CC BY

doi:10.5281/zenodo.14174616

CBGP
CENTRO DE BIOTECNOLOGÍA Y GENÓMICA DE PLANTAS
UPM-INIA/CSIC

EXCELENCIA SEVERO OCHOA

Implementing FAIR Principles

Problems and Progress

https://tinyurl.com/WEBIST-FAIR

https://doi.org/10.5281/zenodo.14174616

Mark D Wilkinson
Isaac Peral Senior Investigator
CBGP UPM-INIA/CSIC
Universidad Politécnica de Madrid

mark.wilkinson@upm.es

Keynote to WEBIST 2024
International Conference on Web Information Systems and Technologies
Nov 17, 2024

POLITÉCNICA

INIA Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria

CSIC Consejo Superior de Investigaciones Científicas

ETSIAAB ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA ALIMENTARIA Y DE BIOSISTEMAS

www.cbgp.upm.es

Centro de Biotecnología y Genómica de Plantas
(CBGP, UPM-INIA/CSIC)

Framing my presentation

FAIR is intended to help machines help people!

Replace time spent in data discovery/manipulation with time spent on thoughtful investigation of global data and knowledge

IT WORKS! We can now prove it!

We have also experienced some notable failures…

Start with the Problems: Measuring FAIRness

Am I FAIR?

How can I know?

SCIENTIFIC DATA 2016

Now with ~15,000 citations

More detail »

Comment | OPEN

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons - Show fewer authors

The FAIR Guiding Principles...

"This necessitates machines to be capable of autonomously and appropriately acting when faced with the wide range of types, formats, and access-mechanisms/protocols that will be encountered during their self-guided exploration of the global data ecosystem."

https://www.nature.com/articles/sdata201618

When I wrote this paragraph, I was imagining a Web of data discovery and exploration agents

> *"This necessitates machines to be capable of autonomously and appropriately acting when faced with the wide range of types, formats, and access-mechanisms/protocols that will be encountered during their self-guided exploration of the global data ecosystem."*
>
> https://www.nature.com/articles/sdata201618

Notable consequence #1

FAIR is, first, a mechanism to guide **automated agents** to discovery of task-relevant data

As a consequence, FAIR is, before all else, about ***metadata***.

> *"This necessitates machines to be capable of autonomously and appropriately acting when faced with the wide range of types, formats, and access-mechanisms/protocols that will be encountered during their self-guided exploration of the global data ecosystem."*
>
> https://www.nature.com/articles/sdata201618

Notable consequence #2

Creating a Web of data that can be ***appropriately*** (re)used by machines necessitates specific data publishing behaviors

These behaviors can be (must be!) concretely described

Software agents then leverage these behaviors to find, access, and appropriately reuse the correct data

Notable consequence #3

This means that FAIRness must also be

measurable

by similar automated agents

(i.e. are you providing sufficient metadata, and providing it in a manner that can be automatically discovered and interpreted to direct an agent to what it needs?)

The FAIR Evaluator

The first fully-automated “agent” for testing FAIRness of a resource

Resulted from early pressure (2017) from e.g. journal editors who wanted to require FAIRness

https://doi.org/10.1038/s41597-019-0184-5

Acknowledgement of Susanna-Assunta Sansone

Several of the following slides are adaptations of slides she created for a presentation we jointly gave a few weeks ago

Susanna-Assunta Sansone
University of Oxford, UK
- University Academic Lead of Research Practice;
- Professor of Data Readiness, Dept of Engineering Science;
- Director, Oxford e-Research Centre.
- R&D Group Leader: https://datareadiness.eng.ox.ac.uk

- ELIXIR Interoperability Platform Co-Lead
- Author of the FAIR Principles, and contributors to a couple of FAIRness assessment tools
- PI of FAIRsharing, and co-lead of FAIR Cookbook, among other FAIR-supporting resources

Motivation for FAIR testing

- FAIR is a set of principles, not a standard
- The narrative is **insufficient** to circumscribe the valid mechanisms to achieve the behaviours they describe

Nevertheless, we need to trust claims such as

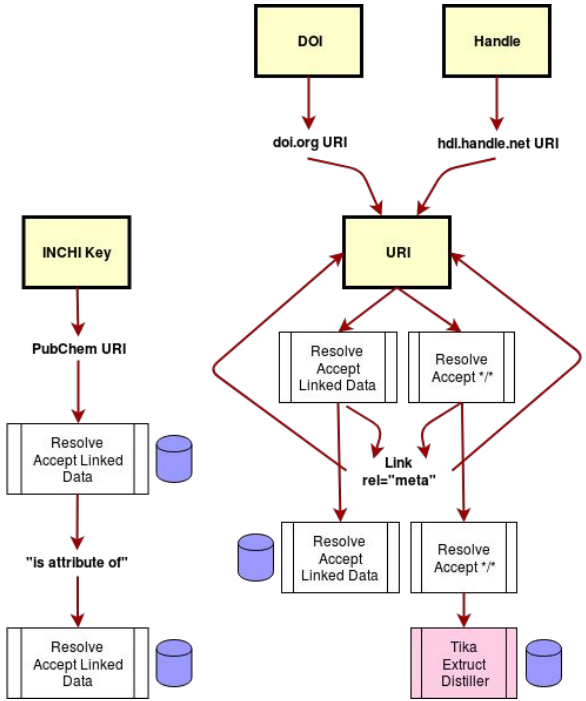
"*I am FAIR*", or "*My data is FAIR*", or "*My repository enables FAIR*"

http://w3id.org/AmIFAIR

Any stakeholder provides the URL of any Digital Object, and ~22 distinct tests of "FAIRness" are executed on that Object

The FAIR Evaluator Harvester

At its core, The Evaluator consists of a “very forgiving” ***metadata*** harvesting workflow & library



The FAIR Evaluator Tests (example)

FAIR Principle F3: metadata explicitly include the identifier of the data it describes

Test Workflow:
Explore the harvested metadata for any metadata facet that *appears to be* a reference to a data record

(there are 18 possibilities that are acceptable)

Evaluator usage

~10,000 FAIRness evaluations run using the public version

Several thousand evaluations run using the private version from my company

Executions of individual tests are not monitored, but do occur frequently

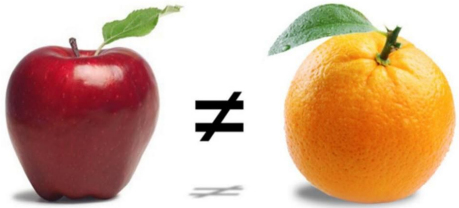
FAIRness is not a competition!
...or is it?

