



WELCOME: DAY 3

Session 1

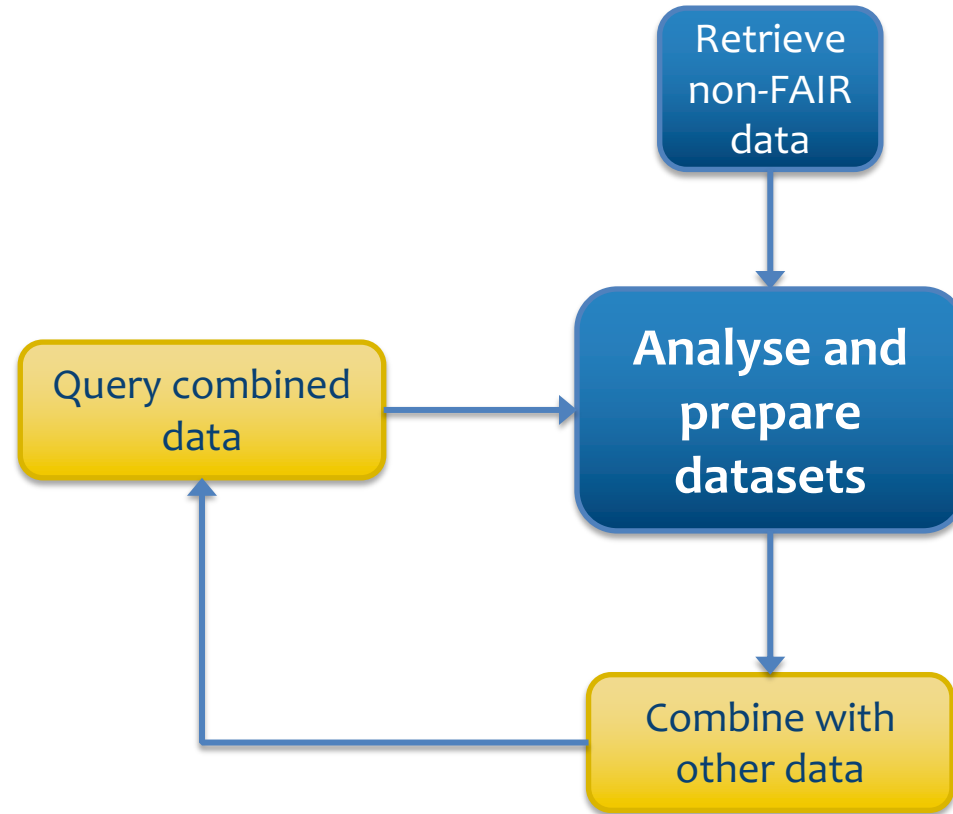
FAIRIFICATION PROCESS AND THE FAIR ECOSYSTEM

FAIR DATA STEWARDSHIP COURSE

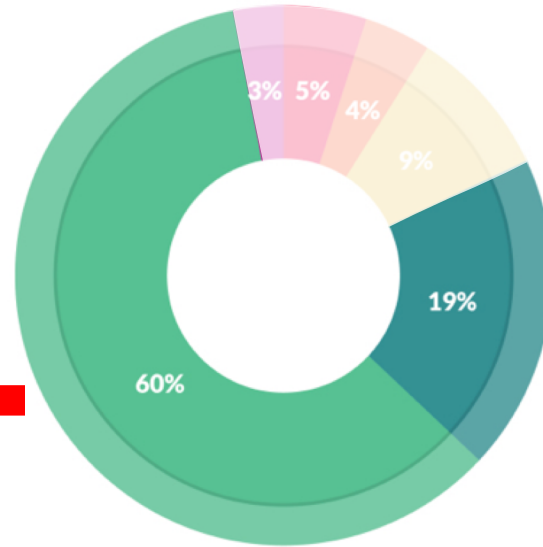
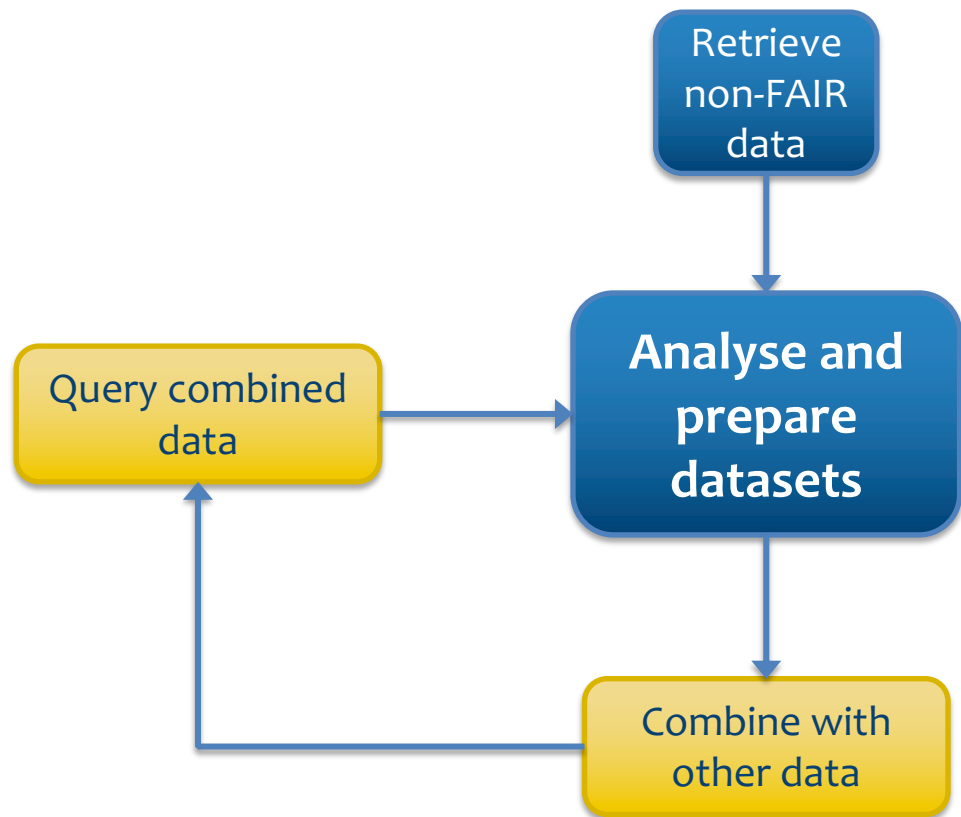
TOWARDS A GO FAIR READINESS PROGRAM

FAIRIFICATION PROCESS

CURRENT SITUATION



CURRENT SITUATION



What data scientists spend the most time doing

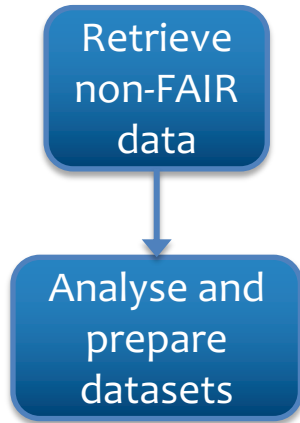
- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

FAIRIFICATION WORKFLOW

Retrieve
non-FAIR
data

Download/locate file
Identify API call
Identify data access protocol

FAIRIFICATION WORKFLOW



Standard format (XML, RDF, relational DB API, VCF, DICOM, etc.)?

What is the content?

Column/field names?

Relations?

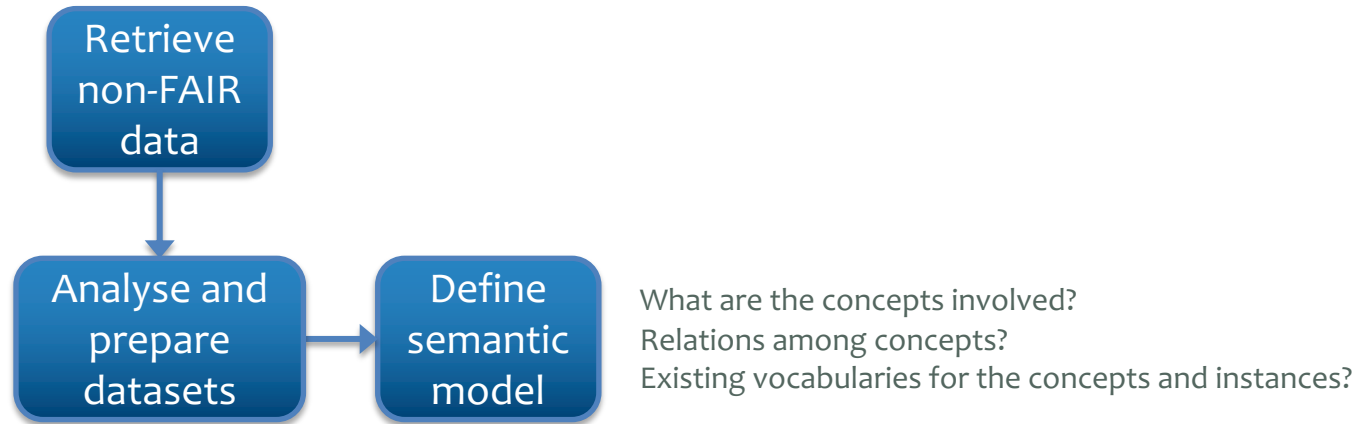
Data domain and range?

Understand the data

Data munging

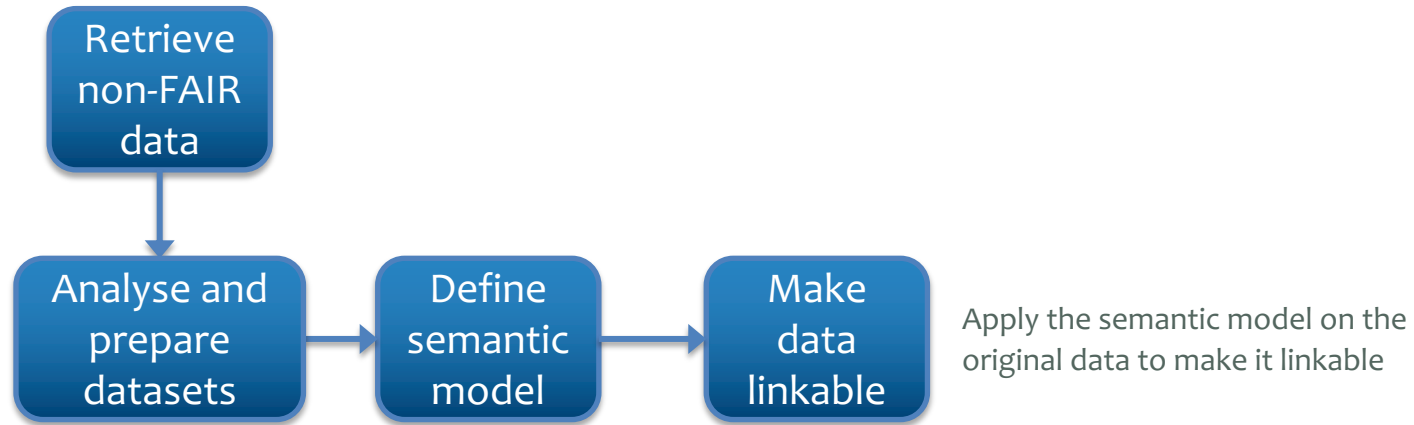
...

