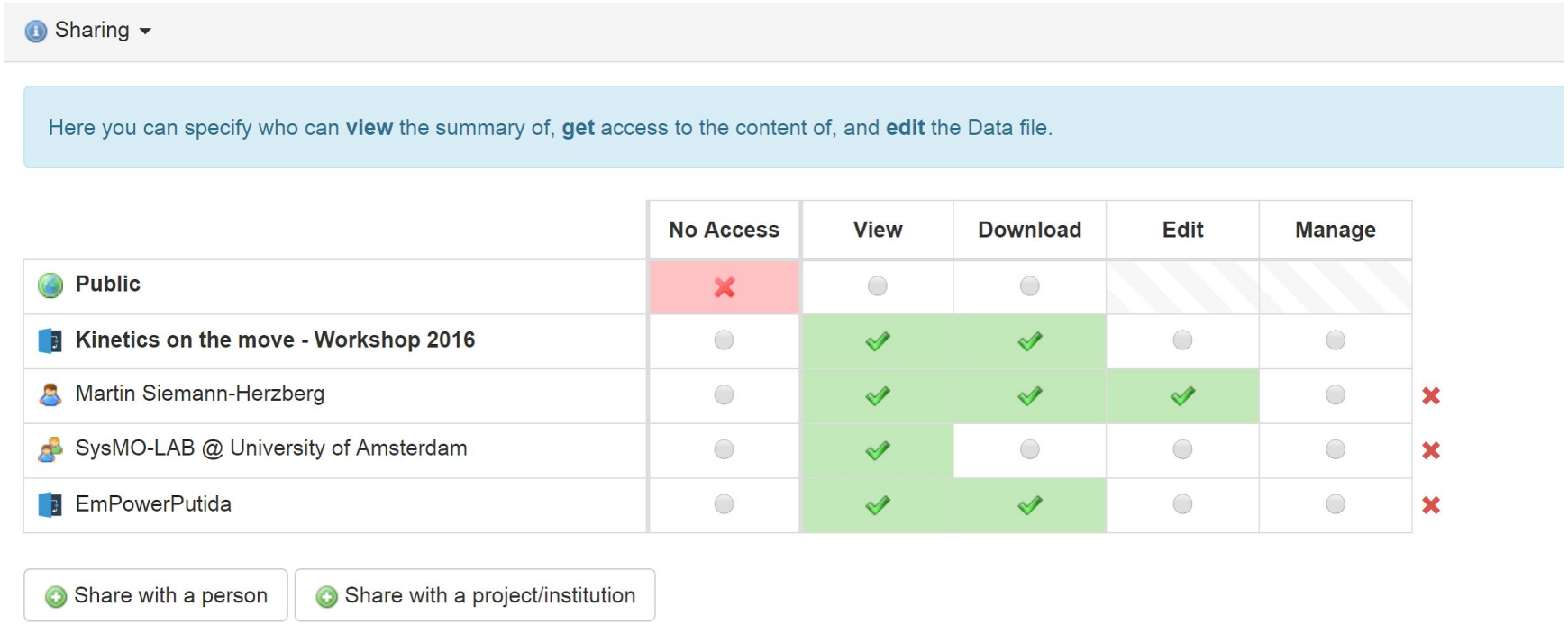

# SEEK aggregates as well as stores, so encourages domain specific publishing too.



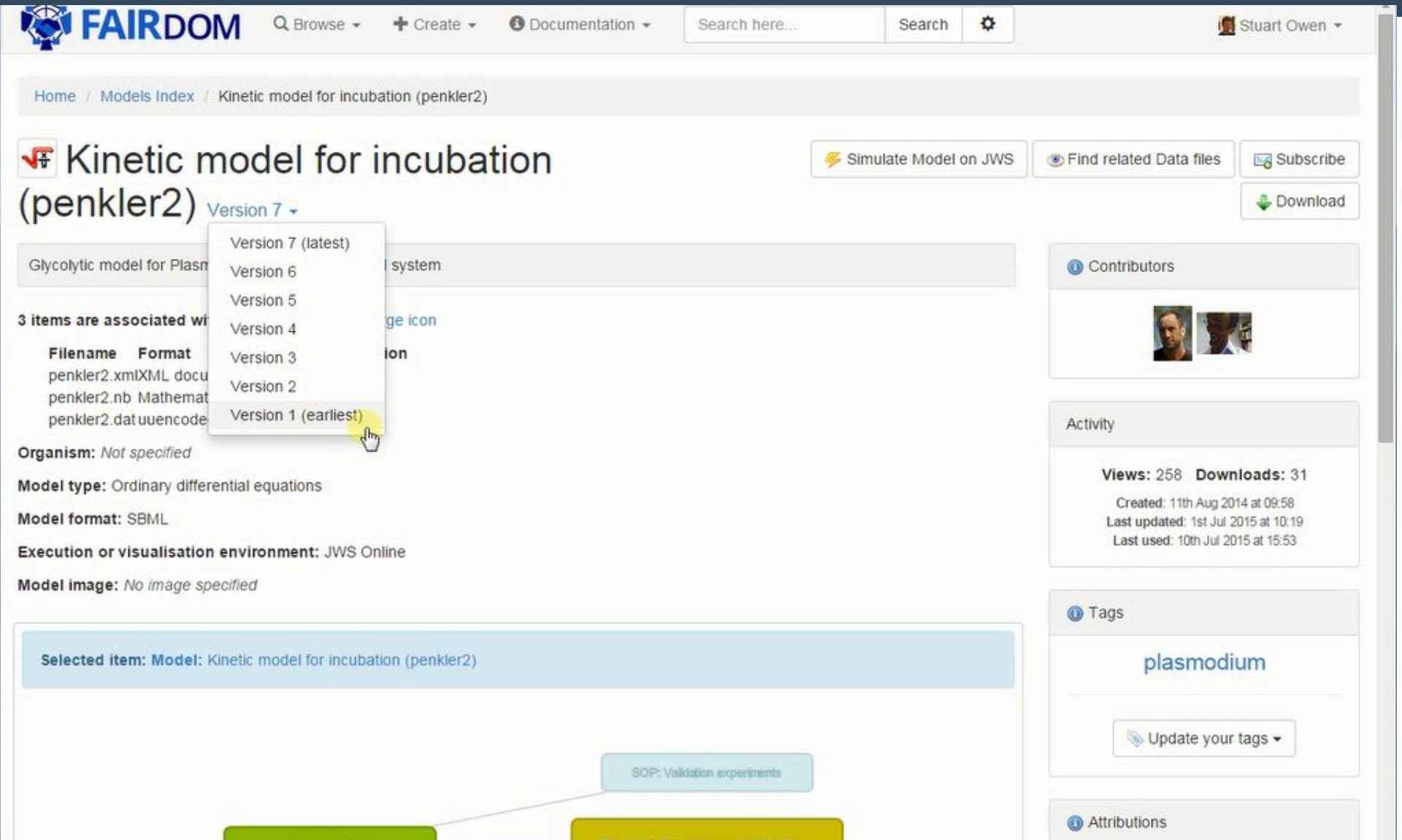
# It allows published work and all associated data and files to be organised in an ISA (Investigation, Study, Assay) format.



# Controlled sharing permissions.

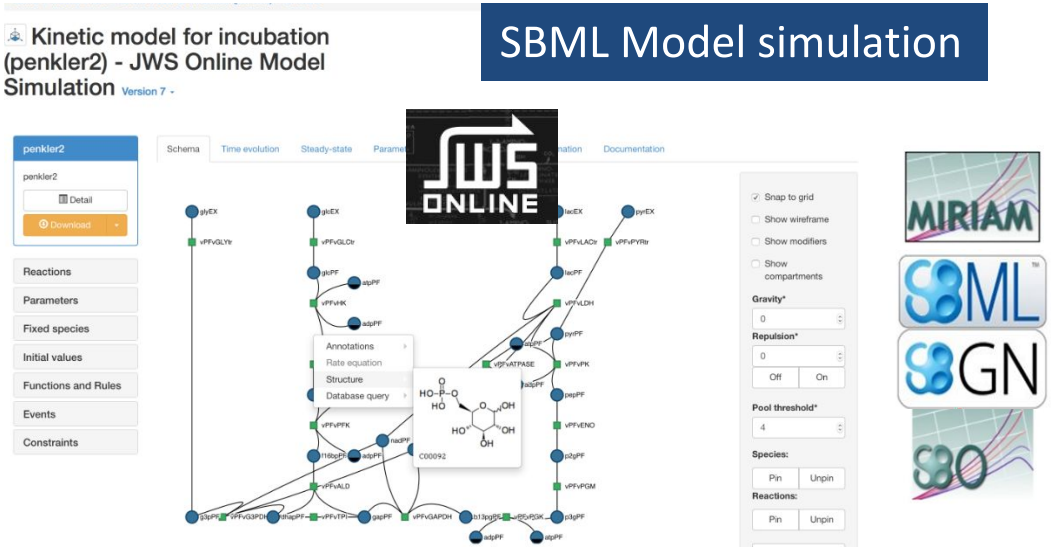


# Versioning


FAIRDOM
Browse
Create
Documentation
Search here...
Search
Stuart Owen
Home / Models Index / Kinetic model for incubation (penkler2)
Kinetic model for incubation (penkler2)
Version 7
Simulate Model on JWS
Find related Data files
Subscribe
Download
Version 7 (latest)
Version 6
Version 5
Version 4
Version 3
Version 2
Version 1 (earliest)
Glycolytic model for Plasm
system
3 items are associated wi
ge icon
Filename Format
ion
penkler2.xml XML docu
penkler2.nb Mathemat
penkler2.dat uuencode
Organism: Not specified
Model type: Ordinary differential equations
Model format: SBML
Execution or visualisation environment: JWS Online
Model image: No image specified
Selected item: Model: Kinetic model for incubation (penkler2)
SOP: Validation experiments
Contributors
Activity
Views: 258 Downloads: 31
Created: 11th Aug 2014 at 09:58
Last updated: 1st Jul 2015 at 10:19
Last used: 10th Jul 2015 at 15:53
Tags
plasmodium
Update your tags
Attributions