Resource	Execution Type
5 Star Data Rating Tool	Manual - questionnaire
Data Stewardship Wizard	Predictive; based on a manually filled questionnaire
F-UJI	Automated
FAIR Data Self-Assessment Tool	Manual - questionnaire
FAIR Evaluator	Automated
FAIR enough?	Manual - checklist
FAIR-Aware (BETA)	Manual - questionnaire
FAIR-Checker	Automated
FAIRdat	Manual - questionnaire
FAIRness self-assessment grids	Manual - checklist
FAIRshake	Manual - questionnaire, Semi-manual
GARDIAN FAIR Metrics	Manual - checklist
RDA Maturity Model	Manual - checklist

As of Aug 2024, there are **28 independent evaluation, assessment, assistance tools**

(see list at https://fairassist.org)

The tests used and the results given are inconsistent, and not comparable

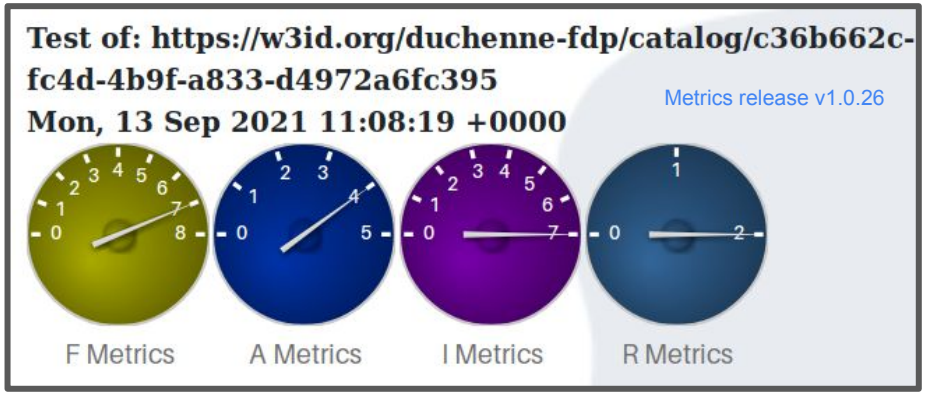


The FAIR Shop

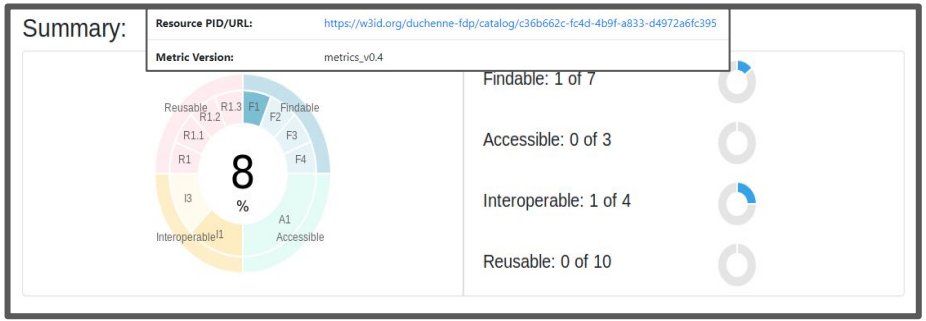
- Mostly questionnaires, few are (semi)automated
- Mostly focused on **metadata**
- They use distinct, often subjective ways of measuring FAIRness
  `
- We lack a set of common, trusted FAIR metrics, and their reproducible FAIR tests

How different can they be?

Comparison of The Evaluator** with F-UJI, on the same Digital Object
(a Catalog record in the Duchenne Parent Project Patient Registry)



20/22 Tests Pass



2/24 Tests Pass

What is “metadata”?

What is “metadata”?

When should links be followed to search for metadata, and which ones?

What is “metadata”?

When should links be followed to search for metadata, and which ones?

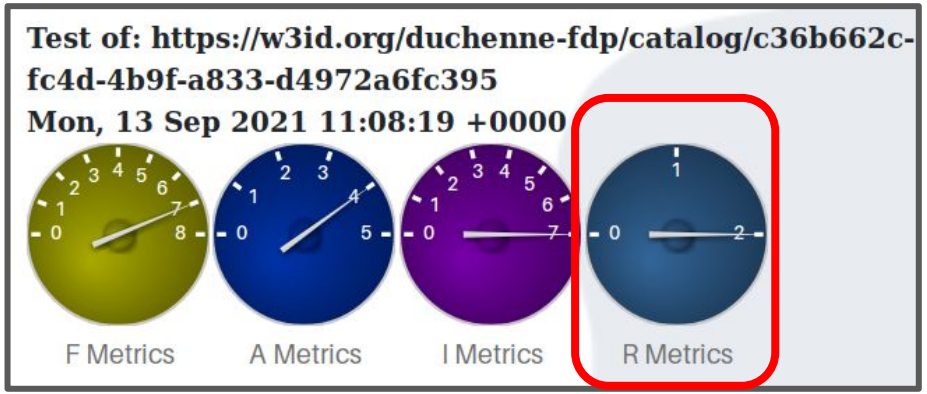
To what extent must a standard be followed?

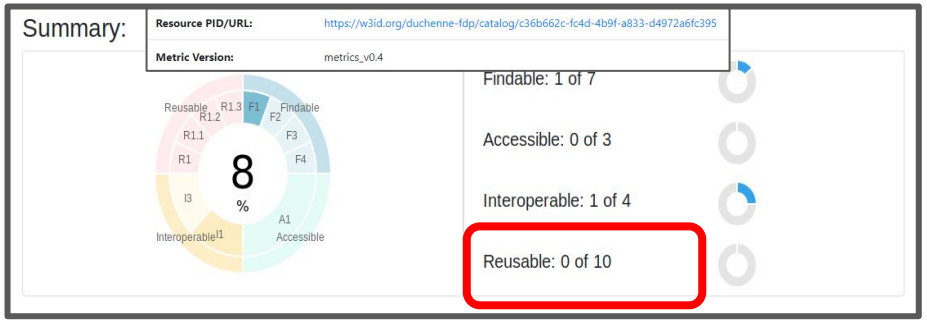
What is “metadata”?

When should links be followed to search for metadata, and which ones?

To what extent must a standard be followed?

Is it acceptable to mix-n-match standards, rather than using one that covers everything?





Not even agreement on what the Principles ***mean***!

EOSC calls for an investigation of the FAIR assessment discrepancies

EOSC FAIR Metrics and Data Quality Task Force

Co-chair

PHASE 1 (2021-2023)

FAIR Metrics and Data Quality TF:
https://eosc.eu/advisory-groups/fair-metrics-and-data-quality

eosc

FAIR-related Task Forces

Outputs formally approved by eosc

DOI 10.5281/zenodo.10797765
Report on FAIR Evaluation community survey
March 8, 2024

DOI 10.5281/zenodo.10490289
Report on FAIR Signposting and its Uptake by the Community
January 11, 2024

DOI 10.5281/zenodo.7515816
TOWARDS A DATA QUALITY FRAMEWORK FOR EOSC
January 9, 2023

DOI 10.5281/zenodo.7463421
FAIR Assessment Tools: Towards an "Apples to Apples" Comparisons
December 20, 2022

DOI 10.5281/zenodo.7390482
Community-driven Governance of FAIRness Assessment: An Open Issue, an Open Discussion
December 1, 2022

Co-chair

Chris Schubert
Head of Data Centre
Climate Change Centre Austria

Three key TF outputs v.v. FAIR Testing

DOI 10.5281/zenodo.7463421

FAIR Assessment Tools: Towards an "Apples to Apples" Comparisons

December 20, 2022

DOI 10.5281/zenodo.10490289

Report on FAIR Signposting and its Uptake by the Community

January 11, 2024

eosc FAIR Metrics and Data Qu https://doi.org/10.5281/zenodo.7390482
Task Force

Community-driven Governance of FAIRness Assessment: An Open Issue, an Open Discussion