FAIRIFICATION WORKFLOW



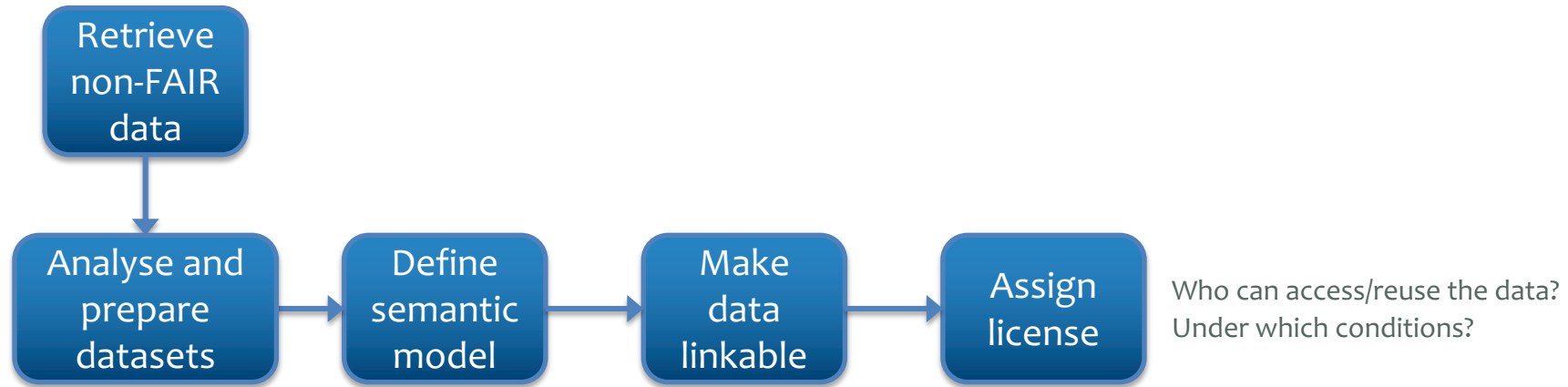
Interoperability
Reusability

FAIRIFICATION WORKFLOW



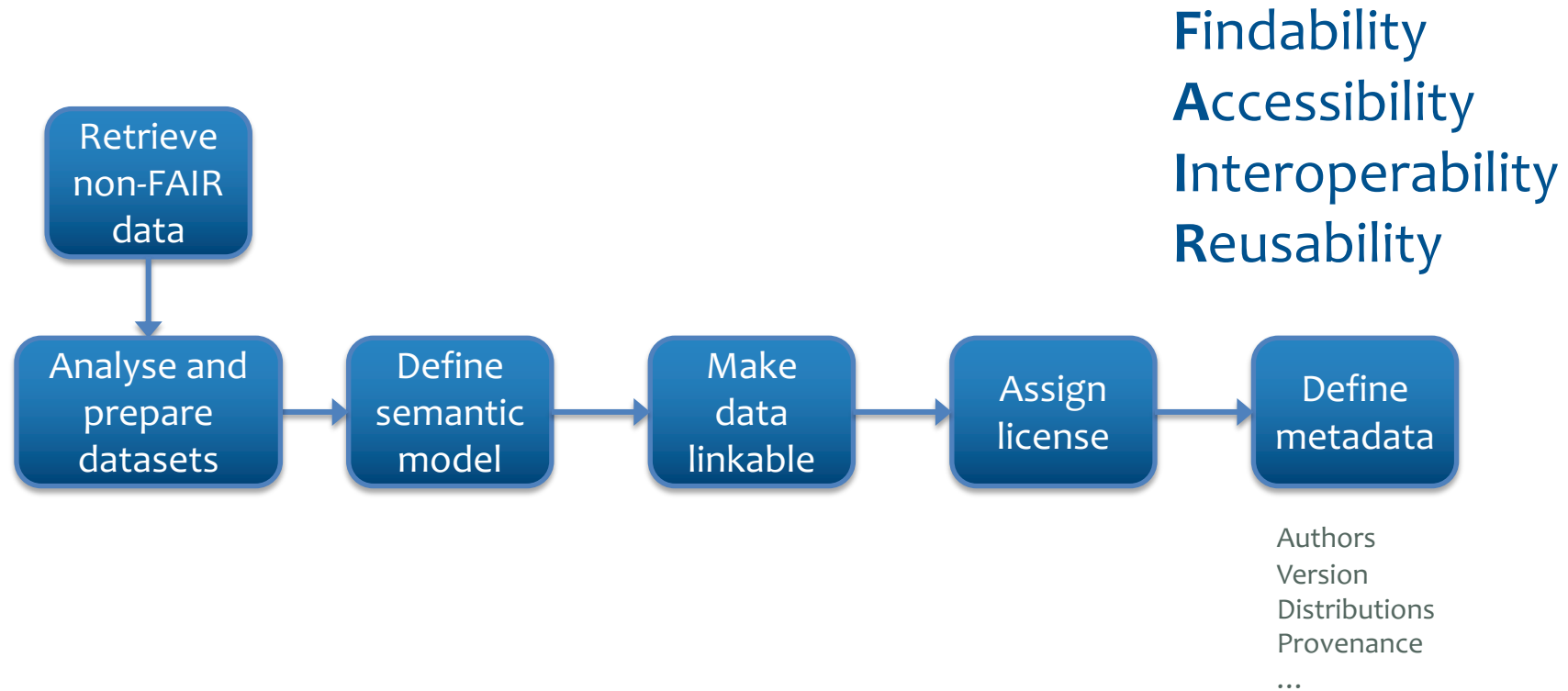
Interoperability
Reusability

FAIRIFICATION WORKFLOW

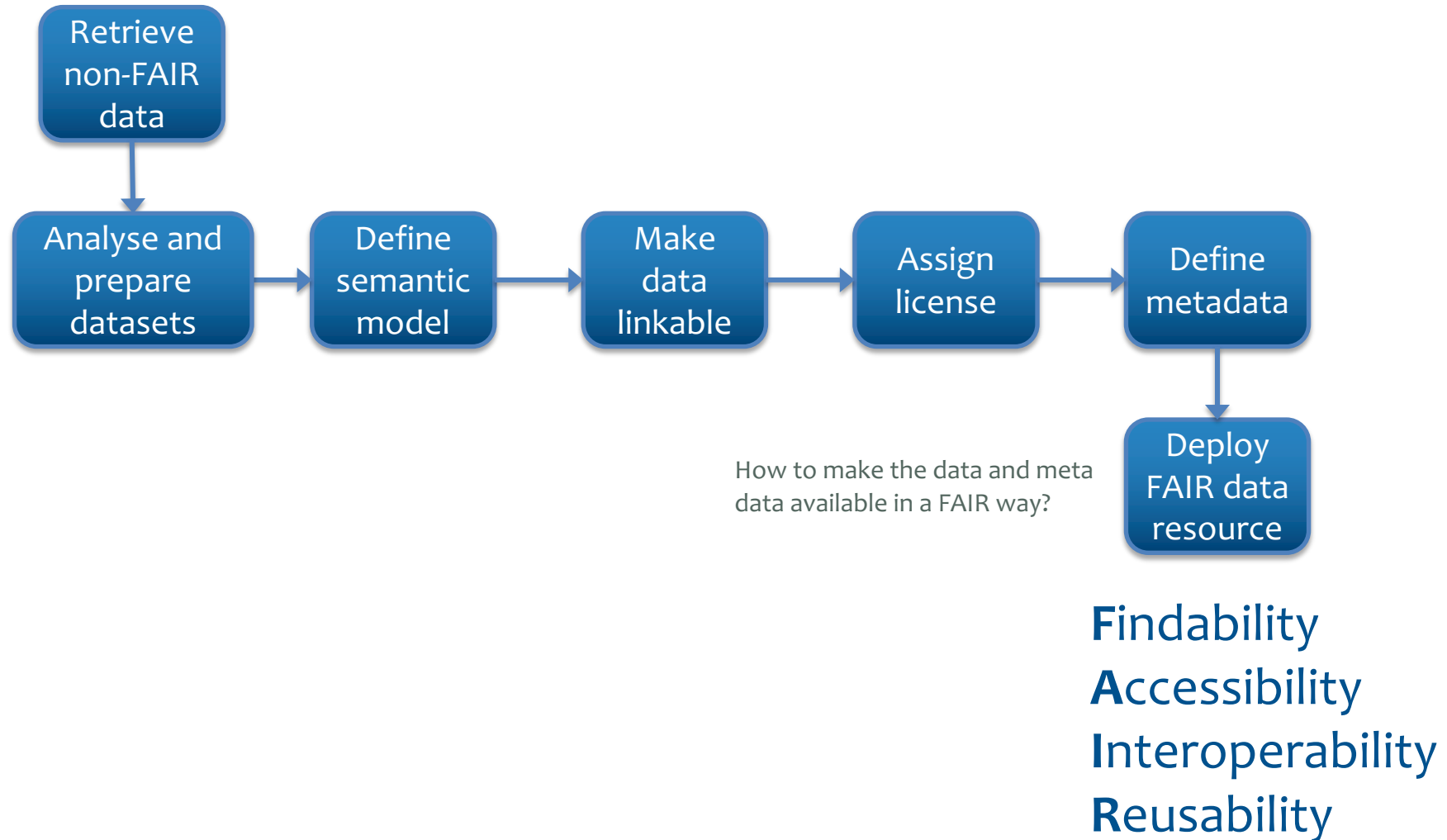


Accessibility
Reusability

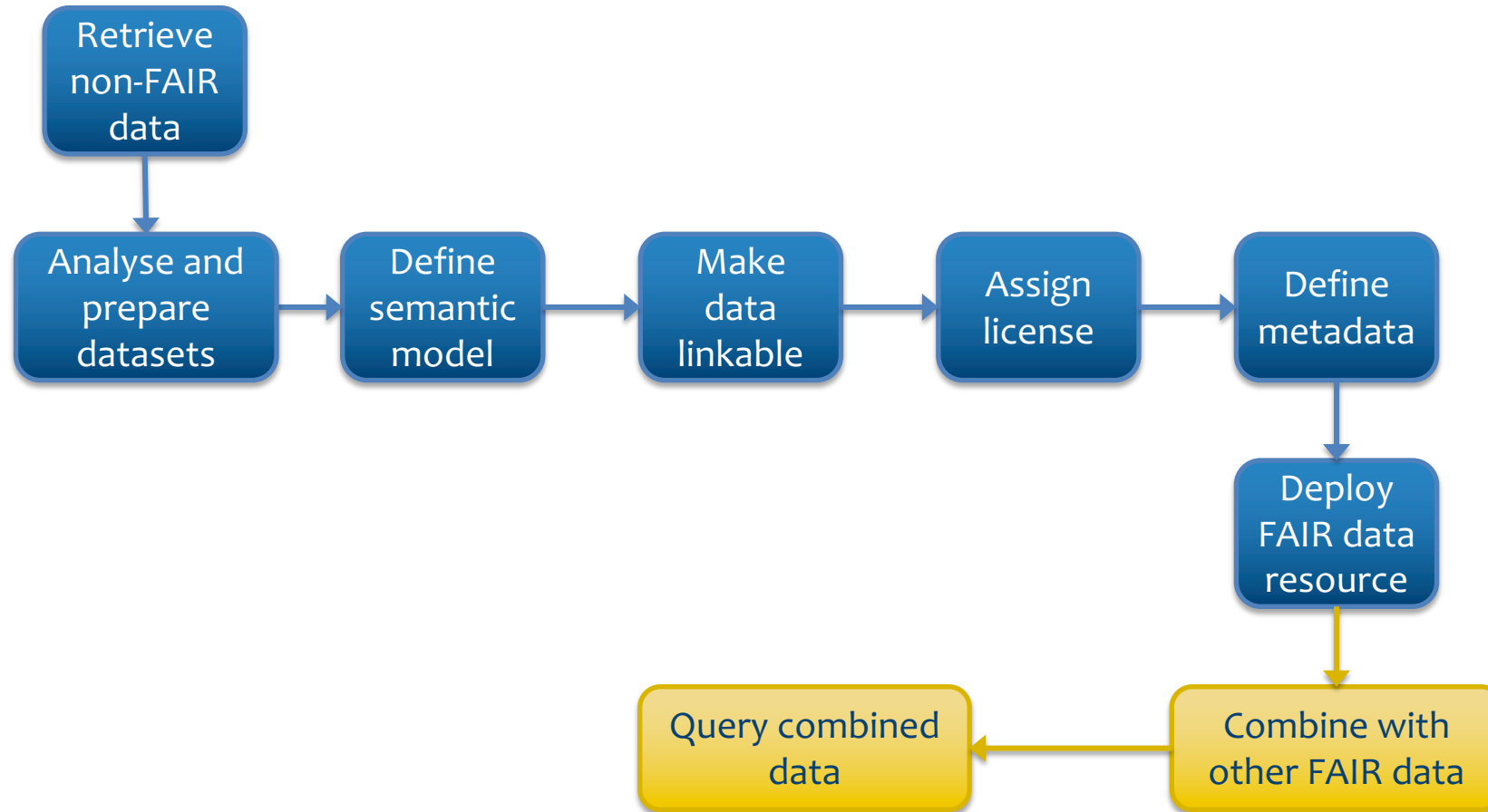
FAIRIFICATION WORKFLOW



FAIRIFICATION WORKFLOW



FAIRIFICATION WORKFLOW



FAIR ECOSYSTEM

FAIR DATA ECOSYSTEM

Plan

Model

Create

Publish

Find

Evaluate



Training

Ontology
Modeling

BYOD

Metadata
Management

Metadata
Registry

FAIR Hackathon

<https://github.com/DTL-FAIRData>

FAIR DATA ECOSYSTEM

Create

FAIRIFIER



Publish

FAIR
DATA POINT



Find

FAIR
DATA 



Annotate

ORKA 
Open RDF
Knowledge Annotator

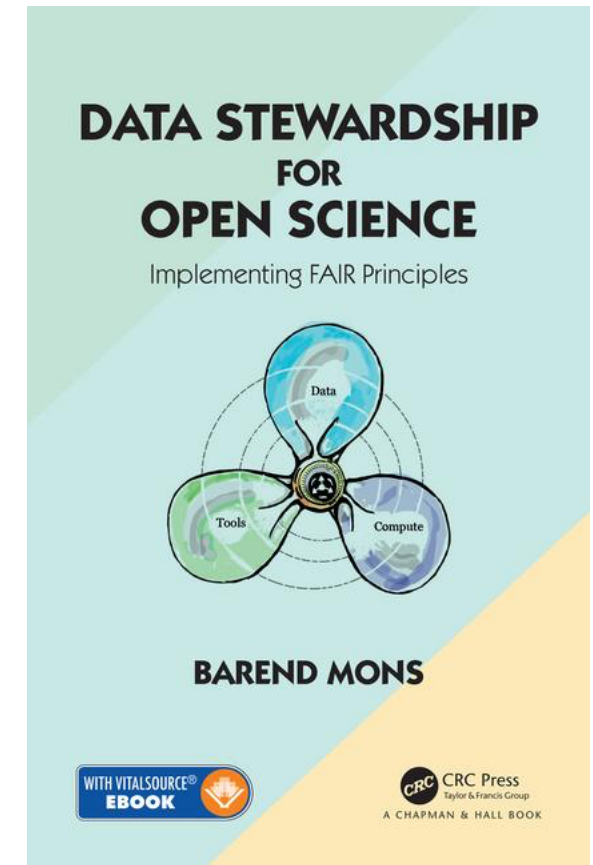


Data FAIRport

DATA STEWARDSHIP WIZARD – DSW.FAIRDATA.SOLUTIONS

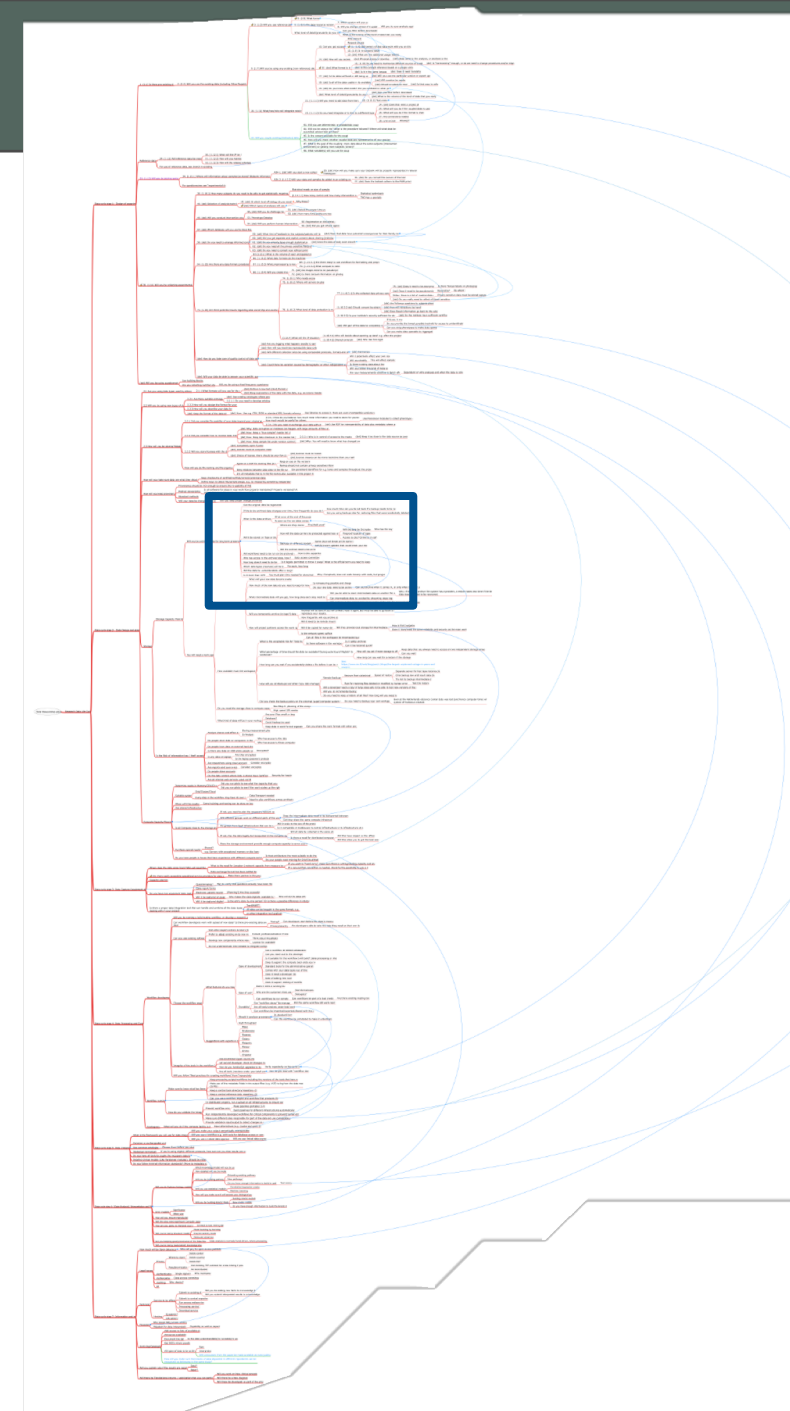
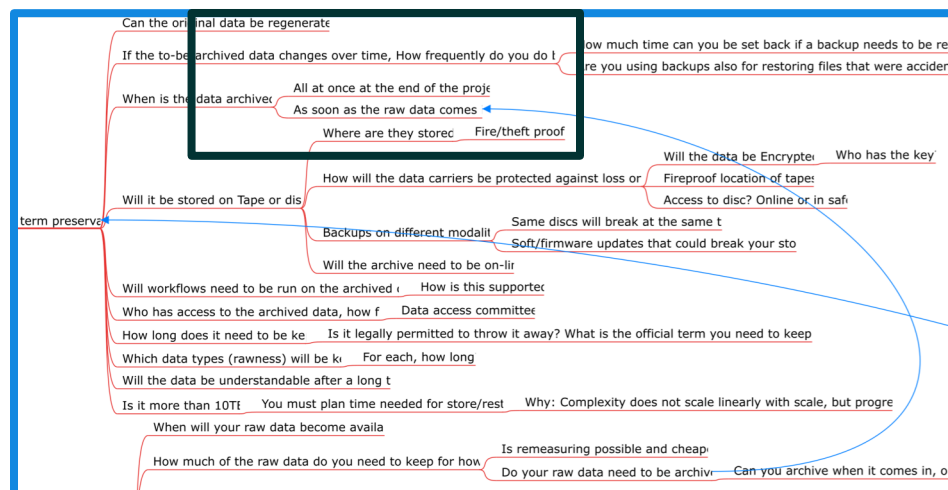
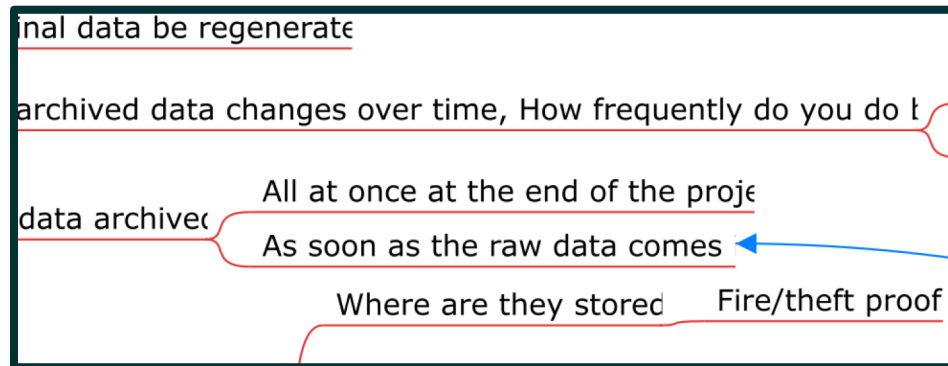
- Dynamic hierarchical questionnaire
- Customisable DS Knowledge Model
- Integration with the Data Stewardship for Open Science book (FAIR) metrics evaluation
- Desirability of answers
- Integration with other resources (FAIRSharing, Bio.tools, ...) – upcoming
- Machine actionable DMPs - part of <http://activedmps.org/>

The screenshot shows the 'Knowledge model editor' interface. On the left is a navigation menu with options: Organization, User Management, KM Editor, KM Packages, and DS Planner. The main area is titled 'DS-KM / Design of experiment / Is there any pre-existing data?'. It contains a 'Question' section with a title 'Is there any pre-existing data?' and a 'Text' field with the question 'Are there any data sets available in the world that are relevant to your planned research?'. Below this is a 'Question Type' dropdown set to 'Options'. The 'Answers' section shows 'No' and 'Yes' options, with a '+ Add answer' button. On the right, a 'Current changes' section lists several questions with their respective options and IDs, such as 'Design of experiment / Is there any pre-existing data?' with options 'No' and 'Yes', and 'Will you be using any pre-existing data (including other people's data)?' with options 'No' and 'Yes'.



DATA STEWARDSHIP KNOWLEDGE MODEL

- Mind map by Rob Hooft (ELIXIR-NL)
 - Over 600 nodes
 - Up to 9 levels
 - Yet just a fragment of all knowledge