# … and integrated tooling

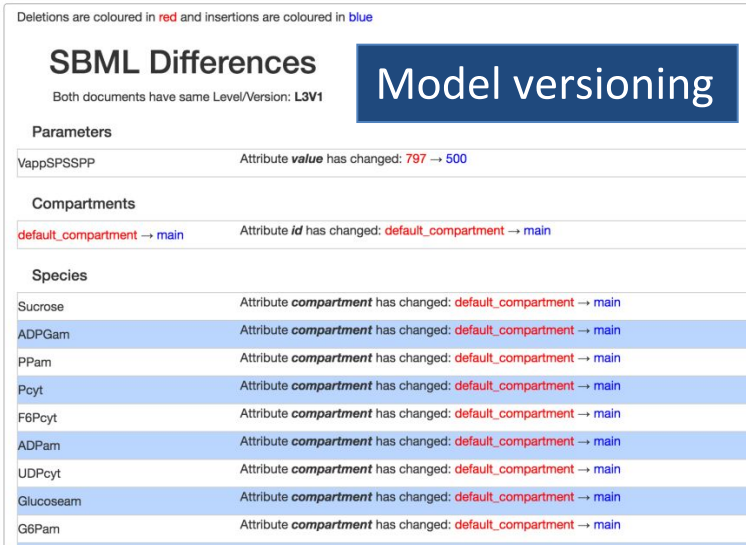
SBML Model simulation



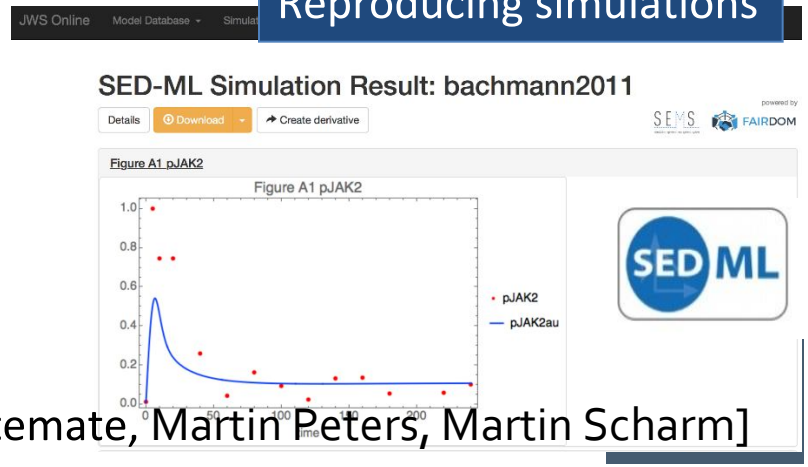




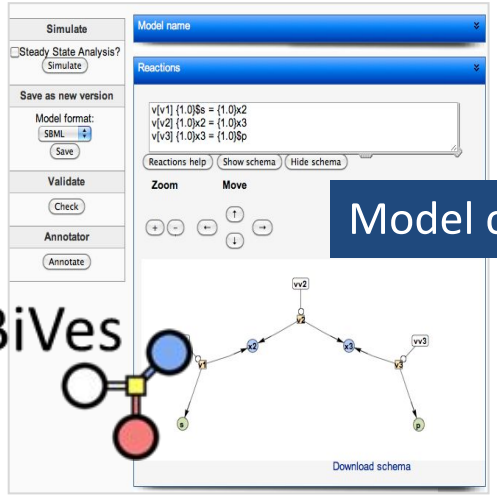
Model versioning




Reproducing simulations

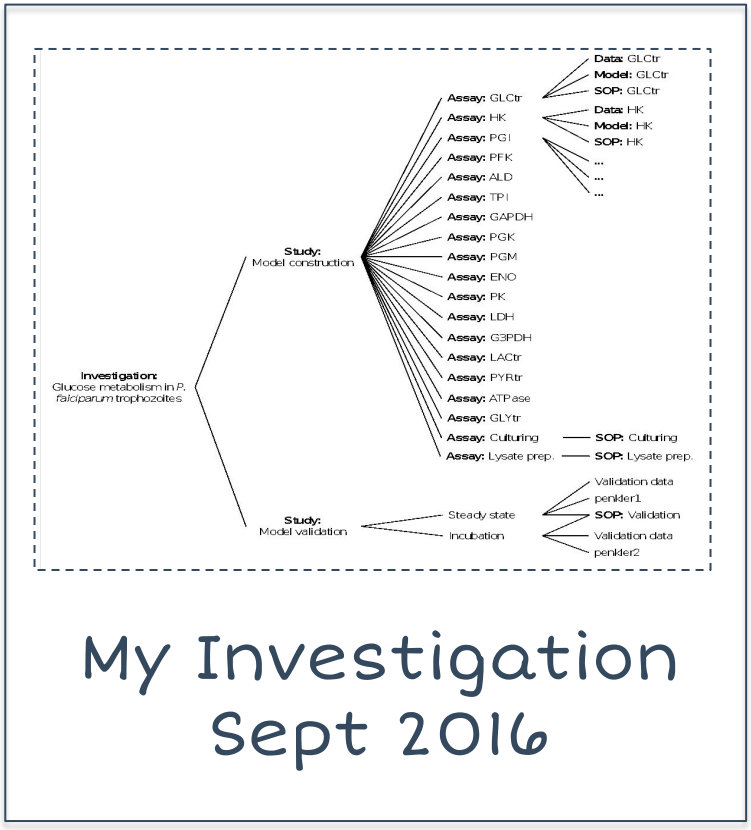



Model comparison

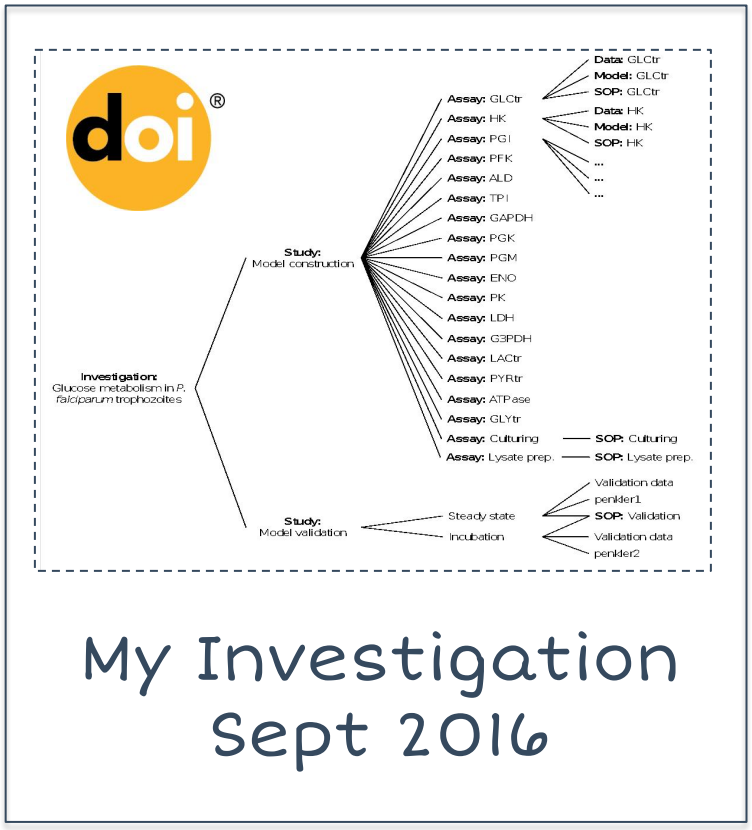



[Jacky Snoep, Dagmar Waltemate, Martin Peters, Martin Scharm]