Authorship Community:

Mark D. Wilkinson[1,3]
Susanna-Assunta Sansone[2,4]
Eva Méndez[5]
Romain David[2,6]
Richard Dennis[2,7]
David Hecker[2,8]
Mari Kleemola[2,9]
Carlo Lacagnina[1,10]
Anastasija Nikiforova[2,11]
Leyla Jael Castro[12]

Three key TF outputs v.v. FAIR Testing

DOI 10.5281/zenodo.7463421

FAIR Assessment Tools: Towards an "Apples to Apples" Comparisons

December 20, 2022

DOI 10.5281/zenodo.10490289

Report on FAIR Signposting and its Uptake by the Community

January 11, 2024

EOSC FAIR Metrics TF “Apples to Apples” workshops & hackathons (2022-2023) Summary

Addressing the problem collaboratively, with key stakeholders:

- Creators of all automated FAIR assessment tools came together over 6 sessions
- Included several generalist repository representatives (e.g. Dataverse, Zenodo)
- Discussed the bases for the differences in FAIR assessment
- Identified that **metadata discovery and harvesting** was approached differently
  - Different expectations, and different levels of “tolerance” between tools
    - what is considered an “error” versus an “acceptable bit of weirdness”
  - This results in different “substrates” that are then submitted to the test

EOSC FAIR Metrics TF “Apples to Apples” workshops & hackathons (2022-2023) Summary

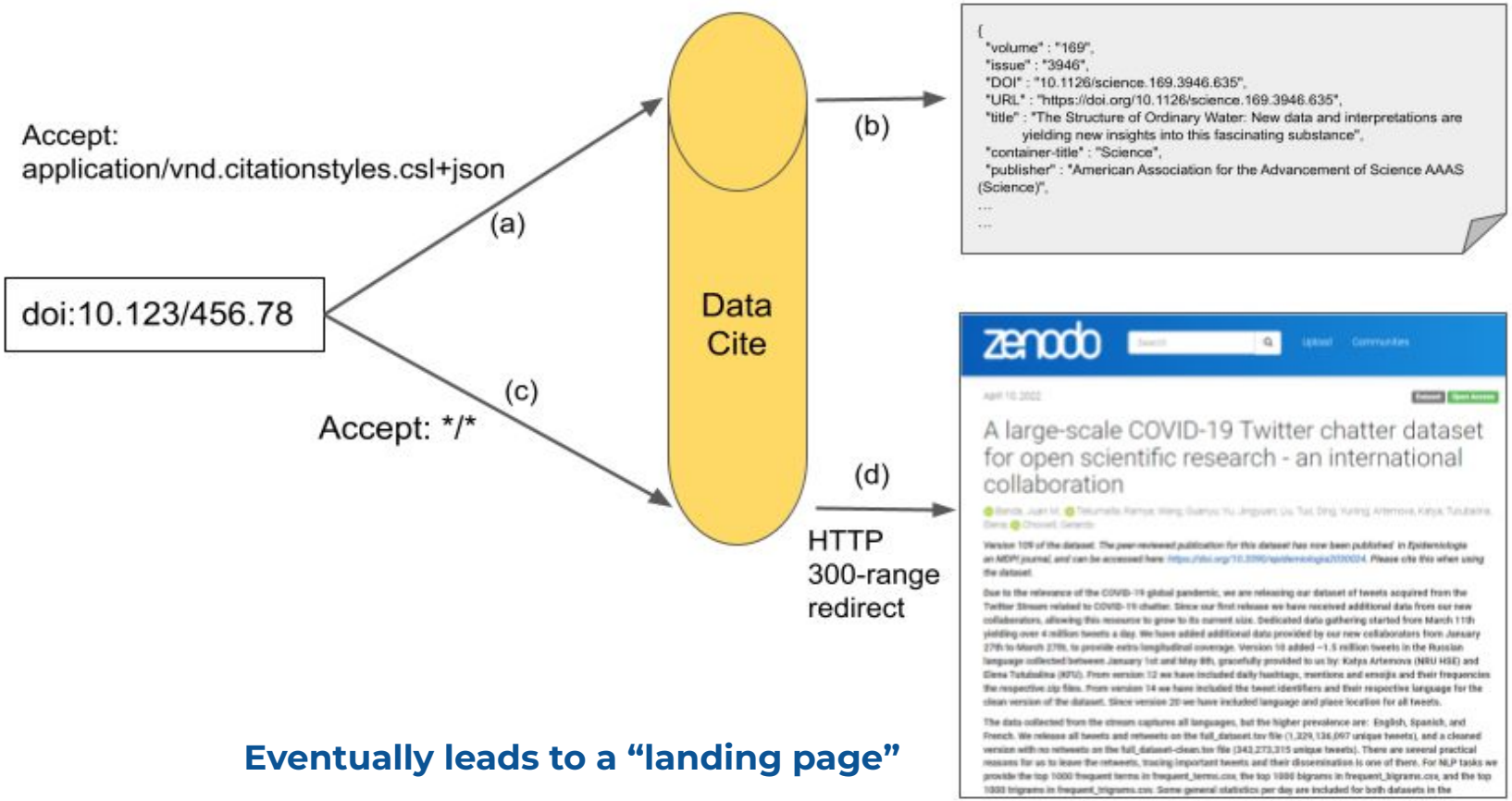
Addressing the problem collaboratively, with key stakeholders:

- Creators of all automated FAIR assessment tools came together over 6 sessions
- Included several generalist repository representatives (e.g. Dataverse, Zenodo)
- Discussed the bases for the differences in FAIR assessment
- Identified that **metadata discovery and harvesting** was approached differently
  - Different expectations, and different levels of “tolerance” between tools
    - what is considered an “error” versus an “acceptable bit of weirdness”
  - This results in different “substrates” that are then submitted to the test

How metadata harvesting becomes a source of discord between the FAIR assessment systems