DATA STEWARDSHIP WIZARD – DSW.FAIRDATA.SOLUTIONS



FAIRIFIER

Based on OpenRefine

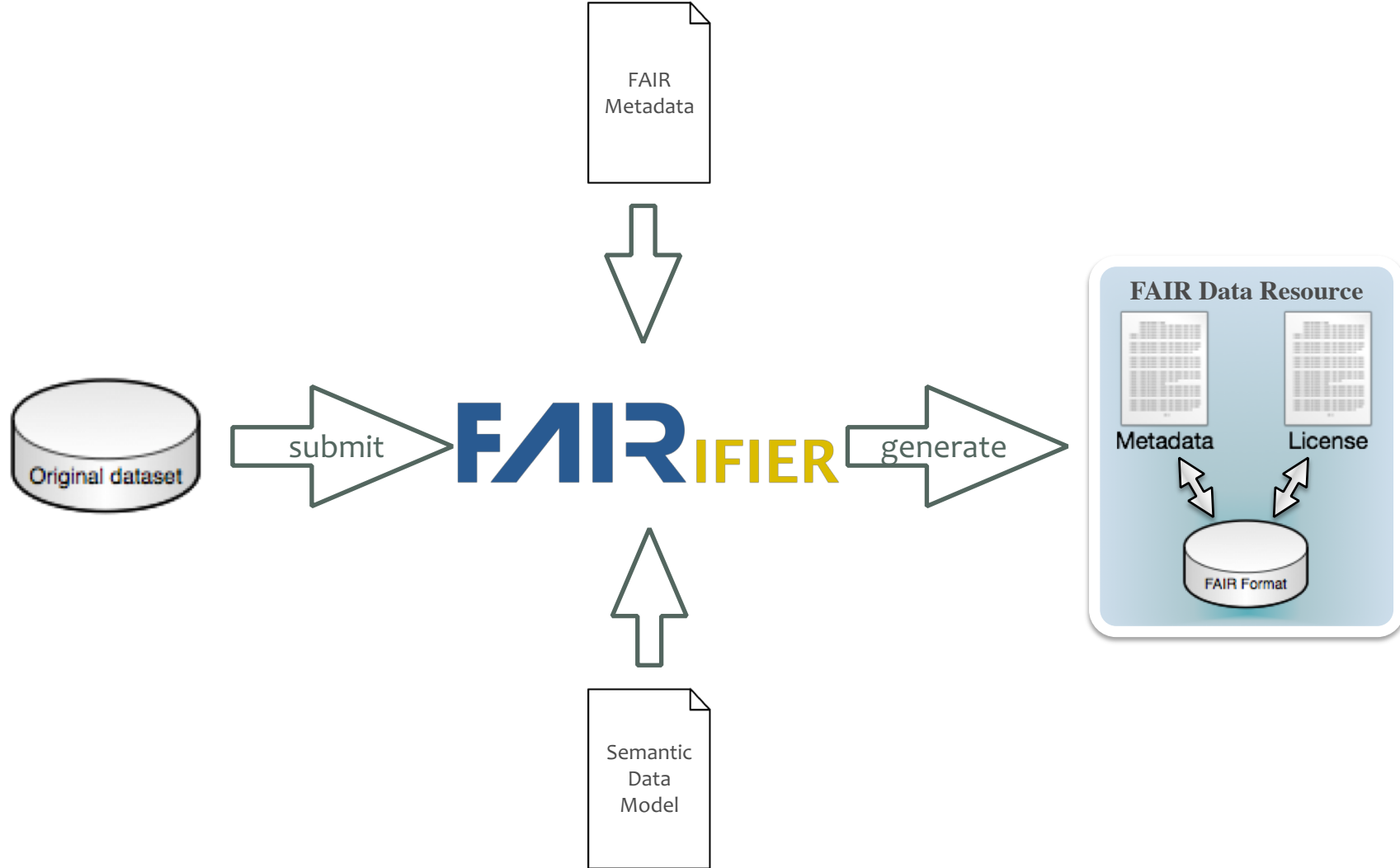
FAIRIFIER – MAJOR FUNCTIONALITY BLOCKS

Data Wrangling

**Semantic Data
Modeling**

**Metadata Definition
and Extraction**

FAIR Publication



Facet / Filter Undo / Redo

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? [Watch these screencasts](#)

23562 rows Extensions: RDF ▾

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 10 next > last »

All	Column 1
1. ☆	101M,R.D.SMITH,J.S.OLSON,G.N.PHILLIPSJR.
2. ☆	102L,D.W.HEINZ,B.W.MATTHEWS
3. ☆	102M,R.D.SMITH,J.S.OLSON,G.N.PHILLIPSJR.
4. ☆	103L,D.W.HEINZ,B.W.MATTHEWS
5. ☆	103M,R.D.SMITH,J.S.OLSON,G.N.PHILLIPSJR.
6. ☆	104L,D.W.HEINZ,B.W.MATTHEWS
7. ☆	104M,R.D.SMITH,J.S.OLSON,G.N.PHILLIPSJR.
8. ☆	105M,R.D.SMITH,J.S.OLSON,G.N.PHILLIPSJR.
9. ☆	106M,R.D.SMITH,J.S.OLSON,G.N.PHILLIPSJR.
10. ☆	107L,M.BLABER,B.W.MATTHEWS

POST to FairDatapoint

The created RDF schema provided can now be uploaded to a FairDatapoint.

Base URI <http://localhost:8084/fdp> edit

catalogs

+ add catalog

▾

OK Cancel

Add new catalog to FAIRDataPoint

identifier edit

title edit

label edit

version edit

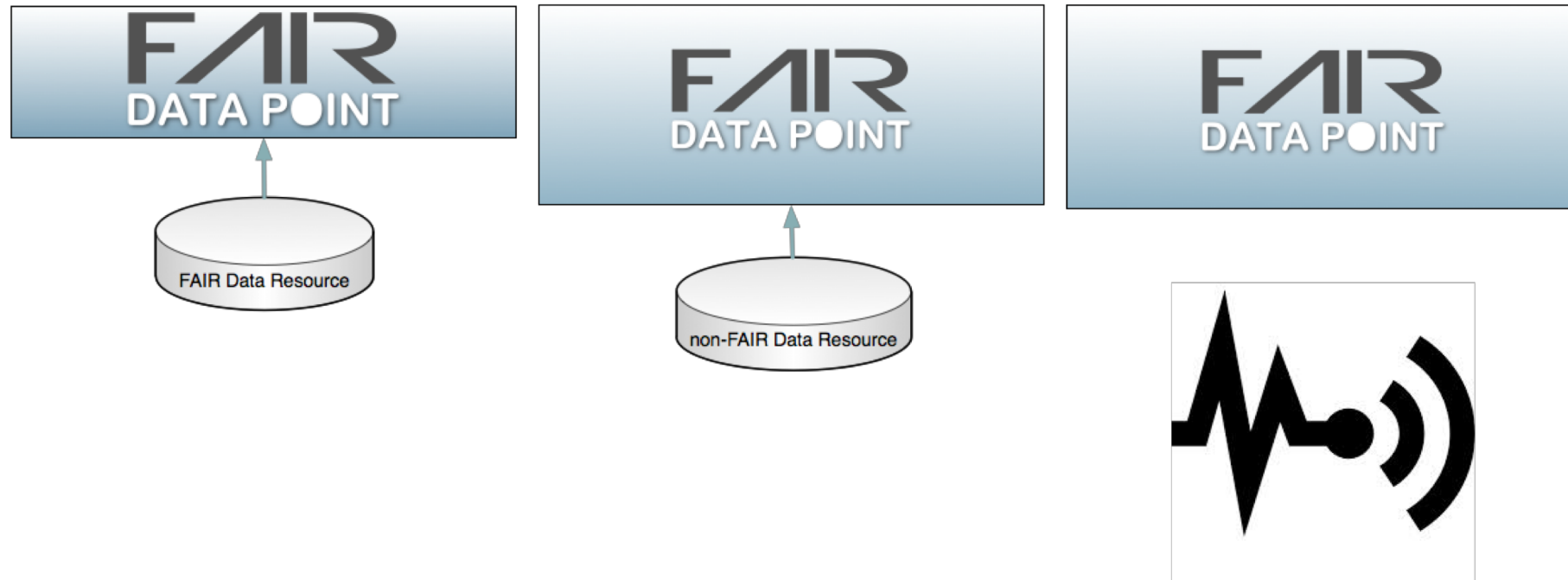
OK Cancel

FAIR

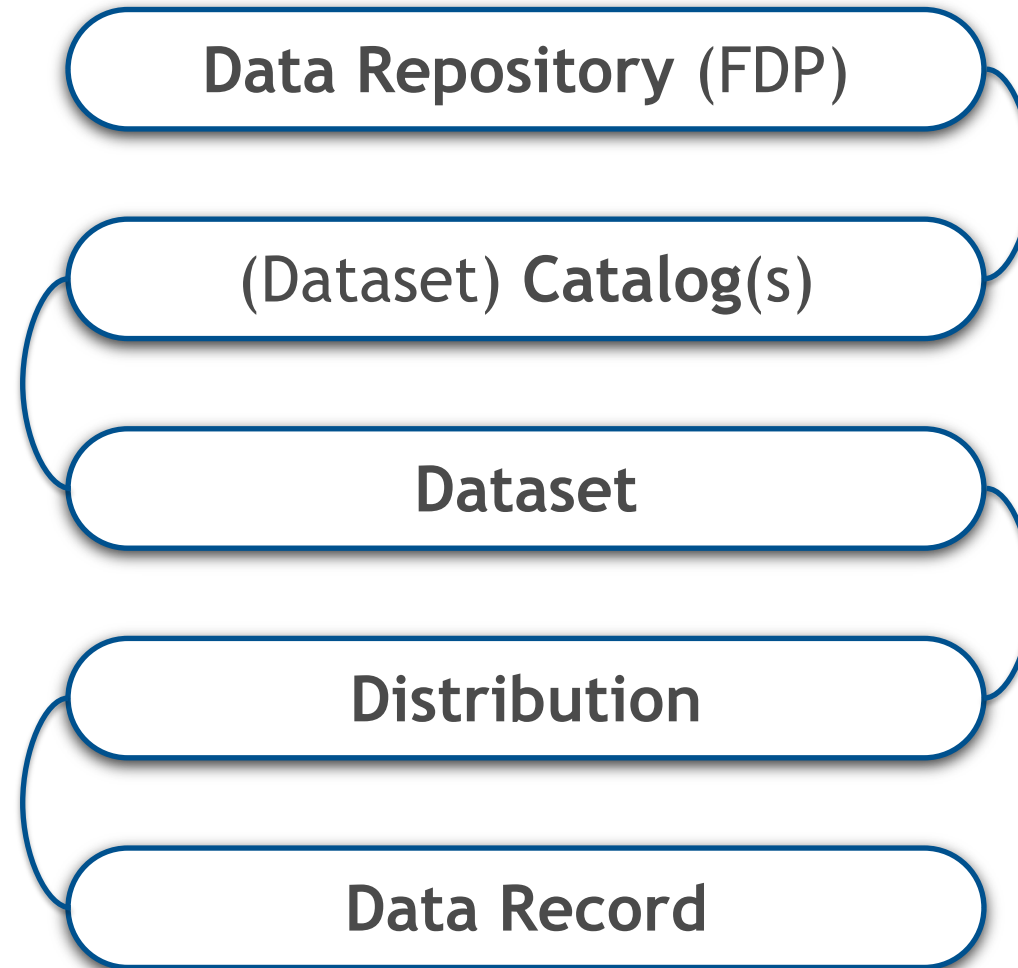
DATA POINT

FAIR DATA POINT

A particular class of FAIR Data System that provides access to datasets in a FAIR manner. The datasets can be external or internal to the FAIR Data Point. Also, the source data can be a non-FAIR dataset or a FAIR Data Resource. If the source data is non-FAIR, the FAIR Data Point needs to made the necessary FAIR transformations on the fly.



METADATA LAYERS



Data Repository (FDP)

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

METADATA LAYERS

Data Repository (FDP)

(Dataset) Catalog(s)

Dataset

Distribution

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

W3C[®]

DCAT/HCLS

METADATA LAYERS

Data Repository (FDP)

(Dataset) Catalog(s)

Dataset

Distribution

Data Record

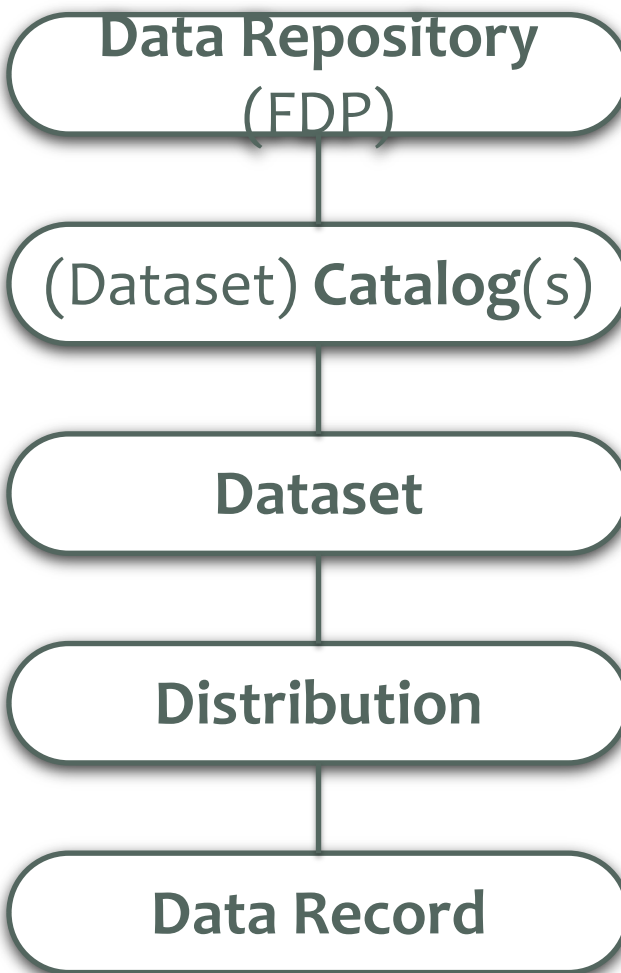
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

W3C[®]

DCAT/HCLS

RML

EXTENDING THE METADATA LAYERS - VOCABULARIES



EXTENDING THE METADATA LAYERS - VOCABULARIES

W3C[®] DCAT

dcat:publisher



Bank for the Diagnosis
and Research on ...