# You can generate snapshots.

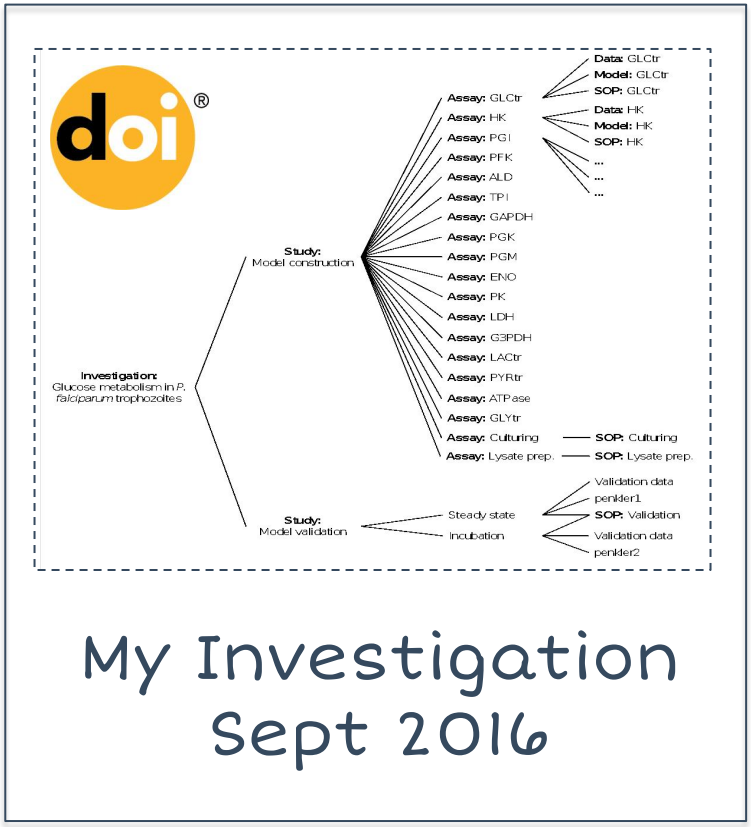


# And assign DOIs to snapshots.

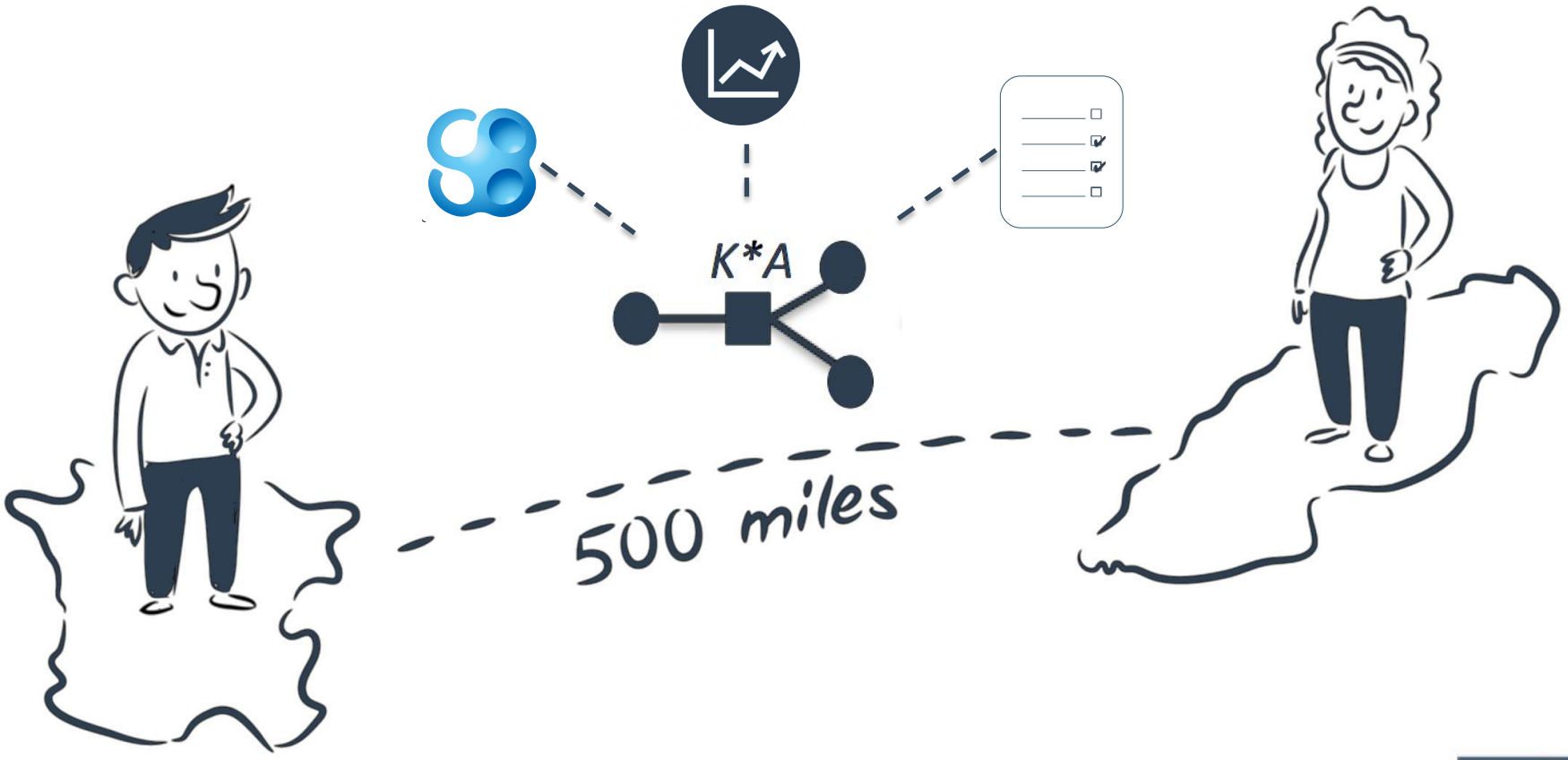



Author List: Joe Bloggs; Jane Doe
Title: My Investigation
Date: September 2016
DOI: https://doi.org/10.15490/seek##

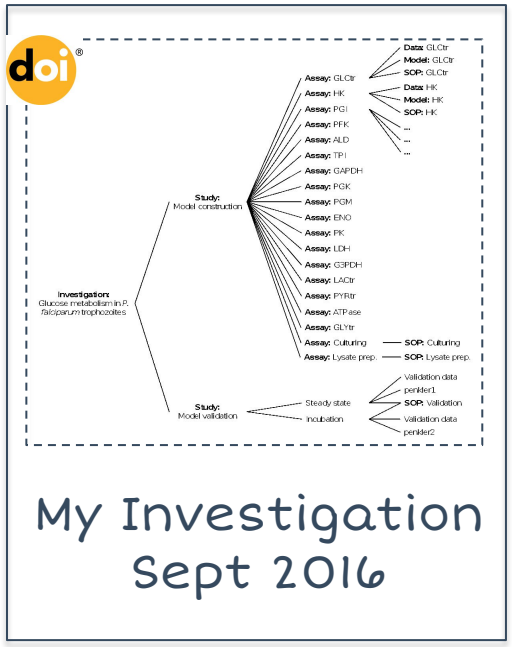
# The snapshot can be packaged as a Research Object




# This information then travels with the data and models



# Ensuring the original context of the data can always be understood.


FINDABLE ✔

ACCESSIBLE ✔

INTEROPERABLE ✔

REUSABLE ✔

# We are working and integrating with other initiatives to support this work.







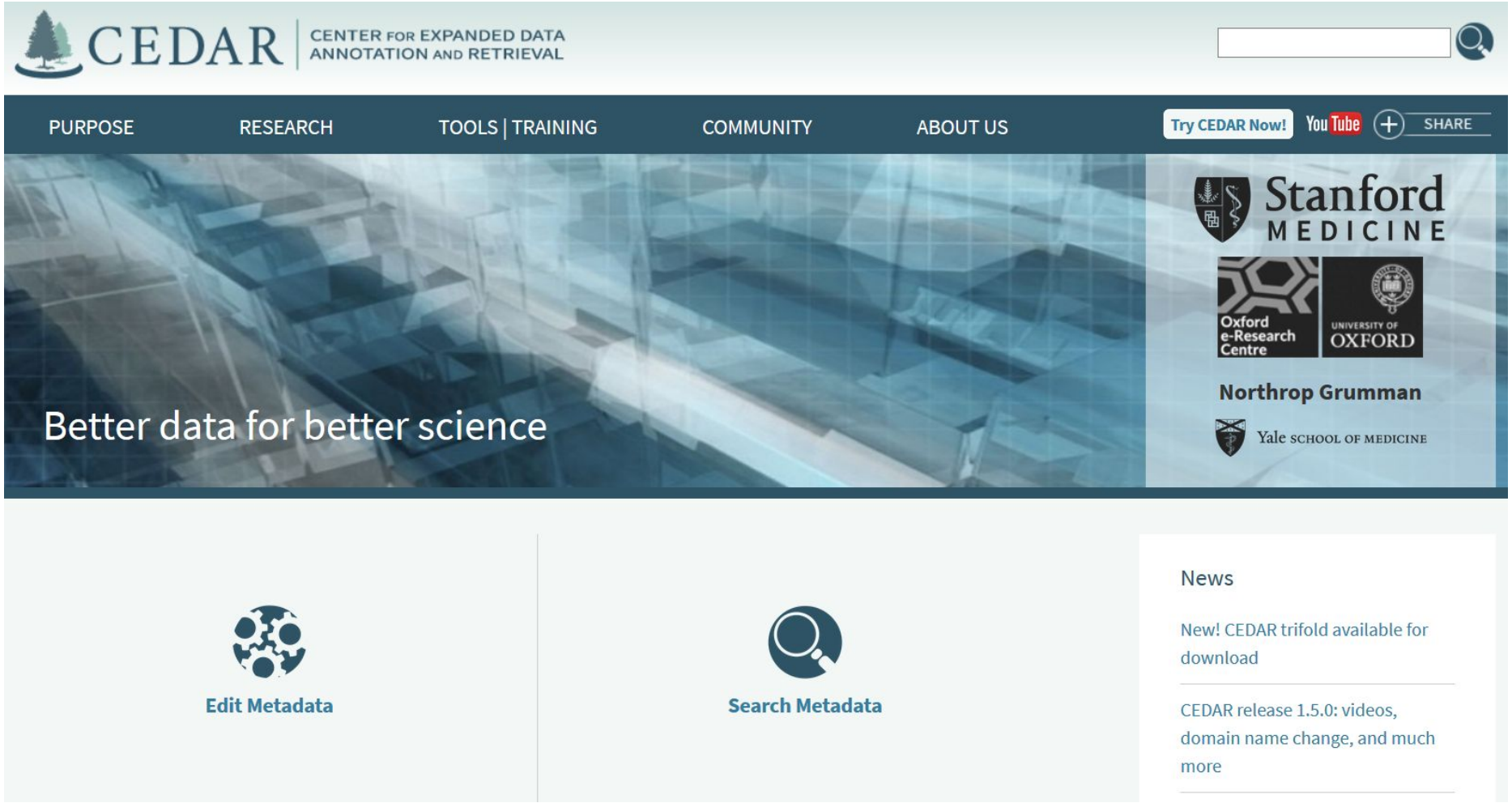


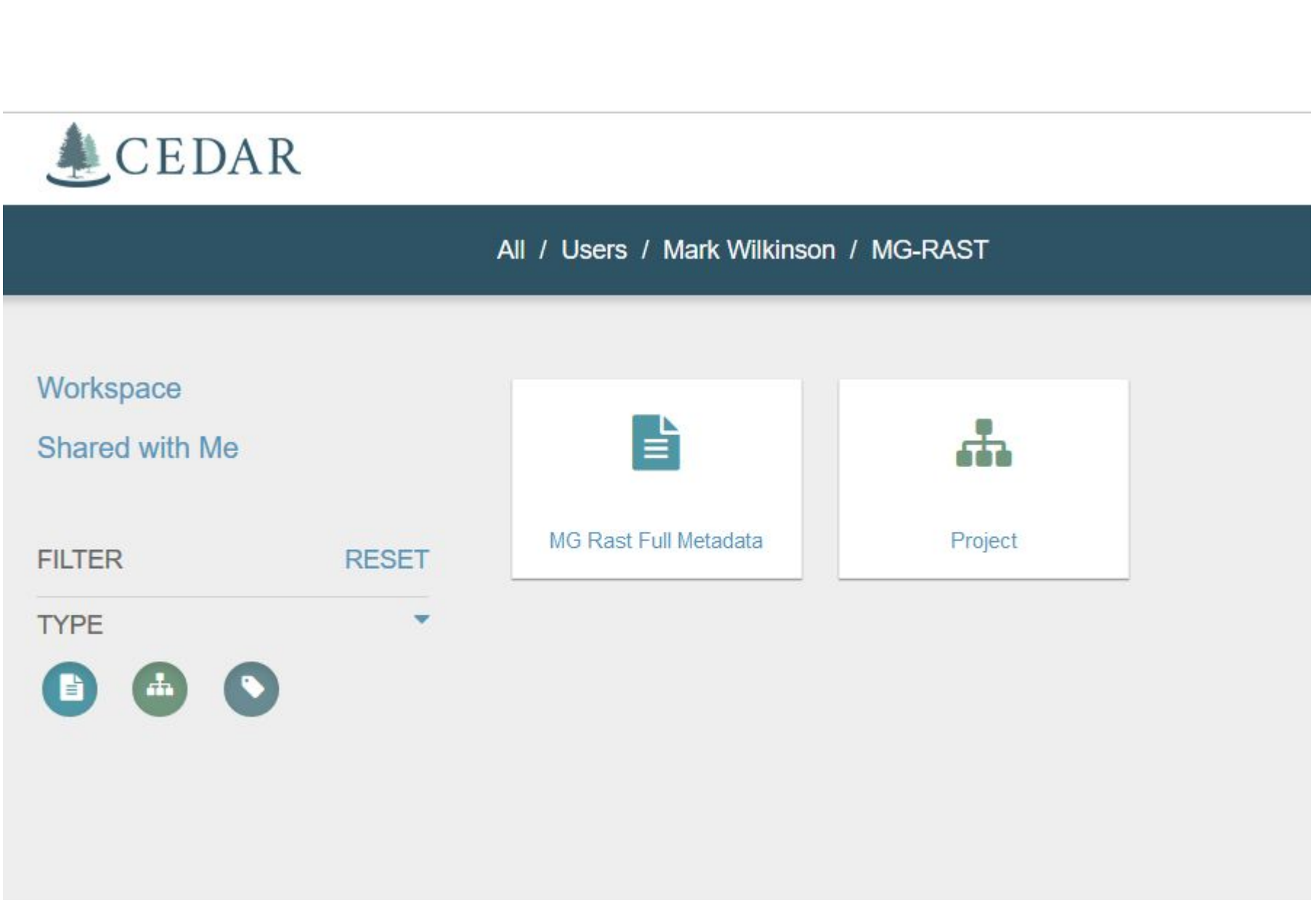

# FAIR Metadata

# CEDAR

Stanford Medical Center

http://metadatacenter.org




CEDAR
All / Users / Mark Wilkinson / MG-RAST
Workspace
Shared with Me
FILTER
RESET
TYPE
MG Rast Full Metadata
Project