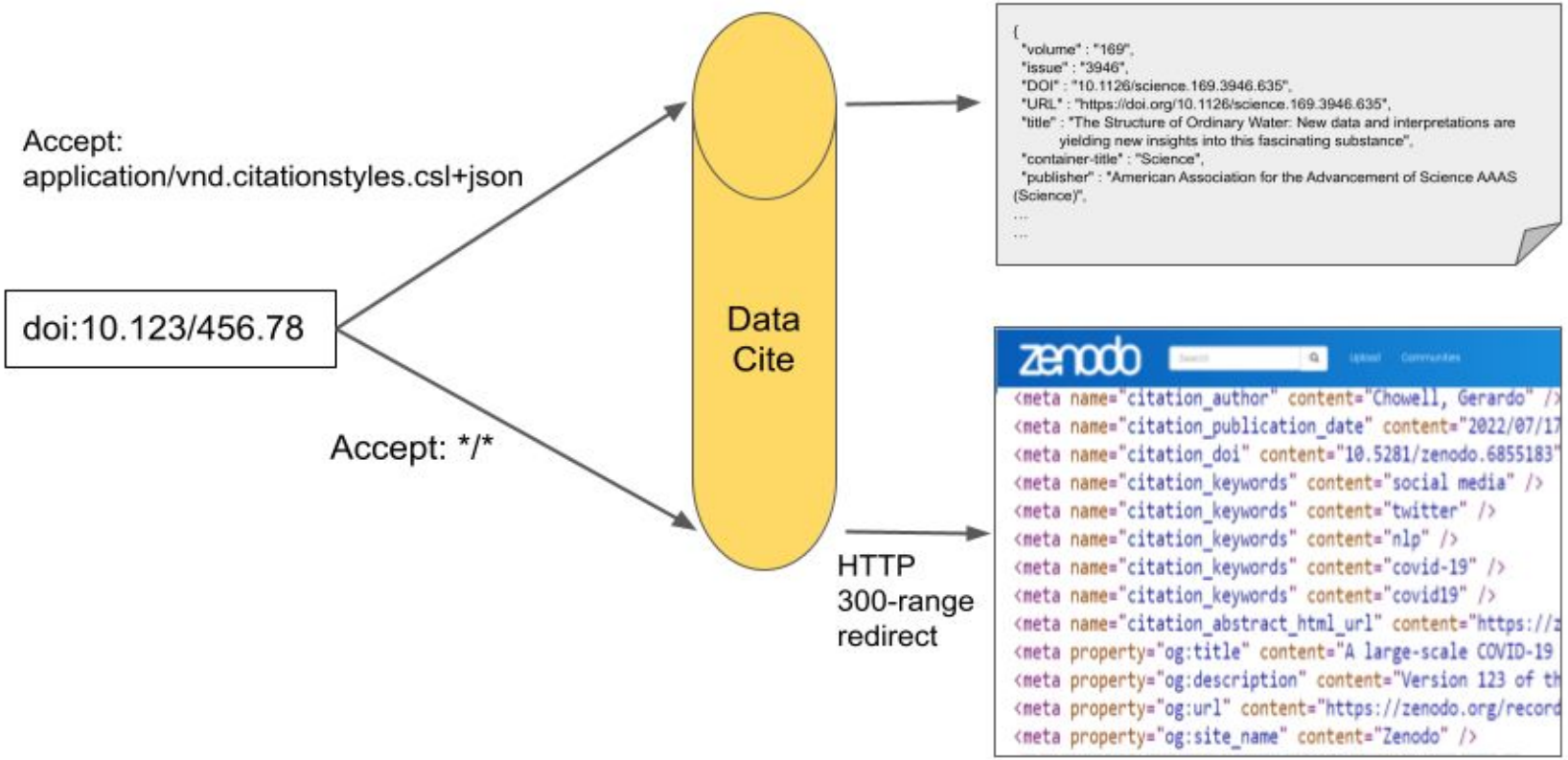
An exemplar case using Zenodo

DOI resolution to metadata



Eventually leads to a “landing page”

DOI resolution to metadata



Landing page embedded metadata

DOI resolution to metadata

HTML “Typed Links”

Pointers to the data records! Excellent!! ...but...

DOI resolution to metadata

HTML “Typed Links”

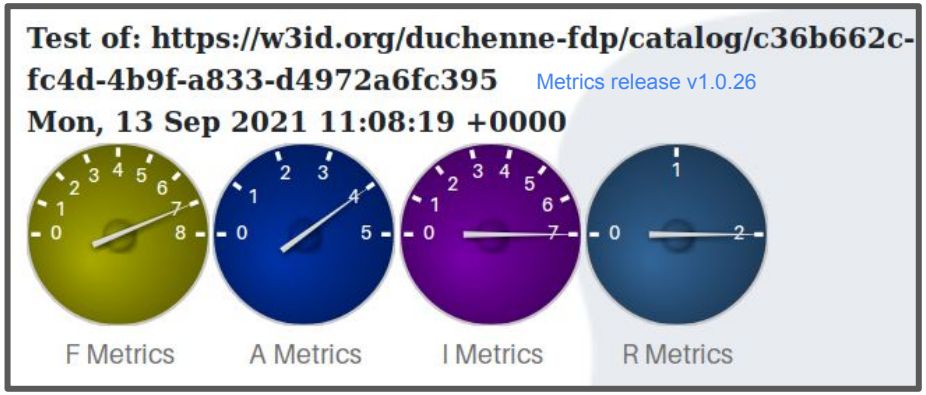
Pointers to the data records! Excellent!! ...but...

*“If the **alternate** keyword is used with the **type** attribute, it indicates that the referenced document is a reformulation of the current document in the specified format.”*

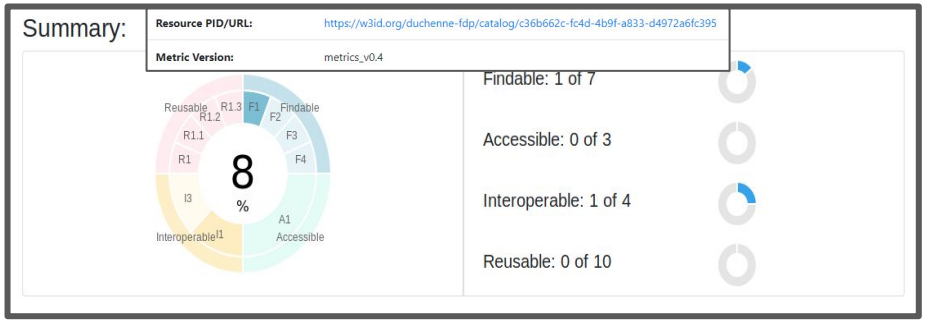
Many sources of ambiguity

- A metadata harvesting agent has to guess (or be coded to know) what to do with a GUID
- There is partial overlap between the DataCite-sourced metadata and Repository metadata
- The use of typed links is a source of ambiguity due to **different interpretations of the spec**
- The interpretation of a “landing page” itself is ambiguous (“**the current document**”?)
  - Some DOIs resolve directly to data, this one resolves to a landing page
  - **What, then, does the DOI represent? The landing page, or the data?**
  - **What is the interpretation of landing-page embedded metadata?**
- **There is no consistent way to discover the data*****
- **There is no consistent way to discover provider-sourced metadata*****
  - **the most important stuff, IMO!!**
- **This is one example, with a widely accepted identifier system and a mature repository**
  - In the “Wild Web” - the “long-tail” of data publishing - unexpected things happen
  - Testing needs to be applied to all kinds of publishers, with widely different expertise

And that's how you end up with this!



20/22 Tests Pass



2/24 Tests Pass

How do you resolve fundamental disagreement among experts?

Hint: Ignore it!

Three key TF outputs v.v. FAIR Testing

DOI 10.5281/zenodo.7463421

FAIR Assessment Tools: Towards an "Apples to Apples" Comparisons

December 20, 2022

DOI 10.5281/zenodo.10490289

Report on FAIR Signposting and its Uptake by the Community

January 11, 2024

The choice of Signposting for achieving FAIR

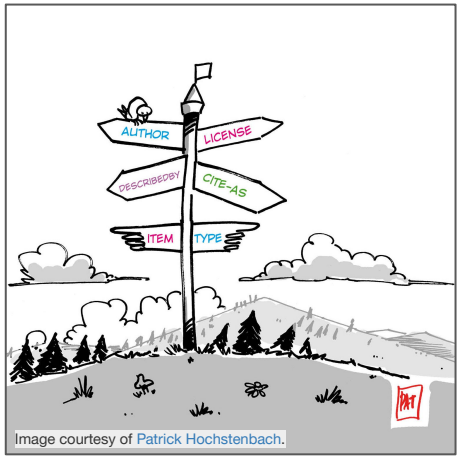


Image courtesy of Patrick Hochstenbach.