Data Point -> RD connect's patient biobank catalog

Diagnosis and Research on Neuromuscular Disorders (NHMGB)
Diagnosis and Medical Ge

Second University of Naples - Cardiology and Medical Genetics
P.zza Miraglia
80138 Naples
Italy

Dataset containing autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit a hendrerit eu feugiat nulla facilisis.
Lorem ipsum dolor sit amet, cons ectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt.

Published on 05-02-2017
Last modified on 05-02-2017

CC BY NC ND
Catalog(s)

X R D T

 Bioschemas
biosch:organization

Keywords: Bank for the diagnosis and research on neuromuscular dise

DBpedia: biobank:
EDAMt: biobank:
Orphanet: autosomal dominant spastic paraplegia type, autosomal recessive limb-girdle muscular dy Brugada syndrome, familial isolated dilated cardiomyopathy, carnitine palmitoyltransferase II deficiency, autosomal recessive limb-girdle muscular dy Fechtner syndrome, Freeman-Sheldon syndrome, autosomal recessive limb-girdle muscular dy, autosomal recessive limb-girdle muscular dy muscular lipodosis, autosomal recessive limb-girdle muscular dy tubular aggregate myopathy, Emery-Dreifuss muscular dystrophy, limb-girdle muscular dystrophy, autosomal recessive limb-girdle muscular dy, autosomal recessive limb-girdle muscular dy, autosomal recessive limb-girdle muscular dy, autosomal recessive limb-girdle muscular dy, facioscapulohumeral dystrophy, Steinert myotonic dystrophy, Ehlers-Danlos syndrome, hypermobility type, Kaposi sarcoma, autosomal recessive limb-girdle muscular dy, autosomal recessive limb-girdle muscular dy, glycogen storage disease due to acid maltase

```
"@context": "http://schema.org",  
"@type": "NGO",  
"address": {  
  "@type": "PostalAddress",  
  "addressLocality": "Utrecht, The Netherlands",  
  "postalCode": "3511 GC",  
  "streetAddress": "Catharijnesingel 54"  
},  
"email": "info(at)dtls.nl",  
"@type": "Organization",  
"@type": "not-for-profit",  
"name": "Dutch Techncentre for Life Sciences",  
"telephone": "( 31) 85 30 30 711"
```