Template Name

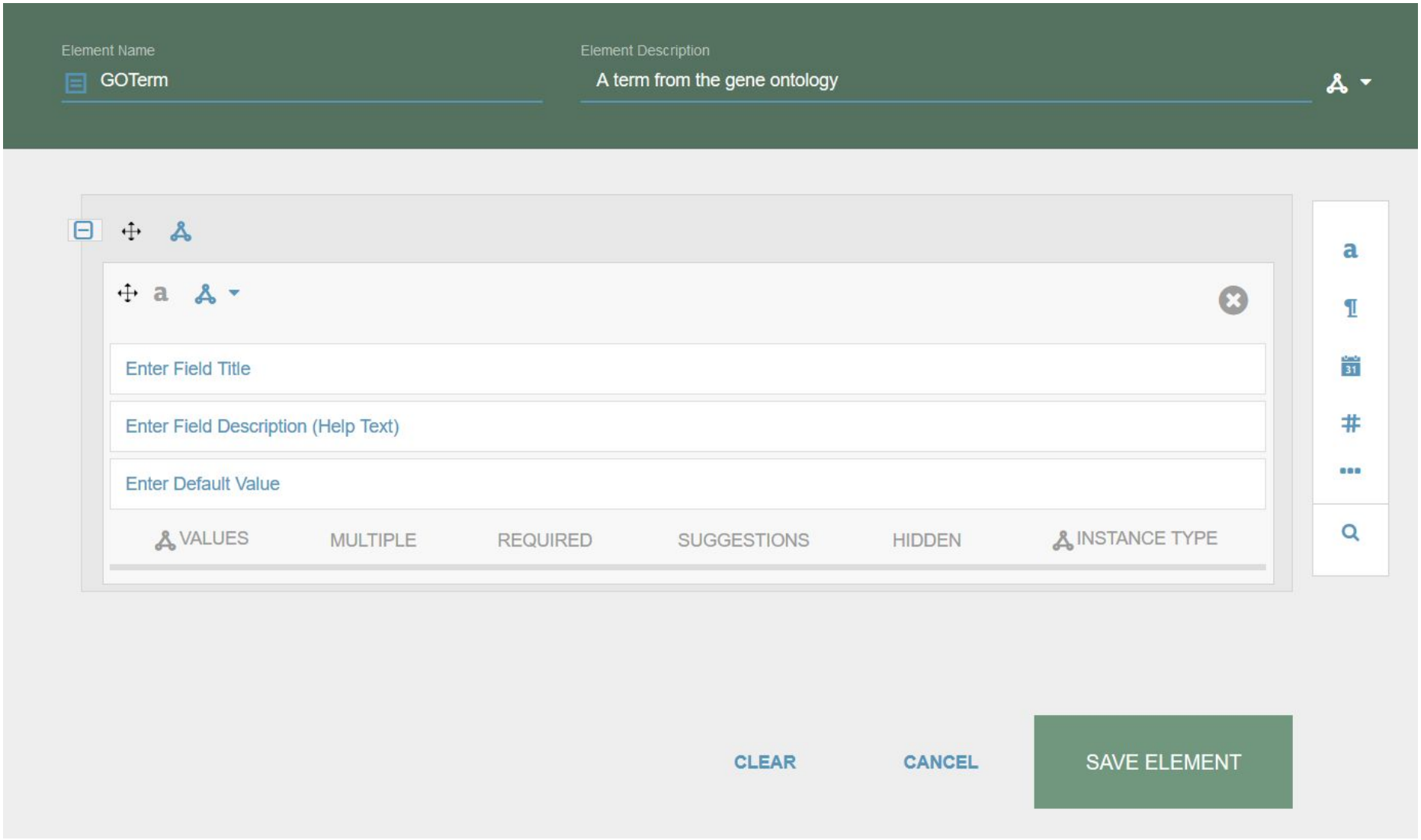
MG Rast Full Metadata

Template Description

All componenets of Metadata recommended for MG-RAST submissions

Project

- project name*
- project description*
- project funding
- project id
- PI email*
- PI first name*
- PI last name*
- PI Organization Name*
- PI Organization URI/URL*


Element Name
GOTerm
Element Description
A term from the gene ontology
Enter Field Title
Enter Field Description (Help Text)
Enter Default Value
VALUES
MULTIPLE
REQUIRED
SUGGESTIONS
HIDDEN
INSTANCE TYPE
CLEAR
CANCEL
SAVE ELEMENT

Find a property in BioPortal or **Enter Property URI**

*Start Over*

Search in BioPortal
has annotation

**Advanced Search Options**

**I want to...**

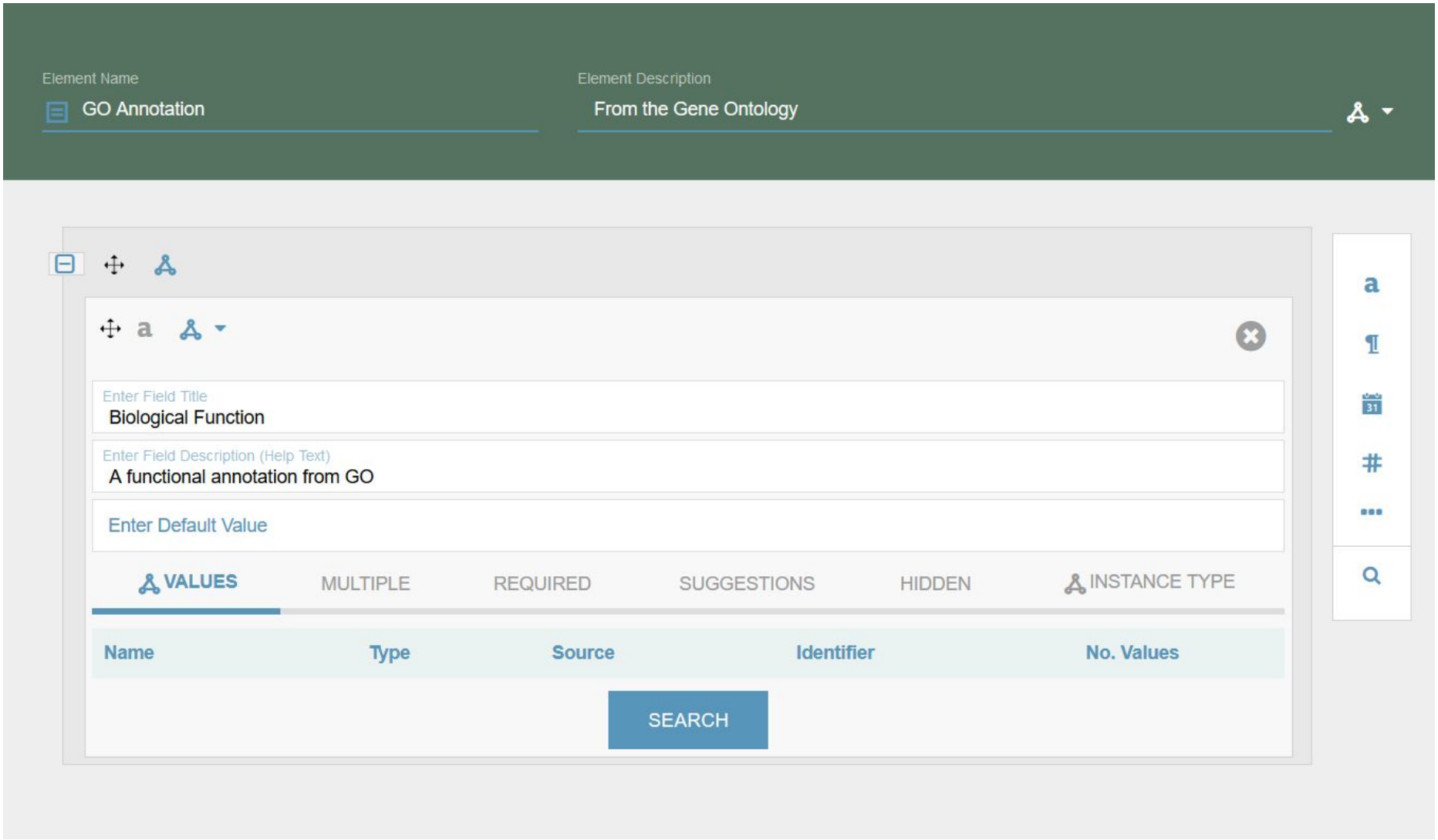
- (●) Search for a property in BioPortal
- ( ) Search for an ontology in BioPortal (e.g. OBI) and explore it

**Narrow your search to specific ontologies**

Add ontologies

493 results for the query 'has annotation'. Click on a property below to select it

| PROPERTY | DEFINITION | TYPE | SOURCE | ID |
|---|---|---|---|---|
| has annotation | has annotation is a relation between an entity and some textual anntotation. | Object Property | SIO | SIO_000255 |
| has annotation | - | Object Property | ORTH | SIO_000255 |
| has annotation | - | Object Property | BIOMO | SIO_000255 |
| has annotation | Links to additional | Object Property | GFVO | hasAnnotation |