With 28 Assessment frameworks there seemed to be too much disagreement

(also no incentive to harmonize!)

The solution, therefore, was to “move the goalposts”!

Rather than trying to harmonize the FAIR Assessment tools, pick a FAIR publishing solution that tool owners agreed **is** completely FAIR

Make sure our assessment tools all score that solution in exactly the same way

FAIR Signposting uses Typed Links

HTML “Typed Links”

FAIR Signposting “Level 1”*

Table 1: Link Relations used by FAIR Signposting	
Relation	**Usage**
cite-as	A one-to-one relationship between the entity and its globally unique identifier
describedby	A one-to-many relationship between the entity and all known metadata records about that entity
item	A one-to-many relationship between an entity representing a deposit and the data file(s) it contains.

These links can appear in:

- The body of the HTML (“Typed Links”)
- The Headers of the HTTP message (“Link Headers”)

Therefore can be used on both Web pages, as well as other non-HTML digital objects

* a subset of “Signposting”, **Van de Sompel & Nelson, 2015: http://doi.org/10.1045/november2015-vandesompel**

FAIR Signposting harvesting workflow

HYPOTHETICAL

I'm showing what we imagine ***could*** be done

Some behaviours are already being implemented

Others will no doubt require more extensive changes

Nevertheless, I think there is some "low hanging fruit" that gives us some critical early wins, particularly w.r.t. FAIR agents and assessment tooling

FAIR Signposting harvesting workflow

Starting Point:

Web Search
Bookmark
DOI resolution
Other ID resolution
…

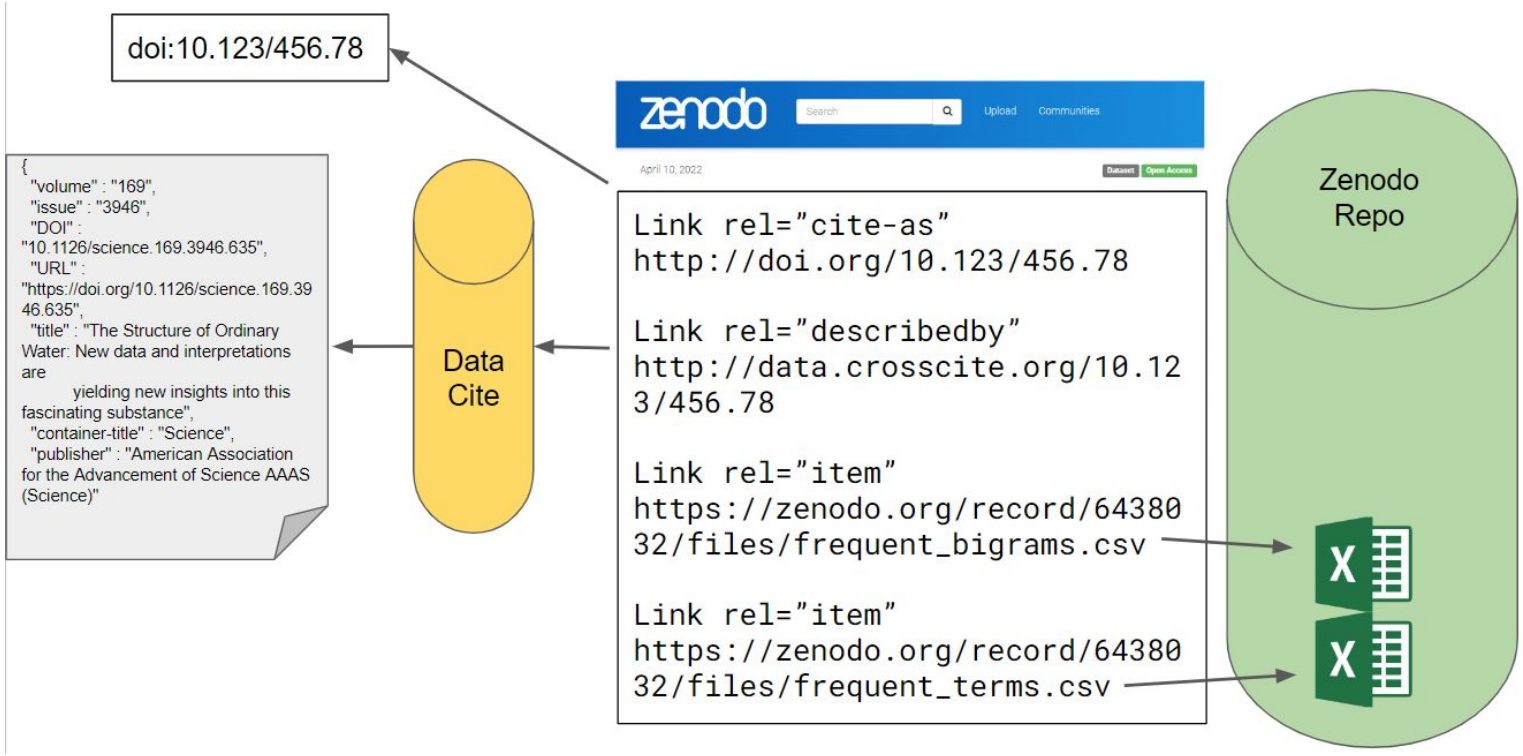
FAIR Signposting harvesting workflow

Starting Point:

Web Search
Bookmark
DOI resolution
Other ID resolution
…

Note that we are not dependent on any specific identifier resolution, beyond HTTP