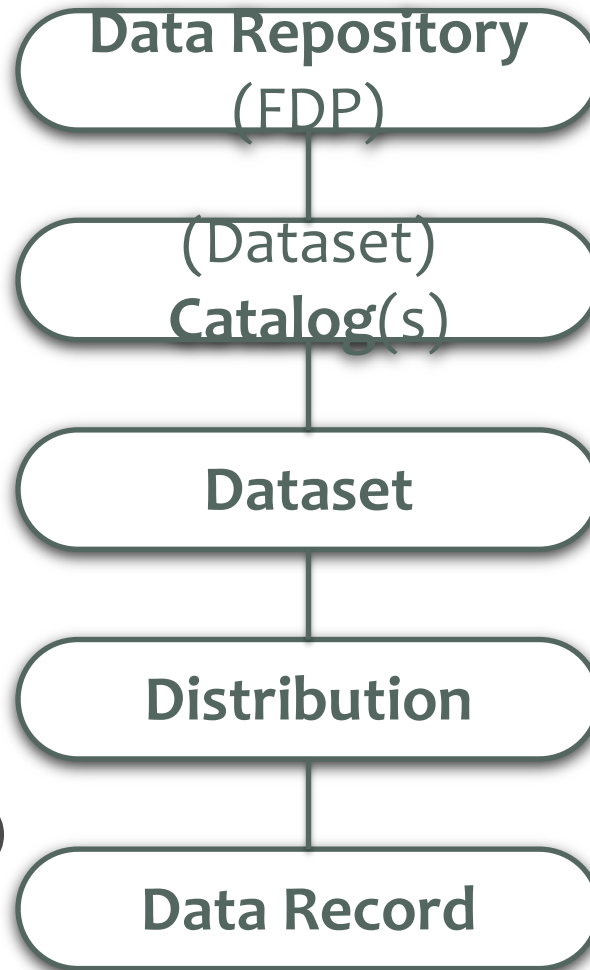
FAIR
DATA POINT

EXTENDING THE METADATA LAYERS – EXTENDED MODEL

**Data Tag Suite
(DATS)**

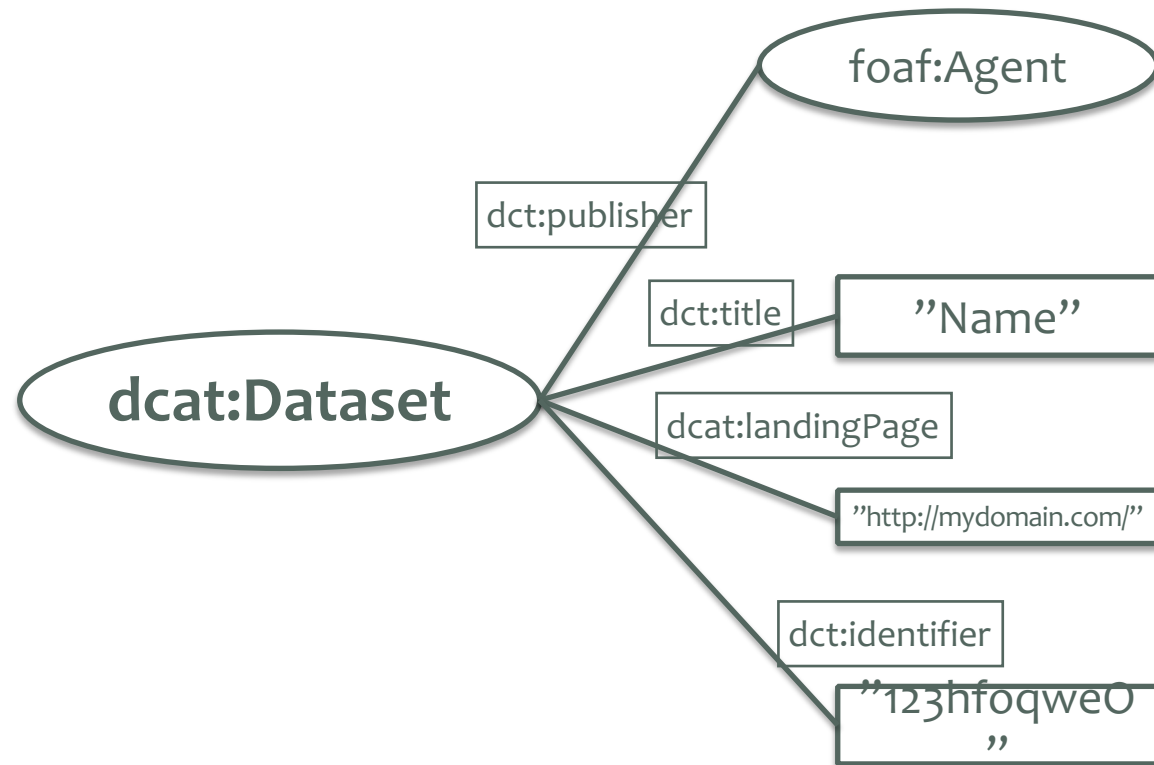
W3C[®] PROV

W3C[®] PROV-O

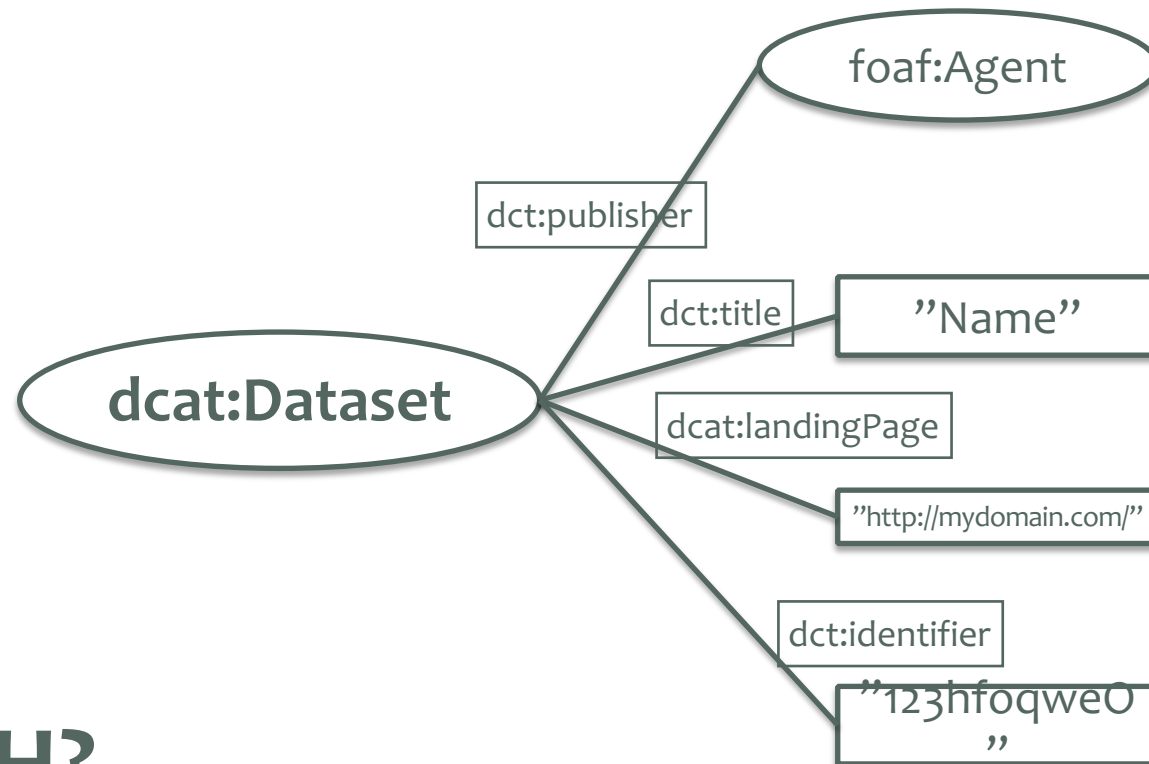


SCHOLIX

MODEL EXTENSION EXAMPLE

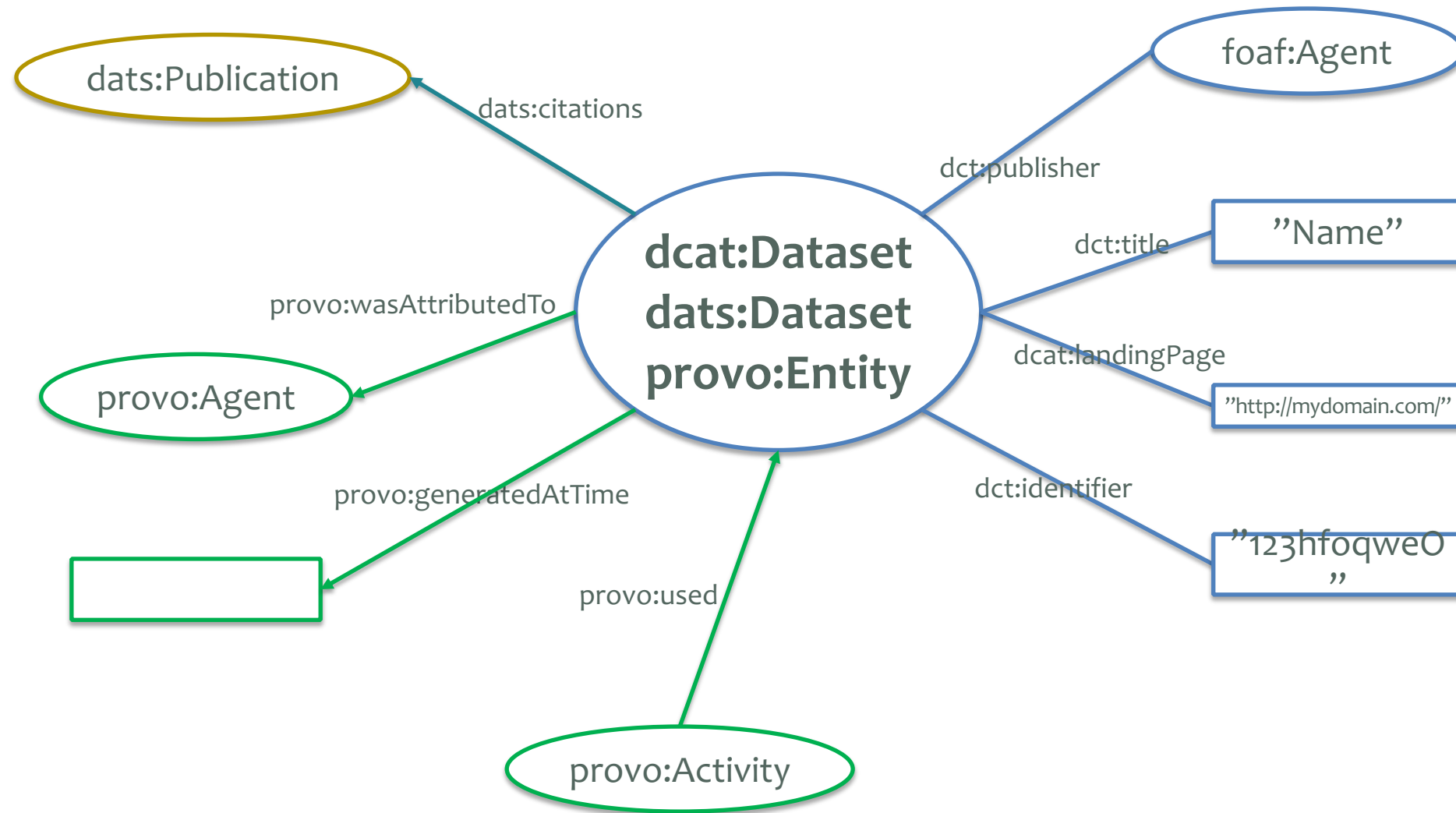


MODEL EXTENSION EXAMPLE



ENOUGH?

MODEL EXTENSION EXAMPLE



DTL FAIR Data Point -> RD connect's patient biobank catalog

Bank for the Diagnosis and Research on Neuromuscular Disorders (NHMGB) Cardiology and Medical Ge

Second University of Naples - Cardiology and Medical Genetics
I Policlinico - Piazza Miraglia
80138 Naples
Italy

Dataset containing autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugait nulla facilisi.

Lorem ipsum dolor sit amet, cons ectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt.

Published on 05-02-2017


Last modified on 05-02-2017



Version: 3.4

Catalog(s)

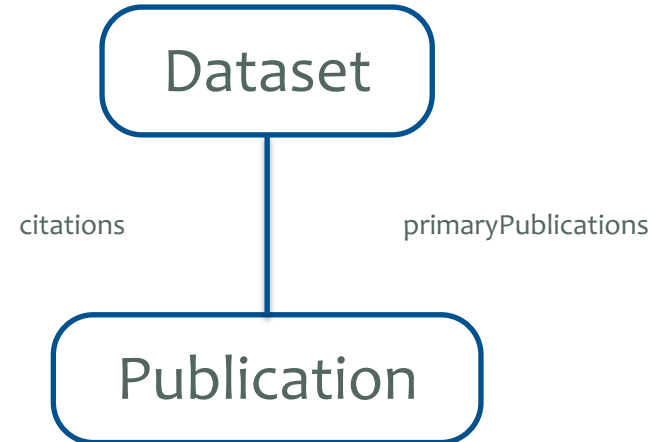


Associated Publications: Iriure dolor in hendrerit in vulputate velit esse molestie consequat.
Iriure dolor in hendrerit in vulputate velit esse molestie consequat.
Iriure dolor in hendrerit in vulputate velit esse molestie consequat. 

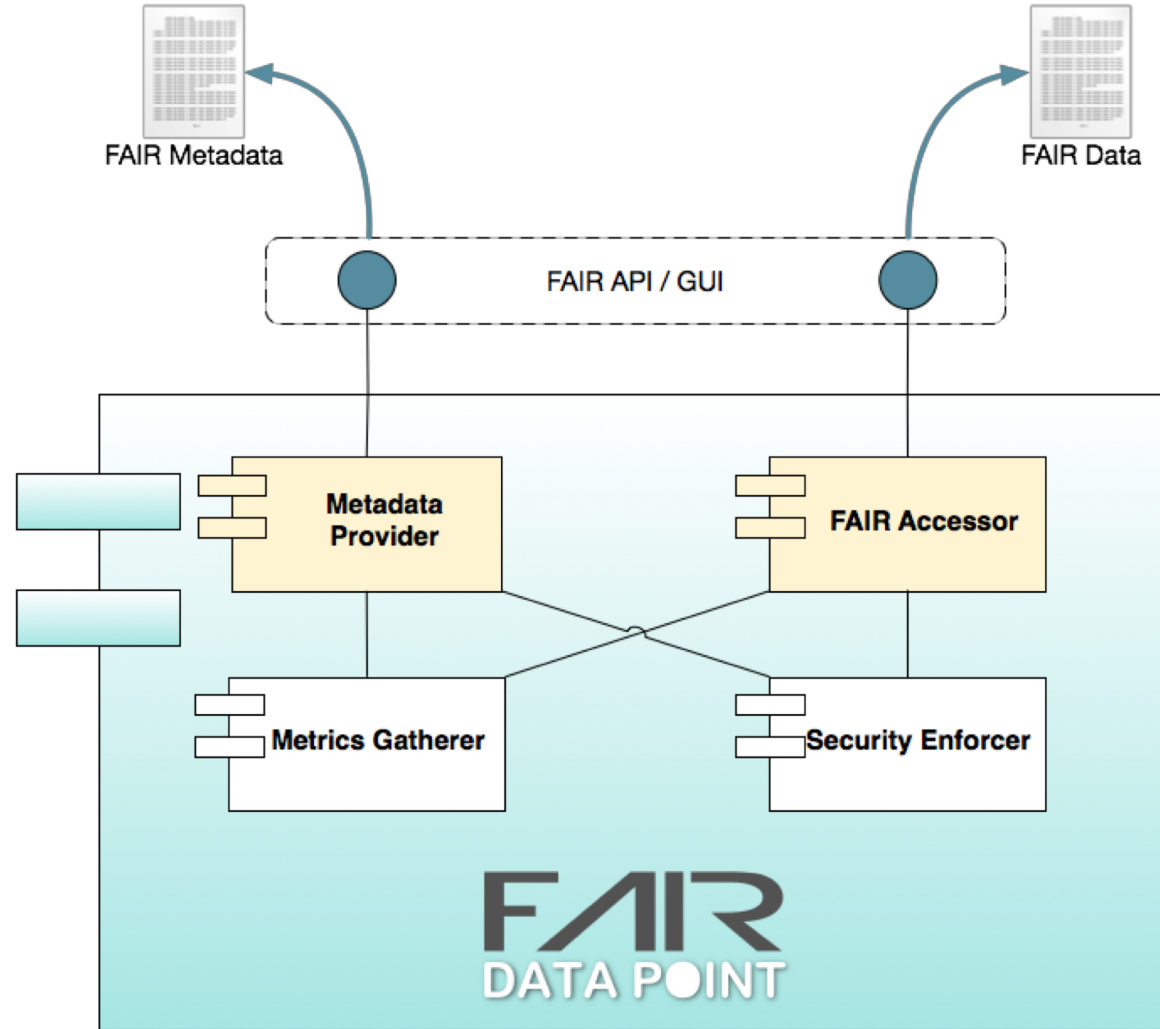
Keywords: Bank for the diagnosis and research on neuromuscular diseases (NHMGB). Cardiology and Ge. biobank

DBpedia: biobank;
EDAM: biobank;
Orphanet: autosomal dominant spastic paraplegia type 4, autosomal recessive limb-girdle muscular dystrophy type 2E, Brugada syndrome, familial isolated dilated cardiomyopathy, carnitine palmitoyltransferase II deficiency, autosomal recessive limb-girdle muscular dystrophy type 2H, Fechtner syndrome, Freeman-Sheldon syndrome, autosomal recessive limb-girdle muscular dystrophy type 2L, autosomal recessive limb-girdle muscular dystrophy type 2N, muscular lipidosis, autosomal recessive limb-girdle muscular dystrophy type 2F, tubular aggregate myopathy, Emery-Dreifuss muscular dystrophy, limb-girdle muscular dystrophy, autosomal recessive limb-girdle muscular dystrophy type 1B, autosomal recessive limb-girdle muscular dystrophy type 1C, autosomal recessive limb-girdle muscular dystrophy type 2A, autosomal recessive limb-girdle muscular dystrophy type 2B, facioscapulohumeral dystrophy, Steinert myotonic dystrophy.


Data Tag Suite (DATS)



FAIR DATA POINT - ARCHITECTURE



FAIR DATA POINT - GUI - FOR TECHIES

 **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 specs](#)
- [Source code](#)

Created by r.kaliyaperumal@lumc.nl
[CC BY-NC-ND 3.0](#)

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: / , API VERSION: 0.1 BETA]

FAIR DATA POINT - GUI - FOR "NORMAL" PEOPLE



The DTL FAIR Data Point stores the datasets made FAIR through BYODs. These datasets lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exercitation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt

Issued: 27-01-2017

license

Last modified: 18-02-2017

Responsible organisation: Dutch Techcenter for Life Sciences

Catalog(s)

RD Connect's Patient Registry Catalog

This catalog contains a number of patient registries' datasets autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui.

Datasets: 284

BBMRI Linked Data task biobank catalog

This catalog contains datasets made FAIR from the BBMRI 2 NL Linked Data task.

Datasets: 14

Repository
metadata

Catalog
metadata
summary

FAIR
DATA POINT

GO FAIR

FAIR DATA POINT - GUI



The DTL FAIR Data Point stores the datasets made FAIR through BYODs. These datasets lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisi ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt

Issued: 27-01-2017 license Last modified: 18-02-2017
Responsible organisation: Dutch Techcenter for Life Sciences

Catalog(s)

RD Connect's Patient Registry Catalog

This catalog contains a number of patient registries' datasets autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui.

Datasets: 284

BBMRI Linked Data task biobank catalog

This catalog contains datasets made FAIR from the BBMRI 2 NL Linked Data task.

Datasets: 14

RD Connect's Patient Registry Catalog

This catalog contains a number of patient registries' datasets autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt

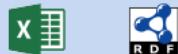
Issued: 29-01-2017 license Last modified: 09-03-2017
Keywords: DBPedia:biobank, EDAM:biobank

Datasets

Bank for the diagnosis and research on neuromuscular disorders (NHMG), Cardiology and Medical GE

Dataset containing autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui.

Published: 05-02-2017 license Version: 3.4



Biobank of the Institute of Rare Diseases/Institute of Health Carlos III (IER-ISCI3)

Dataset containing autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui.

Published: 08-02-2017 license Version: 1.7



Repository metadata

Catalog metadata summary

Catalog metadata

Dataset/distribution metadata summary

FAIR DATA POINT - GUI - DATASET



Bank for the Diagnosis and Research on ...

DTL FAIR Data Point -> RD connect's patient biobank catalog

Bank for the Diagnosis and Research on Neuromuscular Disorders (NHMGB) Cardiomyology and Medical Ge

Second University of Naples - Cardiomyology and Medical Genetics

I Policlinico - Piazza Miraglia
80138 Naples
Italy

Dataset containing autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugait nulla facilisi.

Lorem ipsum dolor sit amet, cons ectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt.