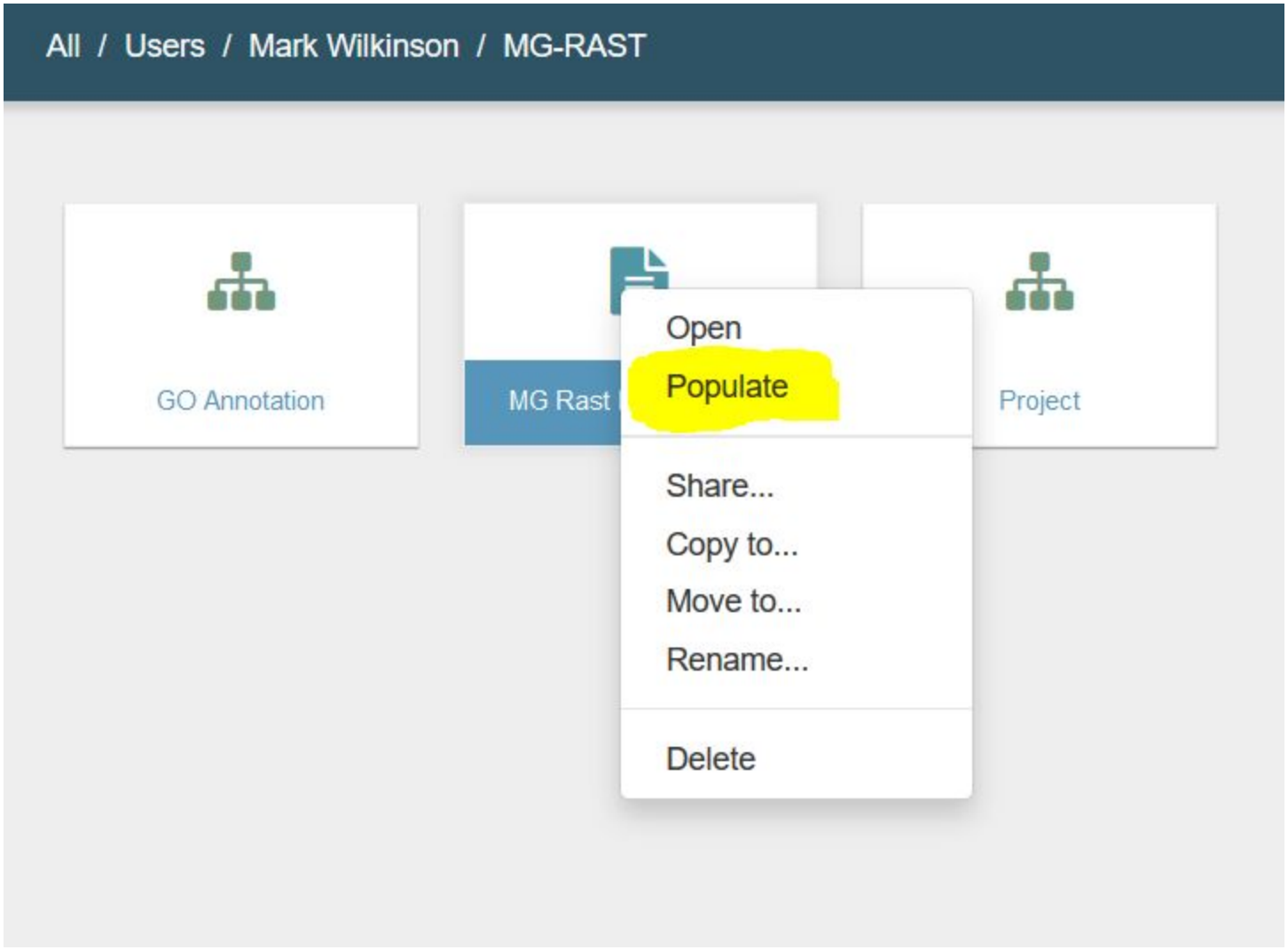

Element Name
GO Annotation
Element Description
From the Gene Ontology
Enter Field Title
Biological Function
Enter Field Description (Help Text)
A functional annotation from GO
Enter Default Value
VALUES
MULTIPLE
REQUIRED
SUGGESTIONS
HIDDEN
INSTANCE TYPE
Name
Type
Source
Identifier
No. Values
SEARCH

MG Rast Full Metadata

All componenets of Metadata recommended for MG-RAST submissions

Project

- project name*
- project description*
- project funding
- project id
- PI email*
- PI first name*
- PI last name*
- PI Organization Name*
- PI Organization URI/URL*
- PI Organization Country*


All / Users / Mark Wilkinson / MG-RAST
GO Annotation
MG Rast
Project
Open
Populate
Share...
Copy to...
Move to...
Rename...
Delete

MG Rast Full Metadata

Project

| | |
|---|---|
| * project name | An example project |
| * project description | To demonstrate CEDAR in Portugal |
| project funding | |
| project id | |
| * **PI email** | mark wilkinson@gmail |
| * PI first name | |
| * PI last name | |
| * PI Organization Name | |
| * PI Organization URI/URL | |
| * PI Organization Country | |
| PI Organization Address | |
| technical contact email | |
| technical contact name | |

OK or Ctrl-Enter

* PI Organization Country

**PI Organization Address**

technical contact email

technical contact name

CANCEL

SAVE

RDF

```
_:b0 <https://schema.metadatacenter.net/properties/project> _:b1 .
_:b1 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.bioontology.org/ontologies/EDAM/classes/http%3A%2F%2Fedamontolc
_:b1 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.bioontology.org/ontologies/JERM/classes/http%3A%2F%2Fwww.mygric
_:b1 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.bioontology.org/ontologies/SIO/classes/http%3A%2F%2Fsemanticsci
_:b1 <https://schema.metadatacenter.net/properties/pIEmail> "mark wilkinson@gmail" .
_:b1 <https://schema.metadatacenter.net/properties/pIOrganizationCountry> _:b2 .
_:b1 <https://schema.metadatacenter.net/properties/projectDescription> "To demonstrate CEDAR in Portugal" .
_:b1 <https://schema.metadatacenter.net/properties/projectName> "An example project" .
```

JSON-LD

PI Organization URI/URL

* PI Organization Country

**PI Organization Address**

technical contact email

technical contact name

CANCEL

SAVE

RDF

JSON-LD


```
{
  "@context": {
    "schema:description": {
      "@type": "xsd:string"
    },
    "rqwerqwe": "https://schema.metadatacenter.net/properties/rqwerqwe",
    "pav": "http://purl.org/pav/",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "xsd": "http://www.w3.org/2001/XMLSchema#",
    "project": "https://schema.metadatacenter.net/properties/project",
    "schema:name": {
      "@type": "xsd:string"
    },
    "schema": "http://schema.org/",
    "oslc:modifiedBy": {
```

Data and templates available via REST API

“All or nothing”

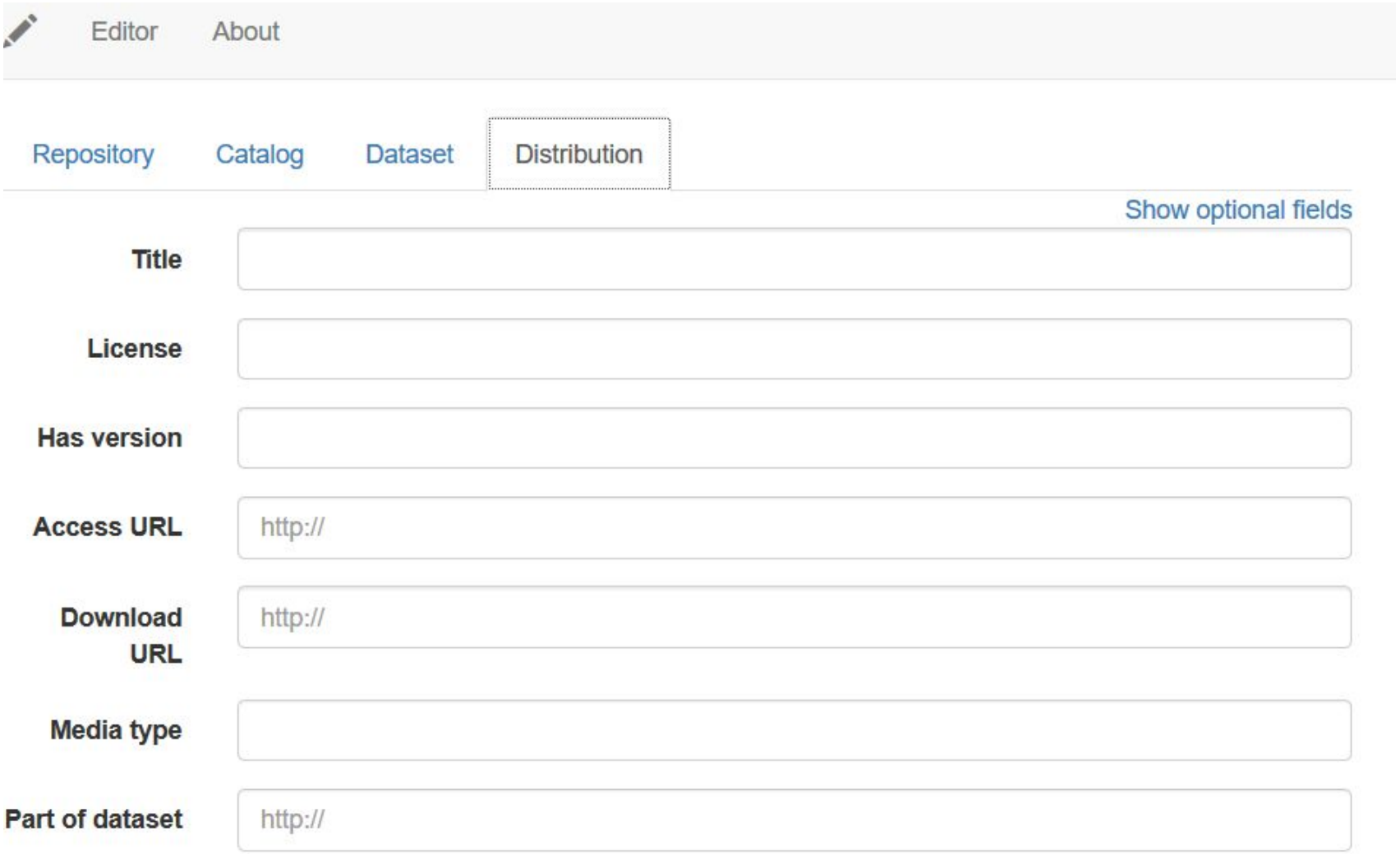
- must install locally for private data;
- individual components not easy to split