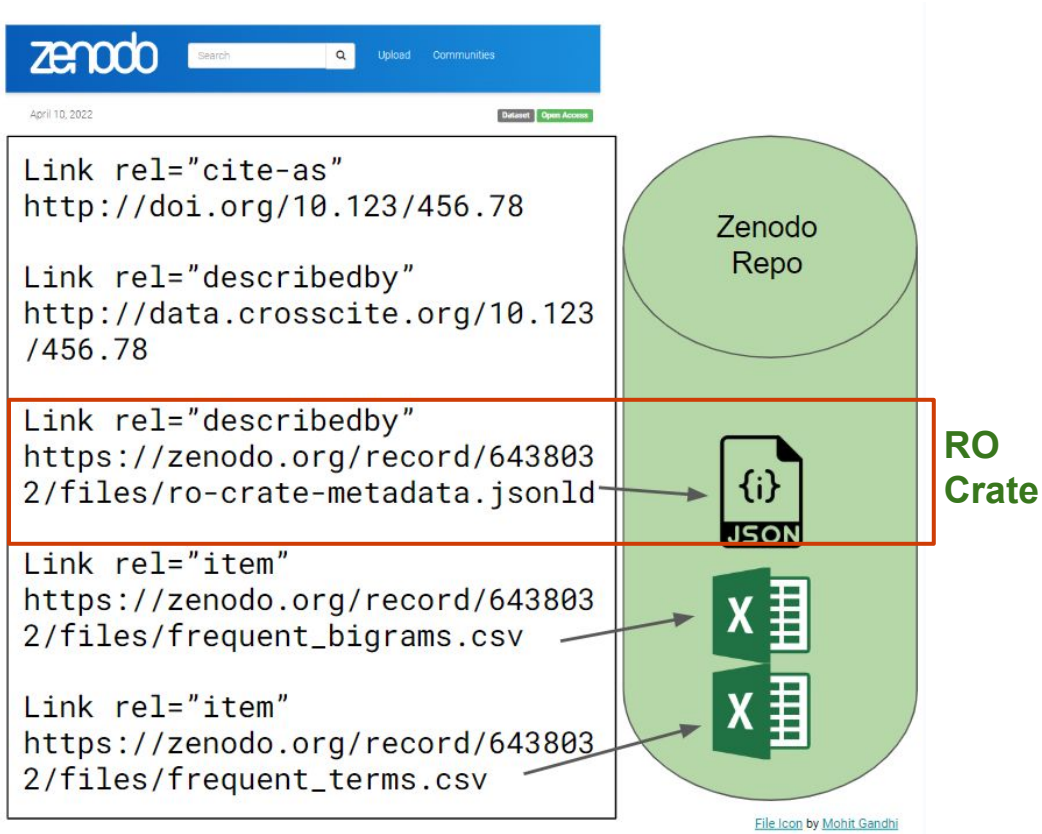
FAIR Signposting harvesting workflow



The *purpose* of the Landing Page is now unambiguous
It is a “broker” pointing at all other entities required in a FAIR record

FAIR Signposting harvesting workflow

Perhaps most importantly, there is an unambiguous **mechanism** for discovering a data provider's own contextual metadata about the record they have deposited



FAIR Signposting harvesting workflow

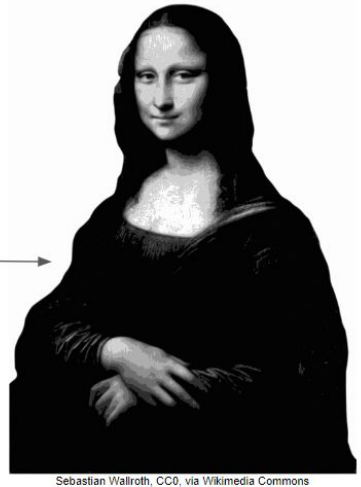
HTTP
Link Headers

```
Link rel="cite-as"
https://upload.wikimedia.org/wikipedia/commons/9/91/Mona_Lisa_vectorized.svg

Link rel="describedby"
https://commons.wikimedia.org/wiki/File:Mona_Lisa_vectorized.svg#metadata
```

Starting Point:

- Web Search
- Bookmark
- DOI resolution
- Other ID resolution
- …



Sebastian Wallroth, CC0, via Wikimedia Commons

We can do the same thing with other kinds of digital objects, through HTTP Link Headers

Most importantly - professionalism!

Benchmarks for Apples-to-Apples FAIR Signposting

These are the Apples-to-Apples FAIR Signposting benchmark tests for tools to verify parsing and compliance with the FAIR Signposting profile.

Benchmarks

- 01-http-describedby-only/
- 02-html-full/
- 03-http-citeas-only/
- 04-http-describedby-iri/
- 05-http-describedby-citeas/
- 06-http-citeas-describedby-item/
- 07-http-describedby-citeas-linkset-json/
- 08-http-describedby-citeas-linkset-txt/
- 09-http-describedby-citeas-linkset-json-txt/
- 10-http-citeas-not-perma/
- 11-http-describedby-iri-wrong-type/
- 12-http-item-does-not-resolve/
- 13-http-describedby-with-type/
- 14-http-describedby-citeas-linkset-json-txt-conneg/
- 15-http-describedby-no-conneg/
- 16-http-describedby-conneg/
- 17-http-citeas-multiple-rels/
- 18-html-citeas-only/

eosc

- We have 78+ Benchmarks
  - positive examples and negative examples
- Challenge metadata harvesting workflows
  - to ensure that they are all working in exactly the same way
- First step in harmonization of FAIR assessments
  - this is leading towards a **governance mechanism** for FAIR assessment

Work in progress but uptake of signposting has begun!

A collaboration between:

https://datascience.nih.gov/data-ecosystem/generalist-repository-ecosystem-initiative

Note that this is being considered as the metadata infrastructure for the European Health Data Space

Returning to this list of ambiguities

- A metadata harvesting agent has to guess (or be coded to know) what to do with a GUID
- There is partial overlap between the DataCite-sourced metadata and Repository metadata
- The use of typed links is a source of ambiguity due to different interpretations of the spec
- The interpretation of the “landing page” itself is ambiguous
  - Some DOIs resolve directly to data, this one resolves to a landing page
  - What, then, does the DOI represent? The landing page, or the data?
  - What is the interpretation of landing-page embedded metadata?
- **There is no consistent way to discover the data*****
- **There is no consistent way to discover provider-sourced metadata*****
  - **the most important stuff, IMO!!**
- This is one example, with a widely accepted identifier system and a mature repository
  - In the “Wild Web” - the “long-tail” of data publishing - unexpected things happen
  - Testing needs to be applied to all kinds of publishers, with widely different expertise

Returning to this list of ambiguities

What are the FAIR evaluation tools actually testing??

- A metadata harvesting agent has to guess (or be coded to know) what to do with a GUID
- There is partial overlap between the DataCite-sourced metadata and Repository metadata
- The use of typed links is a source of ambiguity due to different interpretations of the spec
- The interpretation of the “landing page” itself is ambiguous
  - Some DOIs resolve directly to data, this one resolves to a landing page
  - What, then, does the DOI represent? The landing page, or the data?
  - What is the interpretation of landing-page embedded metadata?
- **There is no consistent way to discover the data*****
- **There is no consistent way to discover provider-sourced metadata*****
  - **the most important stuff, IMO!!**
- This is one example, with a widely accepted identifier system and a mature repository
  - In the “Wild Web” - the “long-tail” of data publishing - unexpected things happen
  - Testing needs to be applied to all kinds of publishers, with widely different expertise

Returning to this list of ambiguities

In most cases, they are testing the repository!

- A metadata harvesting agent has to guess (or be coded to know) what to do with a GUID
- There is partial overlap between the DataCite-sourced metadata and Repository metadata
- The use of typed links is a source of ambiguity due to different interpretations of the spec
- The interpretation of the “landing page” itself is ambiguous
  - Some DOIs resolve directly to data, this one resolves to a landing page
  - What, then, does the DOI represent? The landing page, or the data?
  - What is the interpretation of landing-page embedded metadata?
- **There is no consistent way to discover the data*****
- **There is no consistent way to discover provider-sourced metadata*****
  - **the most important stuff, IMO!!**
- This is one example, with a widely accepted identifier system and a mature repository
  - In the “Wild Web” - the “long-tail” of data publishing - unexpected things happen
  - Testing needs to be applied to all kinds of publishers, with widely different expertise

Is FAIR fulfilling its objectives?

Are we yet supporting the discovery and appropriate reuse of ***data***?

What incentive do data providers have to make their data FAIR if there's no access to it?