Published on 05-02-2017

Last modified on 05-02-2017



Version: 3.4

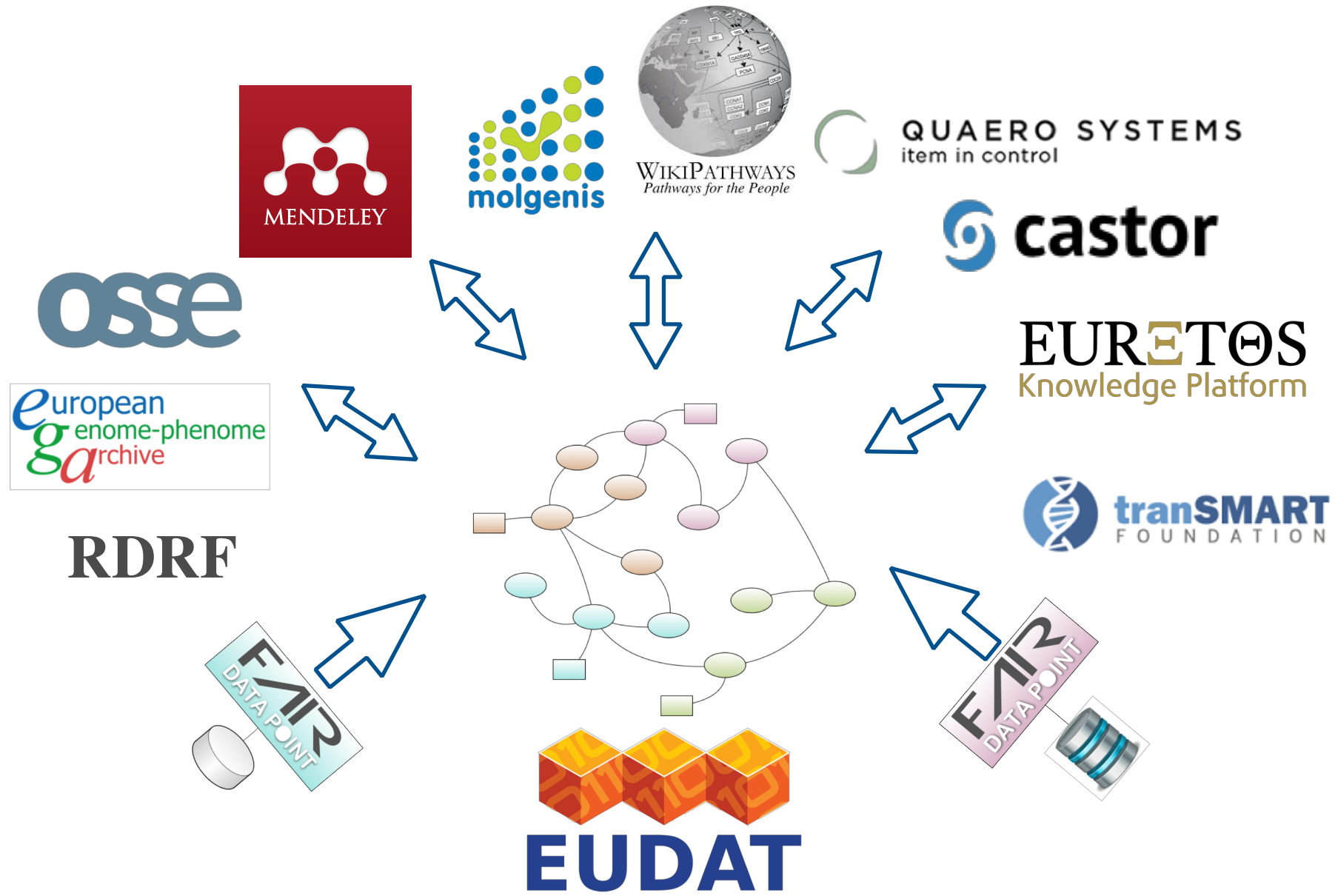
Catalog(s)



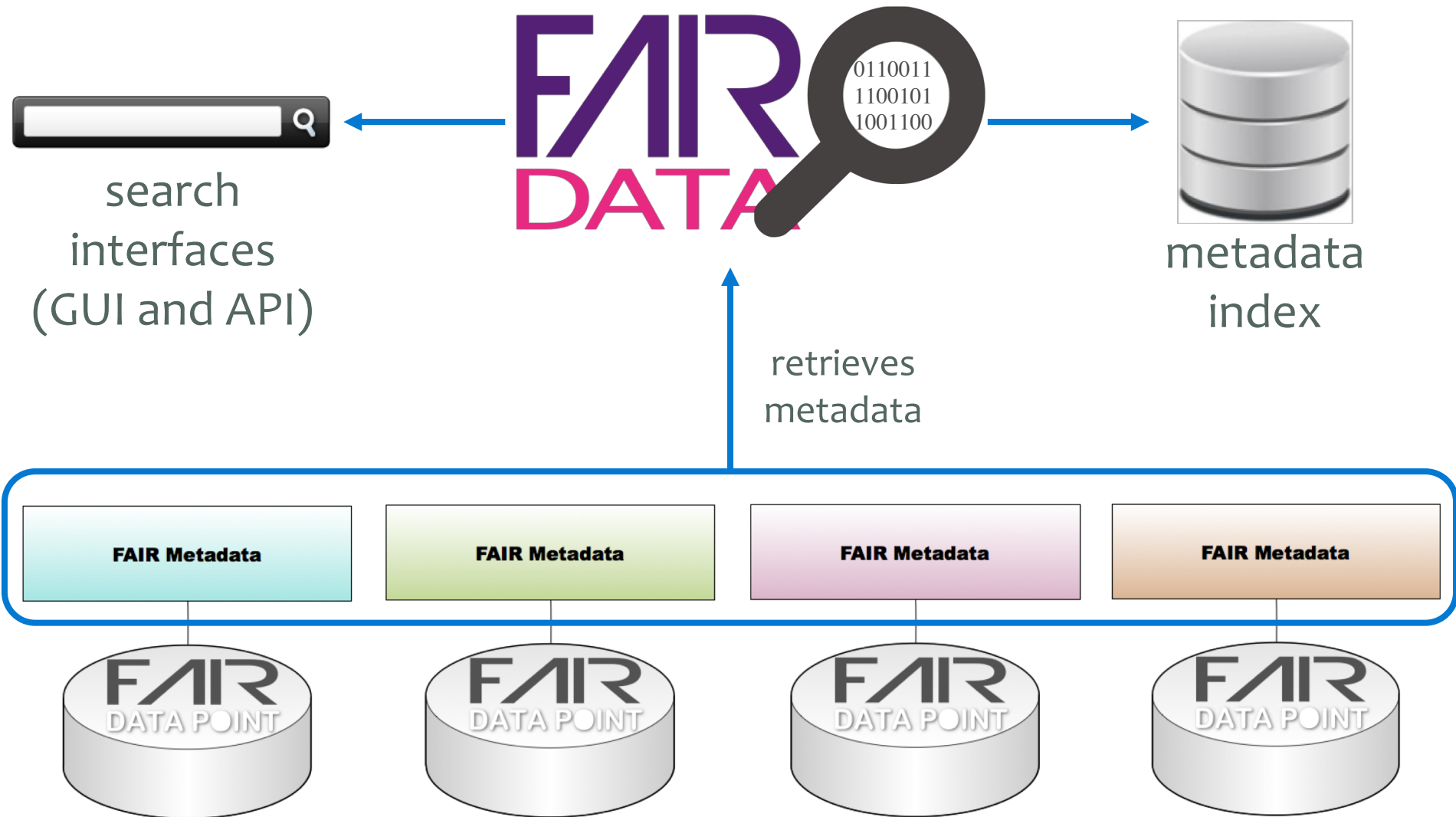
Keywords: [Bank for the diagnosis and research on neuromuscular diseases \(NHMGB\)](#), [Cardiomyology and Ge](#), [biobank](#)

DBpedia: [biobank](#)
EDAM: [biobank](#)
Orphanet: [autosomal dominant spastic paraplegia type 4](#),
[autosomal recessive limb-girdle muscular dystrophy type 2E](#),
[Brugada syndrome](#),
[familial isolated dilated cardiomyopathy](#),
[carnitine palmitoyltransferase II deficiency](#),
[autosomal recessive limb-girdle muscular dystrophy type 2H](#),
[Fechtner syndrome](#),
[Freeman-Sheldon syndrome](#),
[autosomal recessive limb-girdle muscular dystrophy type 2L](#),
[autosomal recessive limb-girdle muscular dystrophy type 2N](#),
[muscular lipodosis](#),
[autosomal recessive limb-girdle muscular dystrophy type 2F](#),
[tubular aggregate myopathy](#),
[Emery-Dreifuss muscular dystrophy](#),
[limb-girdle muscular dystrophy](#),
[autosomal recessive limb-girdle muscular dystrophy type 1B](#),
[autosomal recessive limb-girdle muscular dystrophy type 1C](#),
[autosomal recessive limb-girdle muscular dystrophy type 2A](#),
[autosomal recessive limb-girdle muscular dystrophy type 2B](#),
[facioscapulohumeral dystrophy](#),
[Steinert myotonic dystrophy](#),
[Ehlers-Danlos syndrome, hypermobility type](#),
[Kaposi sarcoma](#),
[autosomal recessive limb-girdle muscular dystrophy type 2I](#),
[autosomal recessive limb-girdle muscular dystrophy type 2C](#),
[glycogen storage disease due to acid maltase deficiency](#)


FDP-COMPLIANT (BETA) SOLUTIONS







search-dev.fair-dtfs.surf-hosted.nl



html (6) apache (10)
gz (4)
unknown (4)

Galliera Genetic Bank
Galliera Genetic Bank dataset

Ring14 biobank turtle distribution (73.9 KiB / APACHE)
Ring14 biobank linked data fragment distribution (/ APACHE)

DTL FAIR Data Point / Rd connect's biobank catalog / Distribution ▾

Ring14 Clinical database
Ring14 patient registry clinical database

Ring14 patient registry turtle distribution (71.4 KiB / APACHE)
Ring14 patient registry linked data fragment distribution (/ APACHE)


DTL FAIR Data Point / Rd connect's patient registry catalog / Distribution ▾

Fantom5 human transcription start site
Fantom5 transcriptomics dataset, which contains info about human transcription start site

Fantom5 nanopubs SPARQL endpoint (/ APACHE) Fantom5 TET web app (/ APACHE)
Fantom5 (human tss) tab separated file, plain text file distribution(gzip) (792.5 MiB / APACHE)

DTL FAIR Data Point / Catalog for transcriptomics datasets / Distribution ▾

GoNL human variants
The dataset contain 27.8k SV calls (>20bp). Calling was realized using 10 differ...
(see below) and a consensus strategy was used to produce this set. The SOURCE field in the



powered by thumbaltr.com

Catalog for transcriptomics datasets

<http://dbpedia.org/resource/Transcriptomics>
http://edamontology.org/topic_3308

Fantom5 nanopubs SPARQL endpoint
Fantom5 TET web app
Fantom5 (human tss) tab separated file, plain text file distribution(gzip)



- Allow third-party annotation on existing knowledge bases
- Capture the provenance of the annotator and the original statement

Open RDF Knowledge Annotator

A service for annotating RDF triples.

Learn more

ORKA

ORKA is an annotation service that lets you annotate any RDF triple. For more information about ORKA, click [here](#).

Demo

To see ORKA in action, click the demo button.

Demo

Nanopublications

The annotations created through ORKA are stored as a nanopublication. You can read more about nanopublications on the official website at nanopub.org.

ODEX4ALL

This application is created in the context of ODEX4ALL. Read all about this project on the [DTL website](#).

Demo

To see ORKA in action, click the demo button.