# FAIR Metadata

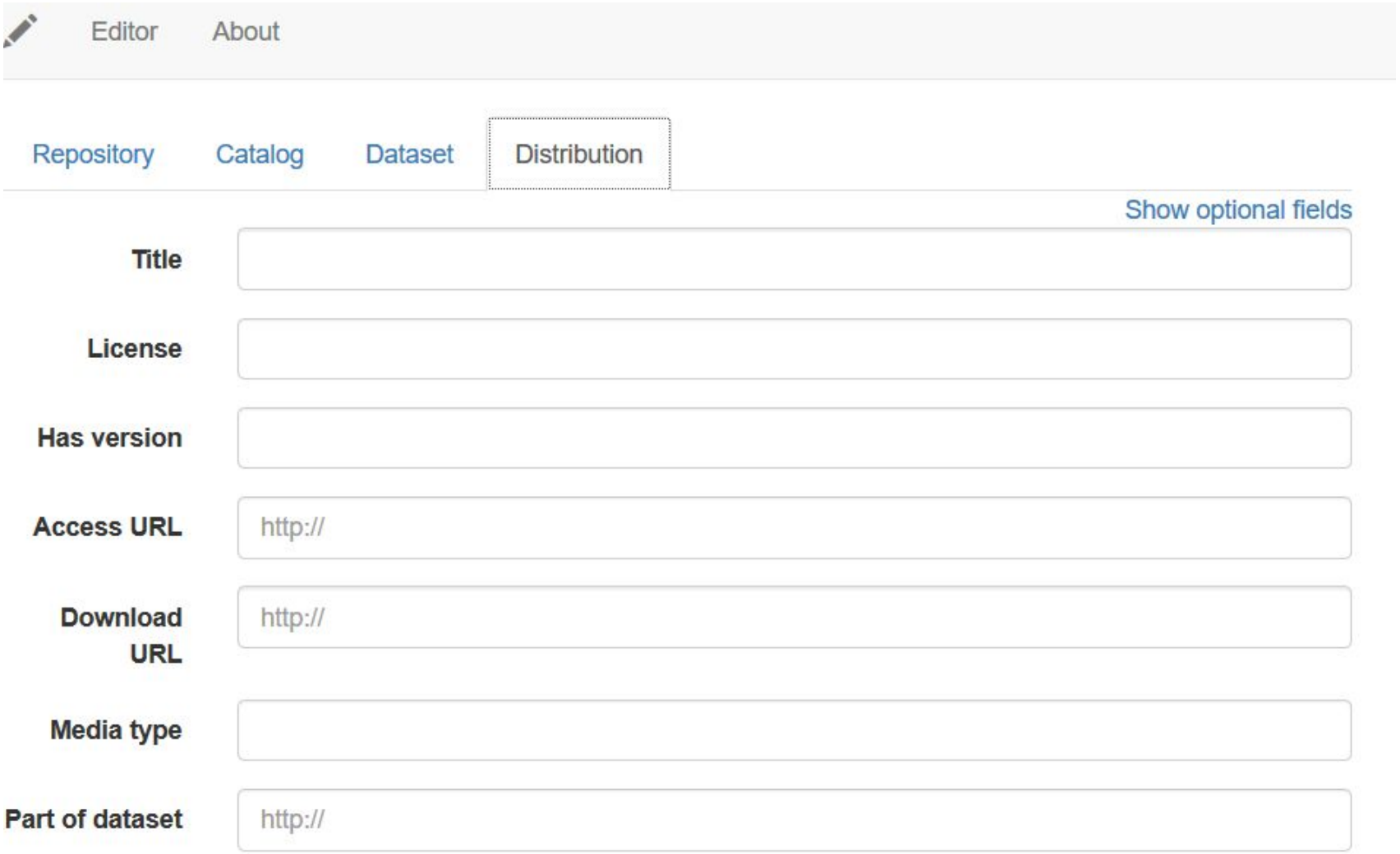
# FAIR Accessor
# FAIR Metadata Editor

From my lab and DTL Netherlands

# https://fair-demo.fair-dtls.surf-hosted.nl/editor/#!/



# https://fair-demo.fair-dtls.surf-hosted.nl/editor/#!/



Build | Share

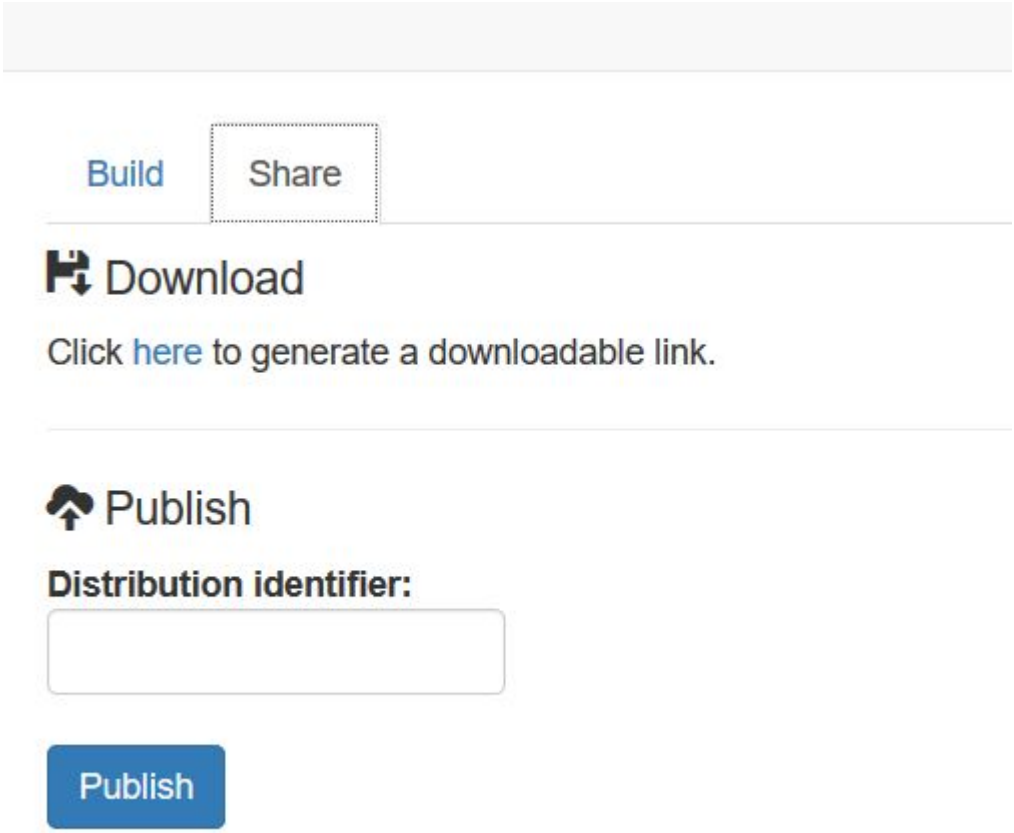
Select a field to read more about it.

## RDF preview

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix dct: <http://purl.org/dc/terms/>.
@prefix dcat: <http://www.w3.org/ns/dcat#>.
@prefix fdp: <http://rdf.biosemantics.org/ontologies/fdp-o#>.
@prefix datacite: <http://purl.org/spar/datacite/>.