EOSC Task Force Phase 2

Co-chair

eosc

FAIR-related Task Forces

Co-Chair

Elli Papadopoulou,
Athena Research Center

PHASE 1 (2021-2023)

FAIR Metrics and Data Quality TF:
https://eosc.eu/advisory-groups/fair-metrics-and-data-quality

Outputs formally approved by eosc

DOI 10.5281/zenodo.10797765
Report on FAIR Evaluation community survey
March 8, 2024

DOI 10.5281/zenodo.10490289
Report on FAIR Signposting and its Uptake by the Community
January 11, 2024

DOI 10.5281/zenodo.7515816
TOWARDS A DATA QUALITY FRAMEWORK FOR EOSC
January 9, 2023

DOI 10.5281/zenodo.7463421
FAIR Assessment Tools: Towards an "Apples to Apples" Comparisons
December 20, 2022

DOI 10.5281/zenodo.7390482
Community-driven Governance of FAIRness Assessment: An Open Issue, an Open Discussion
December 1, 2022

PHASE 2 (2024-2025)

FAIR Metrics and Digital Objects TF:
https://eosc.eu/advisory-groups/fair-metrics-and-digital-objects-task-force

Goals as set by eosc

- Identify the *limitations* of the current FAIR assessment v.v. testing repositories vs. data

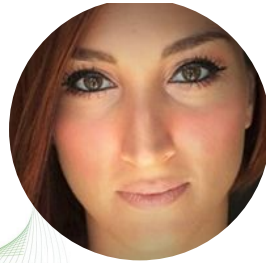
- Identify issues on *data privacy*, considering data usage (machine-actionable data usage policies) and *licensing*

- Examine how provenance is being handled in federated environments

EOSC Task Force Phase 2

Co-chair

FAIR-related
Task Forces



Co-Chair

Elli Papadopoulou,
Athena Research Center

PHASE 2 (2024-2025)

FAIR Metrics and Digital Objects TF:
https://eosc.eu/advisory-groups/fair-metrics-and-digital-objects-task-force

To what degree is FAIRness making a difference at all, v.v. the **reuse of data**?

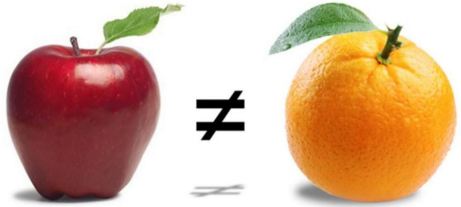
Is there hope on the horizon?

Can we prevent the situation from getting worse?

Resource	Execution Type
5 Star Data Rating Tool	Manual - questionnaire
Data Stewardship Wizard	Predictive; based on a manually filled questionnaire
F-UJI	Automated
FAIR Data Self-Assessment Tool	Manual - questionnaire
FAIR Evaluator	Automated
FAIR enough?	Manual - checklist
FAIR-Aware (BETA)	Manual - questionnaire
FAIR-Checker	Automated
FAIRdat	Manual - questionnaire
FAIRness self-assessment grids	Manual - checklist
FAIRshake	Manual - questionnaire, Semi-manual
GARDIAN FAIR Metrics	Manual - checklist
RDA Maturity Model	Manual - checklist

As of Aug 2024, there are **28 independent tools** (evaluation, assessment, assistance tools, see list at https://fairassist.org

The tests used and the results given are inconsistent, and not comparable



The FAIR Shop

- Mostly questionnaires, few are (semi)automated
- They use distinct, often subjective ways of measuring FAIRness
- We lack a set of common, trusted FAIR metrics, and their reproducible FAIR tests

➡ Along with a **common set of metrics/tests**, we also need a **governance** model to review and adopt new ones, to ensure quality, relevance, value and trust

ostrails

Delivering the Commons to Plan-Track-Assess research in EOSC
(Started in January 2024)

Natalia Manola
CEO OpenAIRE

European Union's Horizon Europe framework programme under grant agreement No. 101130187

The contribution of the EOSC OSTrails project

FAIR Metrics and Tests co-lead

eosc

ostrails

(2024-2027)

Susanna-Assunta Sansone
University of Oxford, UK

FAIR Metrics Governance co-lead

FAIR Metrics and Assessment Workpackage:

- Operationalize the decisions of the EOSC Task Forces
- Harmonize FAIR Assessment environment through standards and APIs
- Design and publish metrics and tests for a range of digital objects beyond “data”, as well as domain-specific assessments
- Provide clarity around the terminology of assessment
- Professionalize the FAIR assessment ecosystem through **good governance**

Harmonize through standards and APIs

Ontology Specification Draft

FAIR Test Results vocabulary

Release: April 8th, 2024

This version:
https://w3id.org/ftr/0.0.1

Latest version:
https://w3id.org/ftr#

Revision:
0.0.1

Authors:
Daniel Garijo
Mark Wilkinson
Rober Huber
Lukas Arnhold
Wim Hugo
Elli Papadopoulou
Leonidas Pispiringas
Allyson Lister
Clement Jonquet
Neil Chue Hong

Source:
https://docs.google.com/document/d/1HusredfHgymRg2ub4L0GnVSRV8IWZvFJyMkE6POejpc/edit?usp=sharing
https://raw.githubusercontent.com/OSTrails/FAIR_assessment_output_specification/main/cqs/cqs.csv

Download serialization:
Format JSON LD | Format RDF/XML | Format N Triples | Format TTL