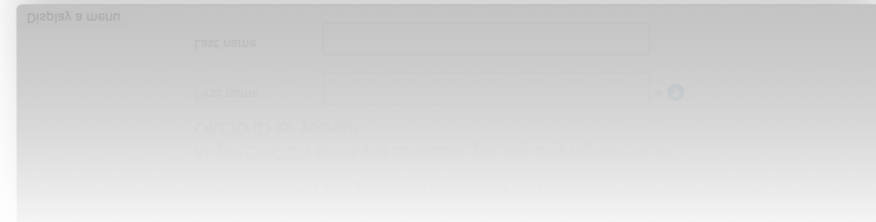
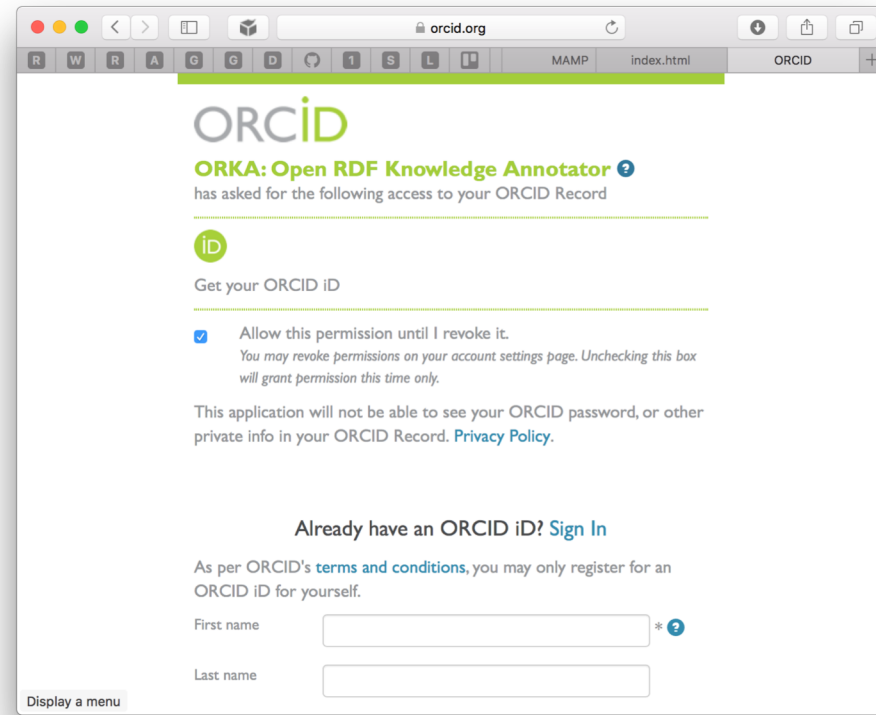
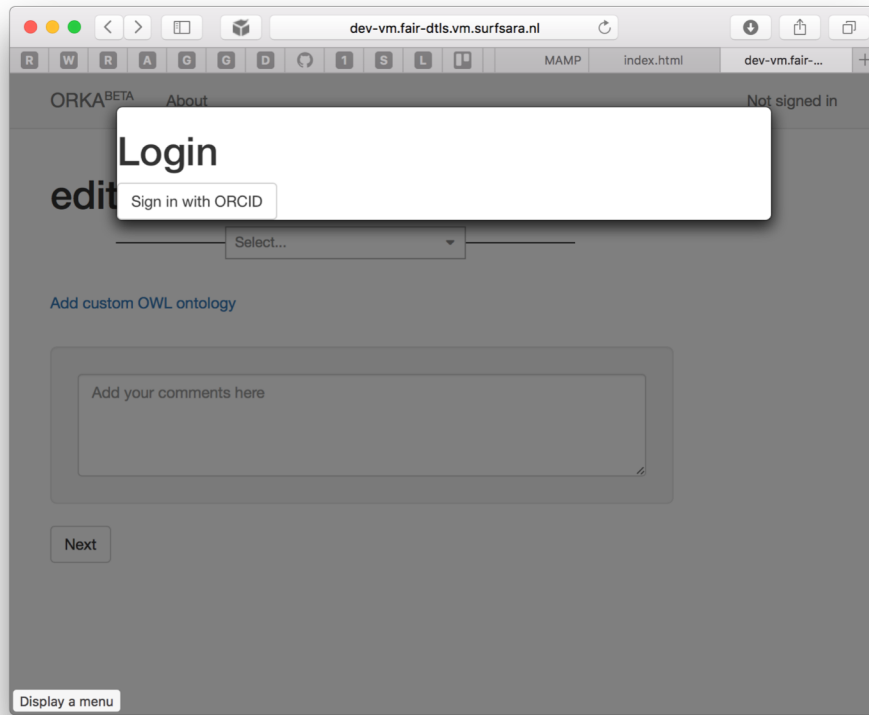
Demo

RDF triple. For more information about ORKA, click [here](#).

nanopublications on the official website at nanopub.org.

nanopublication. You can read more about

Read all about this project on the [DTL website](#).



dev-vm.fair-dtls.vm.surfsara.nl

ORKA^{BETA} About Luiz Olavo Bonino da Silva Santos

edit

aspirin — **treats** — **smoking**

[Add custom OWL ontology](#)

- treats
- shows
- highlights
- treats
- increases
- summons

Add your comments here

Next

ANNOTATIONS GO TO NANOPUB STORE

Display a menu

FAIR METRICS EVALUATOR

- Based on the metrics defined by the FAIR Metrics Group (fairmetrics.org)
- Two forms of evaluation:
 - Automatic – given the URL of the resource, the Evaluator performs an automatic evaluation
 - Manual – the user answers a questionnaire
- Support for core and extended metrics

FAIR
METRICS
EVALUATOR

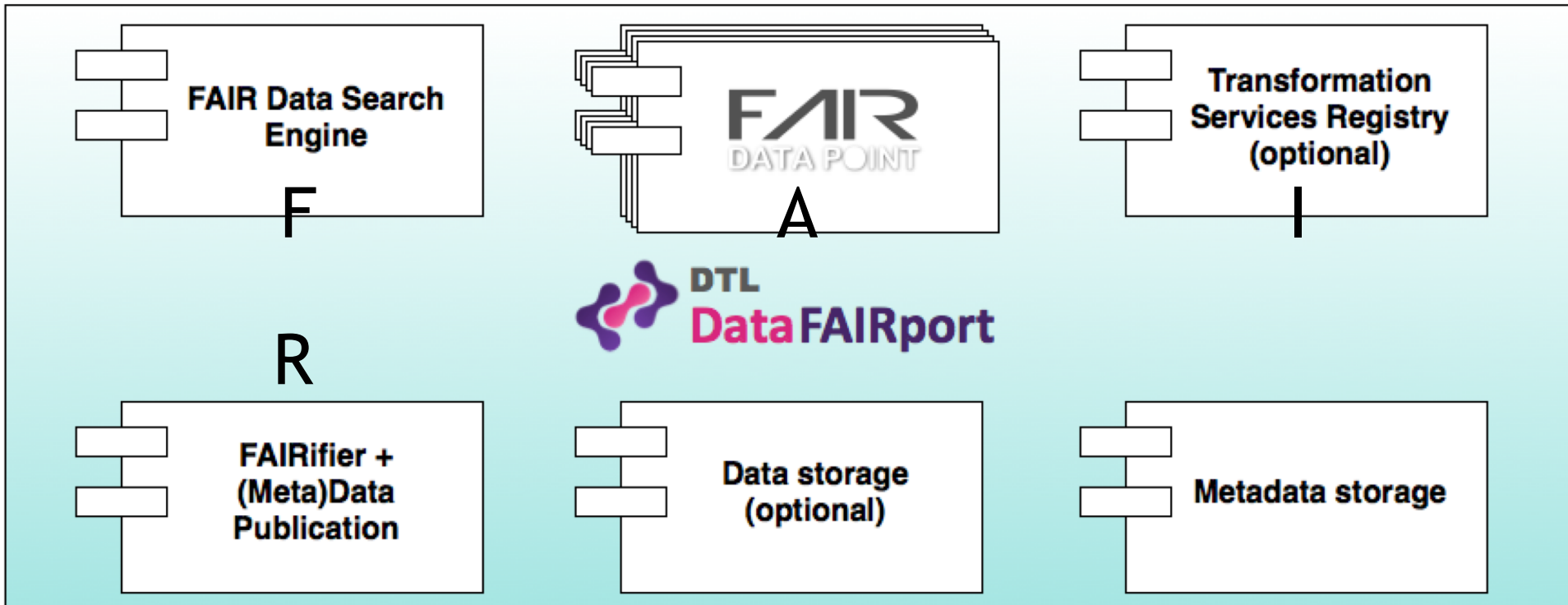
GO FAIR FOUNDATION CERTIFICATION PROGRAM

- Based on the FAIR Metrics Evaluator
- Certificates are also FAIR (persistent identifiers, metadata, ...)
- Detailed report providing information on what to improve

FAIR
METRICS
EVALUATOR

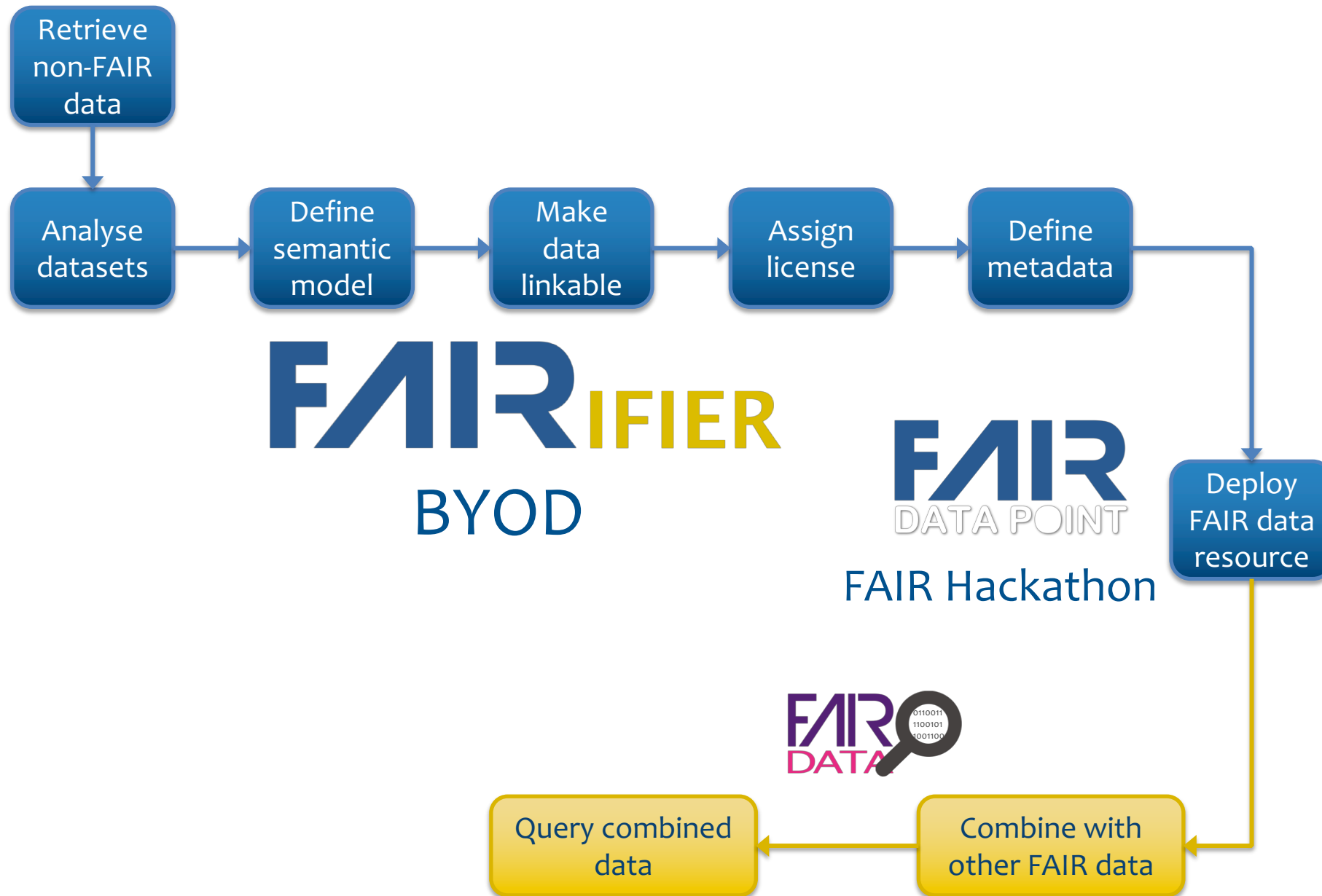


DATA FAIRPORT



FAIR DATA ECOSYSTEM

DSW
Data Stewardship Wizard



FAIR
METRICS
EVALUATOR



Luiz Bonino

International Technology Coordinator – GO FAIR
Associate Professor BioSemantics – LUMC

E-mail: luiz.bonino@go-fair.org

Skype: [luizolavobonino](https://www.skype.com/people/luizolavobonino)

Web: www.go-fair.org