<> rdf:type dcat:Distribution.
```

https://fair-demo.fair-dtls.surf-hosted.nl/editor/#!/



swagger default (/v2/api-docs) Explore

# FDP API Java based

This API is a prototype version, If you find bugs in this api please contact the developer.API specsSource code

Created by rr.kaliyaperumal@gmail.com
The MIT License

## metadata-controller : FDP metadata

Show/Hide | List Operations | Expand Operations

| | | |
|---|---|---|
| GET | / | FDP metadata |
| PATCH | / | Update fdp metadata |
| POST | /catalog | POST catalog metadata |
| GET | /catalog/{id} | Catalog metadata |
| POST | /dataset | POST dataset metadata |
| GET | /dataset/{id} | Dataset metadata |
| POST | /distribution | POST distribution metadata |
| GET | /distribution/{id} | Dataset distribution metadata |

[ BASE URL: /fdp , API VERSION: 0.1-beta ]

# My repository on CPAN

## Create FAIR Accessors entirely in code
## (the examples from the talk this morning all used this)

# FAIR Data

# RightField

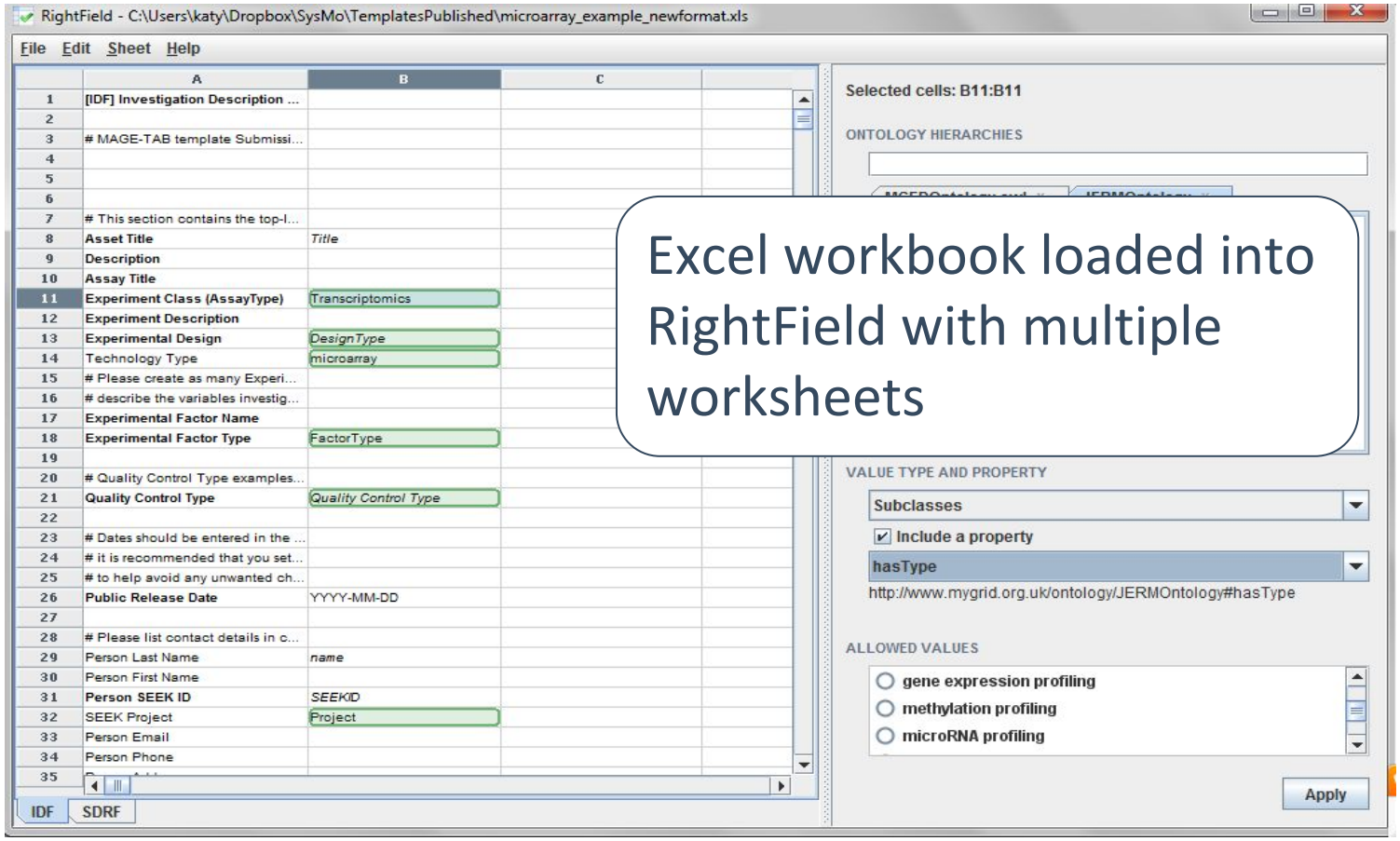
FAIRDOM Project


FAIRIFIER
A power tool for working with messy data.
Create Project
Open Project
Import Project
Language Settings
Version Alpha [TRUNK]
Help
About
« Start Over
Configure Parsing Options
Project name dataoutput csv
Create Project »
Entry
Entry name
Status
Protein names
Gene names
Organism
Length
1. P06213 INSR_HUMAN reviewed Insulin receptor (IR) (EC 2.7.10.1) (CD antigen CD220) [Cleaved into: Insulin receptor subunit alpha; Insulin receptor subunit beta] INSR Homo sapiens (Human) 1382
2. P51787 KCNQ1_HUMAN reviewed Potassium voltage-gated channel subfamily KQT member 1 (IKs producing slow voltage-gated potassium channel subunit alpha KvLQT1) (KQT-like 1) (Voltage-gated potassium channel subunit Kv7.1) KCNQ1 KCNA8 KCNA9 KVLQT1 Homo sapiens (Human) 676
3. P10997 IAPP_HUMAN reviewed Islet amyloid polypeptide (Amylin) (Diabetes-associated peptide) (DAP) (Insulinoma amyloid peptide) IAPP Homo sapiens (Human) 89
4. Q14654 KCJ11_HUMAN reviewed ATP-sensitive inward rectifier potassium channel 11 (IKATP) (Inward rectifier K(+) channel Kir6.2) (Potassium channel, inwardly rectifying subfamily J member 11) KCNJ11 Homo sapiens (Human) 390
Parse data as
CSV / TSV / separator-based files
Line-based text files
Fixed-width field text files
PC-Axis text files
JSON files
MARC files
RDF/N3 files
XML files
Open Document Format
Character encoding
Update Preview
Columns are separated by
commas (CSV)
tabs (TSV)
custom \t
Escape special characters with \
Ignore first 0 line(s) at beginning of file
Parse next 1 line(s) as column headers
Discard initial 0 row(s) of data
Load at most 0 row(s) of data
Parse cell text into numbers, dates, ...
Quotation marks are used to enclose cells containing column separators
Store blank rows
Store blank cells as nulls
Store file source (file names, URLs) in each row

# We provide Tooling for annotating spreadsheets.

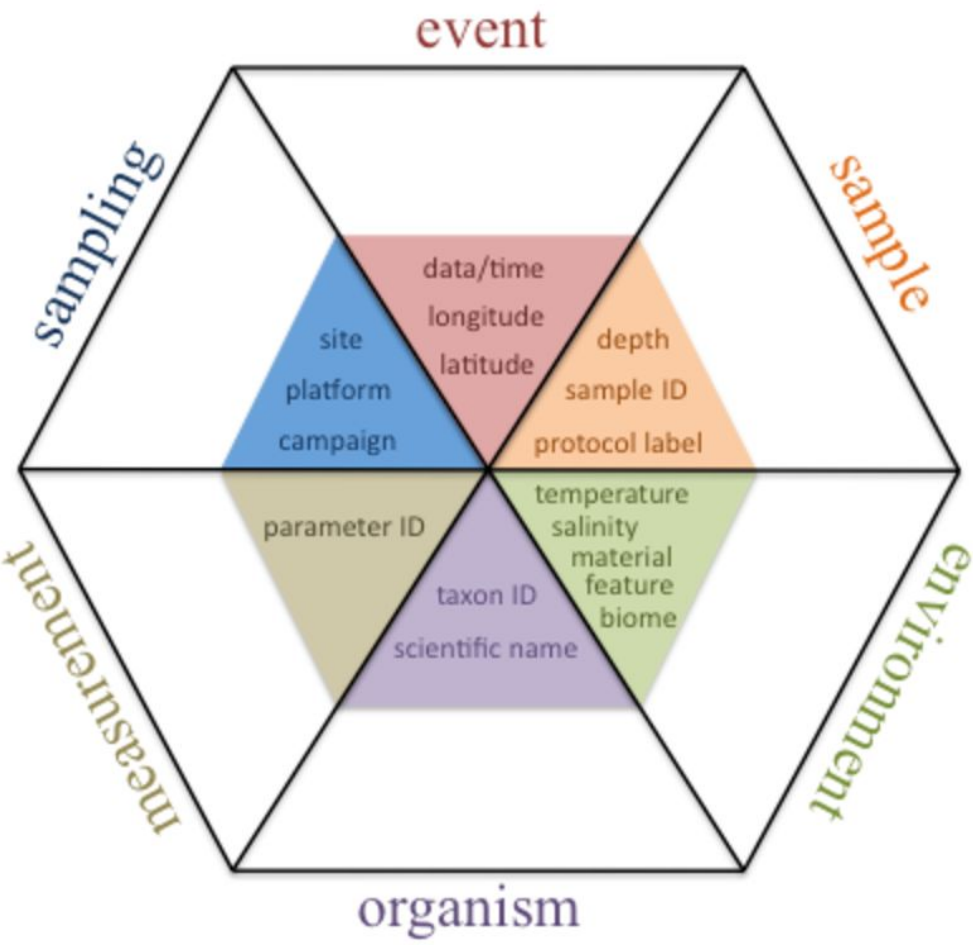
**RightField**

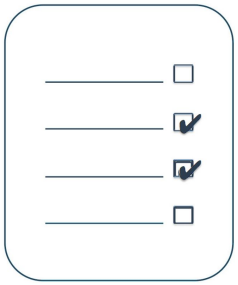
# You can use it to generate templates for different types of assay data.




Excel workbook loaded into RightField with multiple worksheets

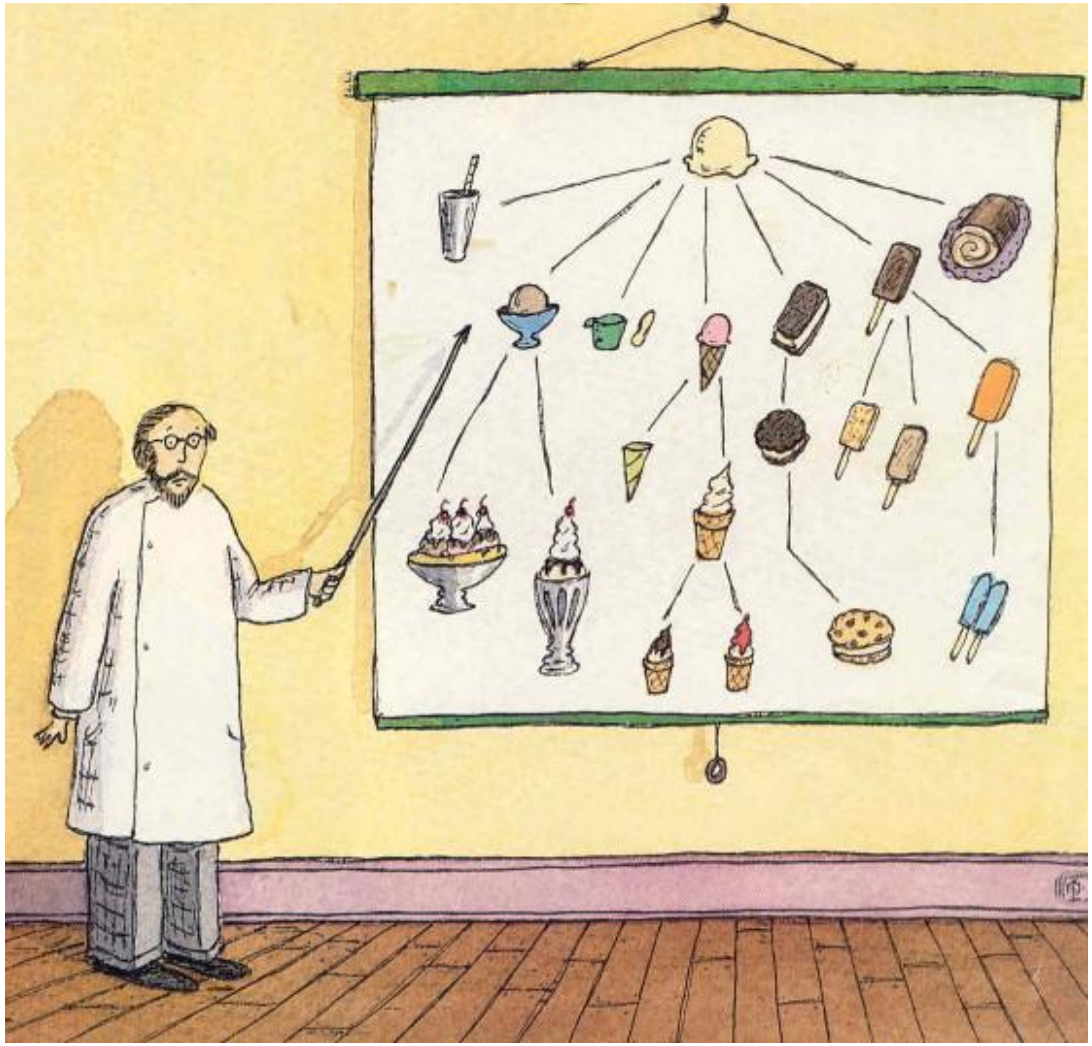
# We advise on checklists of information that assist with understanding and re-use of data



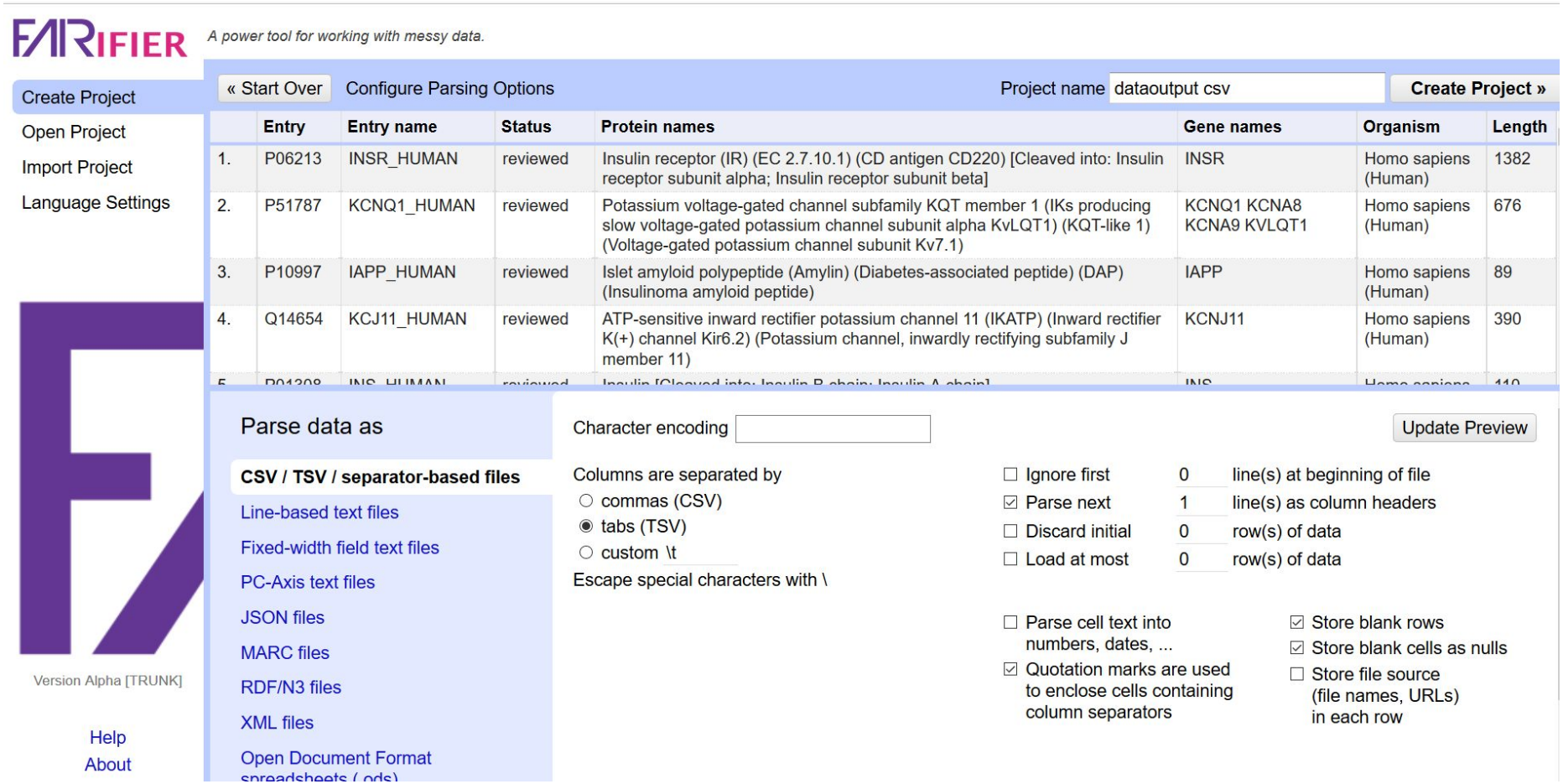
Taken from the Micro B3 Checklist in the European Nucleotide Archive (2016)

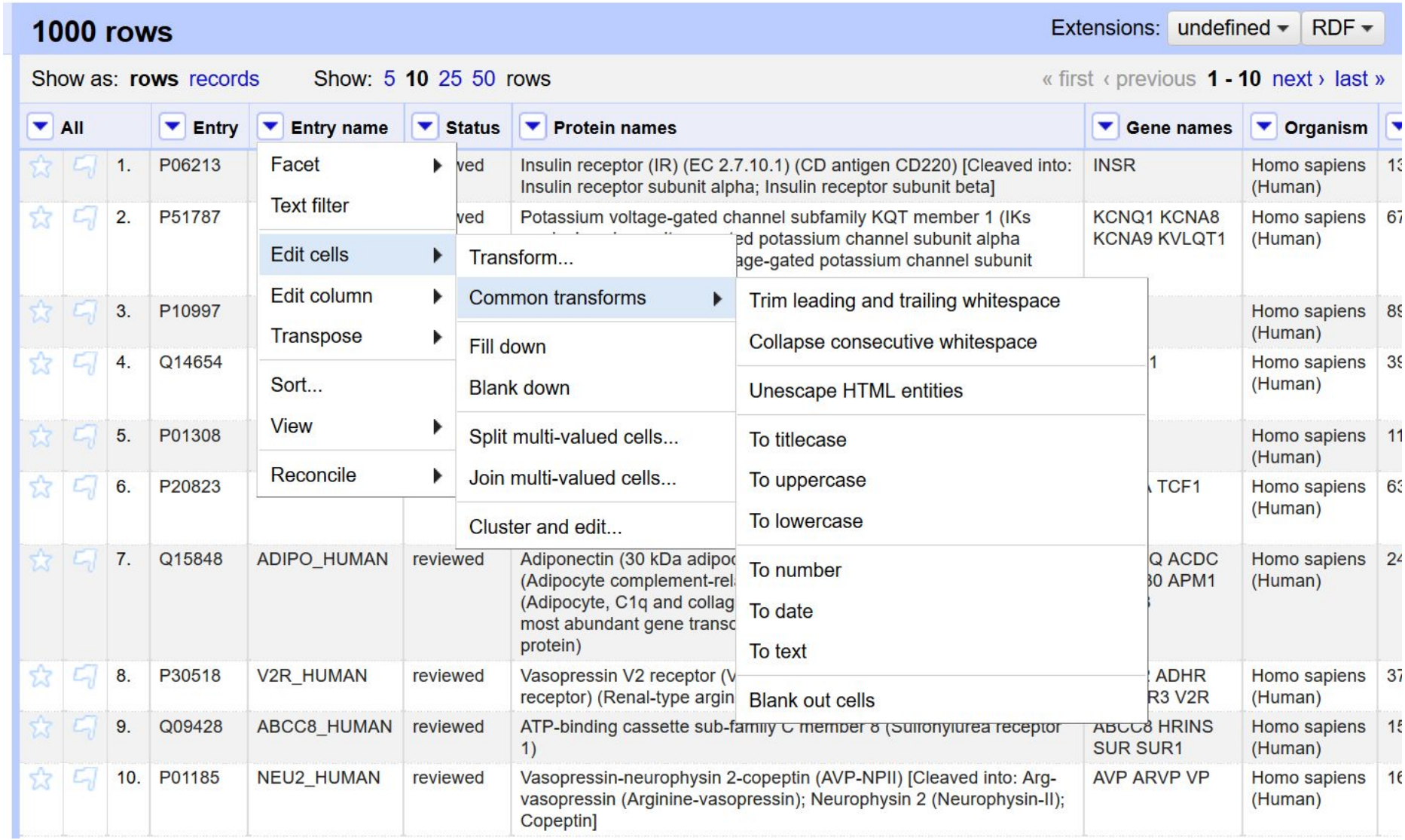
# And Ontologies that can be used in conjunction with checklists



# FAIRifyer

DTL Netherlands


FAIRIFIER
A power tool for working with messy data.
Create Project
Open Project
Import Project
Language Settings
Version Alpha [TRUNK]
Help
About
« Start Over
Configure Parsing Options
Project name
dataoutput csv
Create Project »
Entry
Entry name
Status
Protein names
Gene names
Organism
Length
1. P06213 INSR_HUMAN reviewed Insulin receptor (IR) (EC 2.7.10.1) (CD antigen CD220) [Cleaved into: Insulin receptor subunit alpha; Insulin receptor subunit beta] INSR Homo sapiens (Human) 1382
2. P51787 KCNQ1_HUMAN reviewed Potassium voltage-gated channel subfamily KQT member 1 (IKs producing slow voltage-gated potassium channel subunit alpha KvLQT1) (KQT-like 1) (Voltage-gated potassium channel subunit Kv7.1) KCNQ1 KCNA8 KCNA9 KVLQT1 Homo sapiens (Human) 676
3. P10997 IAPP_HUMAN reviewed Islet amyloid polypeptide (Amylin) (Diabetes-associated peptide) (DAP) (Insulinoma amyloid peptide) IAPP Homo sapiens (Human) 89
4. Q14654 KCJ11_HUMAN reviewed ATP-sensitive inward rectifier potassium channel 11 (IKATP) (Inward rectifier K(+) channel Kir6.2) (Potassium channel, inwardly rectifying subfamily J member 11) KCNJ11 Homo sapiens (Human) 390
Parse data as
CSV / TSV / separator-based files
Line-based text files
Fixed-width field text files
PC-Axis text files
JSON files
MARC files
RDF/N3 files
XML files
Open Document Format
Character encoding
Update Preview
Columns are separated by
commas (CSV)
tabs (TSV)
custom \t
Escape special characters with \
Ignore first 0 line(s) at beginning of file
Parse next 1 line(s) as column headers
Discard initial 0 row(s) of data
Load at most 0 row(s) of data
Parse cell text into numbers, dates, ...
Quotation marks are used to enclose cells containing column separators
Store blank rows
Store blank cells as nulls
Store file source (file names, URLs) in each row


1000 rows
Extensions: undefined RDF
Show as: rows records
Show: 5 10 25 50 rows
« first ‹ previous 1 - 10 next › last »
All
Entry
Entry name
Status
Protein names
Gene names
Organism
Facet
Text filter
Edit cells
Edit column
Transpose
Sort...
View
Reconcile
Transform...
Common transforms
Fill down
Blank down
Split multi-valued cells...
Join multi-valued cells...
Cluster and edit...
Trim leading and trailing whitespace
Collapse consecutive whitespace
Unescape HTML entities
To titlecase
To uppercase
To lowercase
To number
To date
To text
Blank out cells
1. P06213 Insulin receptor (IR) (EC 2.7.10.1) (CD antigen CD220) [Cleaved into: Insulin receptor subunit alpha; Insulin receptor subunit beta] INSR Homo sapiens (Human)
2. P51787 Potassium voltage-gated channel subfamily KQT member 1 (IKs KCNQ1 KCNA8 KCNA9 KVLQT1 Homo sapiens (Human)
3. P10997 Homo sapiens (Human)
4. Q14654 Homo sapiens (Human)
5. P01308 Homo sapiens (Human)
6. P20823 TCF1 Homo sapiens (Human)
7. Q15848 ADIPO_HUMAN reviewed Adiponectin (30 kDa adipoc (Adipocyte complement-rel (Adipocyte, C1q and collag most abundant gene transc protein) Q ACDC 30 APM1 Homo sapiens (Human)
8. P30518 V2R_HUMAN reviewed Vasopressin V2 receptor (V receptor) (Renal-type argin ADHR R3 V2R Homo sapiens (Human)
9. Q09428 ABCC8_HUMAN reviewed ATP-binding cassette sub-family C member 8 (Sulfonylurea receptor 1) ABCC8 HRINS SUR SUR1 Homo sapiens (Human)
10. P01185 NEU2_HUMAN reviewed Vasopressin-neurophysin 2-copeptin (AVP-NPII) [Cleaved into: Arg-vasopressin (Arginine-vasopressin); Neurophysin 2 (Neurophysin-II); Copeptin] AVP ARVP VP Homo sapiens (Human)

**Base URI:** https://fair-demo.fair-dtls.surf-hosted.nl/ edit

**Semantic model** | RDF Preview

Available Prefixes: add prefix | manage prefixes

**(row index)** URI

add rdf:type

- property? → **Entry** cell
- property? → **Entry name** cell
- property? → **Status** cell
- property? → **Protein names** cell
- property? → **Gene names** cell
- property? → **Organism** cell
- property? → **Length** cell

add property

# RDF Node

## Use content from cell...

- ○ (row index)
- ◉ Entry
- ○ Entry name
- ○ Status
- ○ Protein names
- ○ Gene names
- ○ Organism
- ○ Length
- ○ Constant Value

## The cell's content is used ...

- ○ as a URI
- ◉ as text
- ○ as language-tagged text
- ○ as integer number
- ○ as non-integer number
- ○ as date (YYYY-MM-DD)
- ○ as dateTime (YYYY-MM-DD HH:MM:SS)
- ○ as boolean
- ○ as custom datatype (specify type URI)
- ○ as a blank node

## Use custom expression...

```
value
```

preview/edit

OK Cancel

# FAIR Data

# FAIR Projector

# FAIR Projector Builder

Please enter the URL of the tab-delimited file

https://zenodo.org/record/8

Build FAIR Projector

Put the URL for the Zenodo data here

# FAIR Data

# D2RQ Mapper
# (DCBLS Japan)

# http://d2rq.dbcls.jp/

D2RQ Mapper


## Welcome to D2RQ Mapper

To fully utilize data stored in the RDB (relational database) in the Semantic Web context while keeping the database content up-to-date, it is ideal to have a middleware to map RDB datasets to those of RDF (Resource Description Framework) graph and to

## FAIR Data Publishing

FAIR Data Point
(DTL Netherlands)

## FAIR Publishing of Analysis Tools

smartAPI
(Michel Dumontier)

SADI
(Mark Wilkinson)

## FAIR Metrics

(FAIR Metrics
Working Group)