License:
License http://creativecommons.org/licenses/by/2.0/

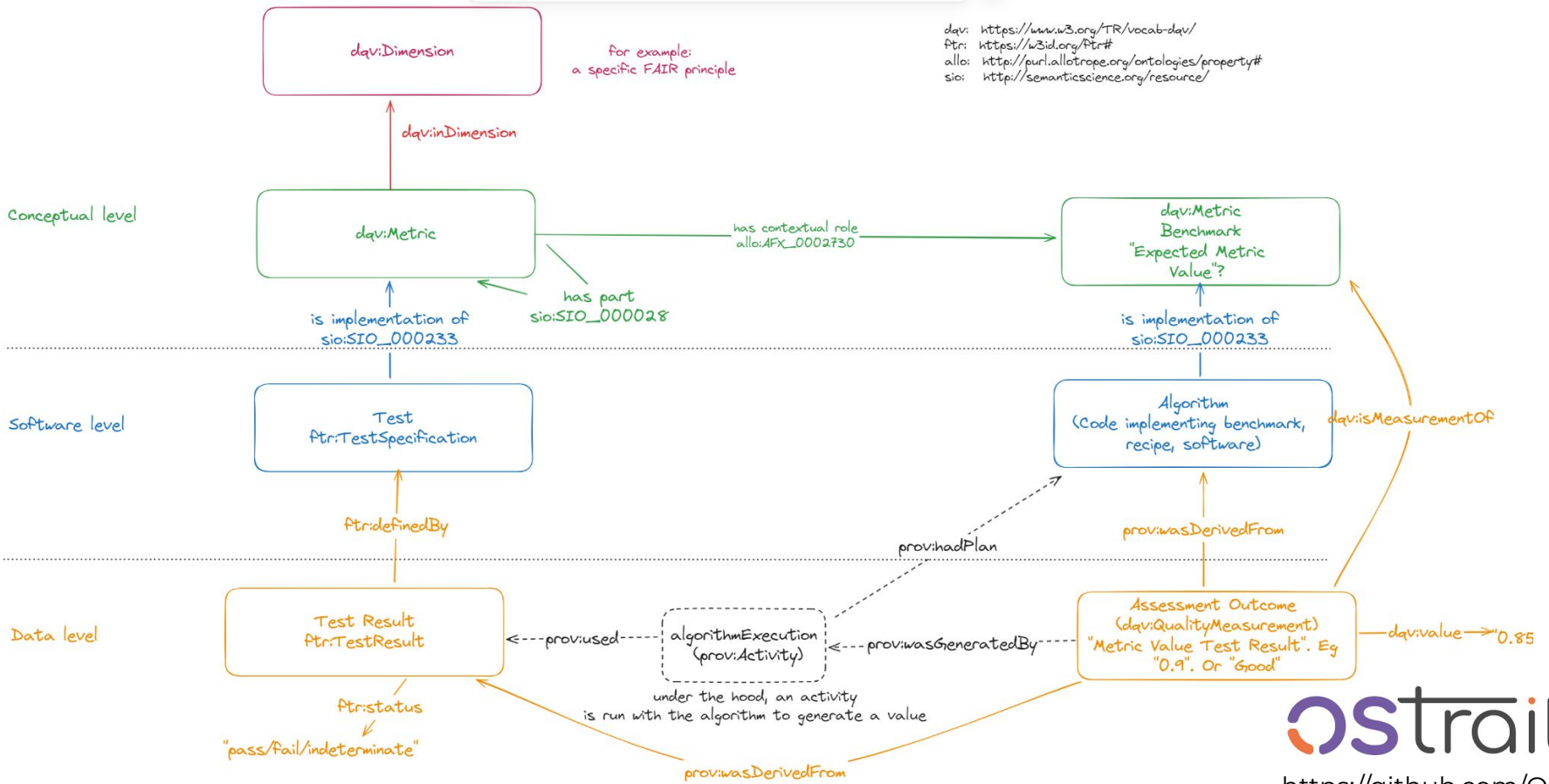
Visualization:
Visualize with WebVowl

Vocabulary maintained at:
https://github.com/OSTrails/FAIR_assessment_output_specification

OSTrails

https://github.com/OSTrails

FAIR Assessment - component terminology

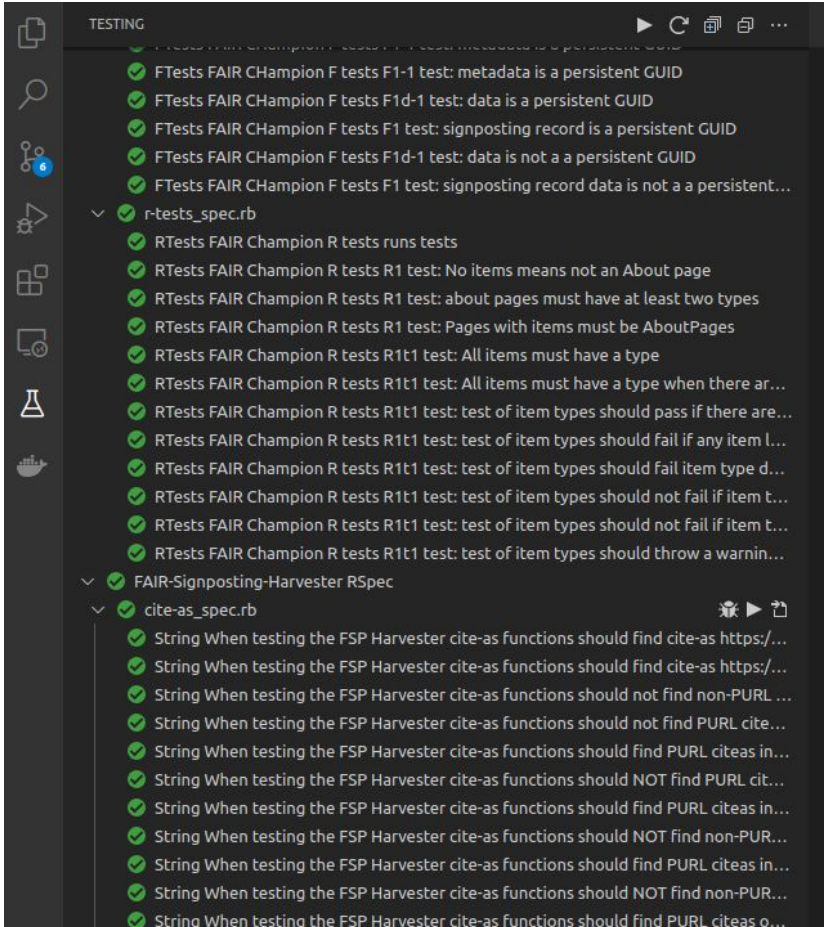


https://github.com/OSTrails

Benchmark Signposting Test Suite for the FAIR Champion* FAIR Assessment Tool

Governance and Professionalism:

ensure that all FAIR tests comply with the standards



*FAIR Champion is the evolution of the FAIR Evaluator, that will be compliant with the OSTrails APIs and standards

https://github.com/OSTrails

Things will get better soon!

End with Success Stories: FAIR works!*

Two examples:

Rare Diseases

Seed Banking

*** with some important caveats**

An example of a successful large-scale FAIRification initiative

The European Joint Programme on Rare Diseases (EJP-RD)

CBGP CENTRO DE BIOTECNOLOGÍA Y GENÓMICA DE PLANTAS UPM-INIA/CSIC

EJP RD

Coordinated by Inserm

EJP RD in numbers

+1800 people

35 participating countries

26 EU MS, 7 associated (AM, CH, GE, IL, NO, RS, TK), UK and CA

ALL 24 ERNs

101 M€ Budget

Union contribution: 55 M€ (70% reimbursement rate)

85% of European RD research community (directly or indirectly) involved in EJP RD

91 beneficiaries

- **10** hospitals
- **12** research institutes
- **31** research funding bodies/ ministries
- **27** universities/hospital universities
- **5** EU infrastructures
- **5** charities/foundations
- EURORDIS

+ 52 linked third parties

+100% associated networks

EURORDIS RARE DISEASES EUROPE · eatris · elixir · INFRAFRONTIER · European Reference Networks · ECRIN EUROPEAN CLINICAL RESEARCH INFRASTRUCTURE NETWORK · BBMRI-ERIC gateway for health

www.cbgp.upm.es

Yanis Mimouni, 2021. https://vascern.eu/wp-content/uploads/2021/10/20211008_EJP-RD_presentation_VASCERN.pdf

EJP RD

Coordinated by Inserm
La science pour la santé From science to health

EJP RD in numbers

+1800 people

35 participa… coun…
26 EU MS, 7 associated (A… GE, IL, NO, RS, TK), UK …

ALL 24 E…

101 Bud…
Union contribution: 55 M€ (70% reimbursement rate)

91 beneficiaries
10 hospitals
12 research institutes
…ding bodies/ ministries
…ospital universities
…res
…dations

…hird parties
…ciated networks

Challenge: Make it all FAIR

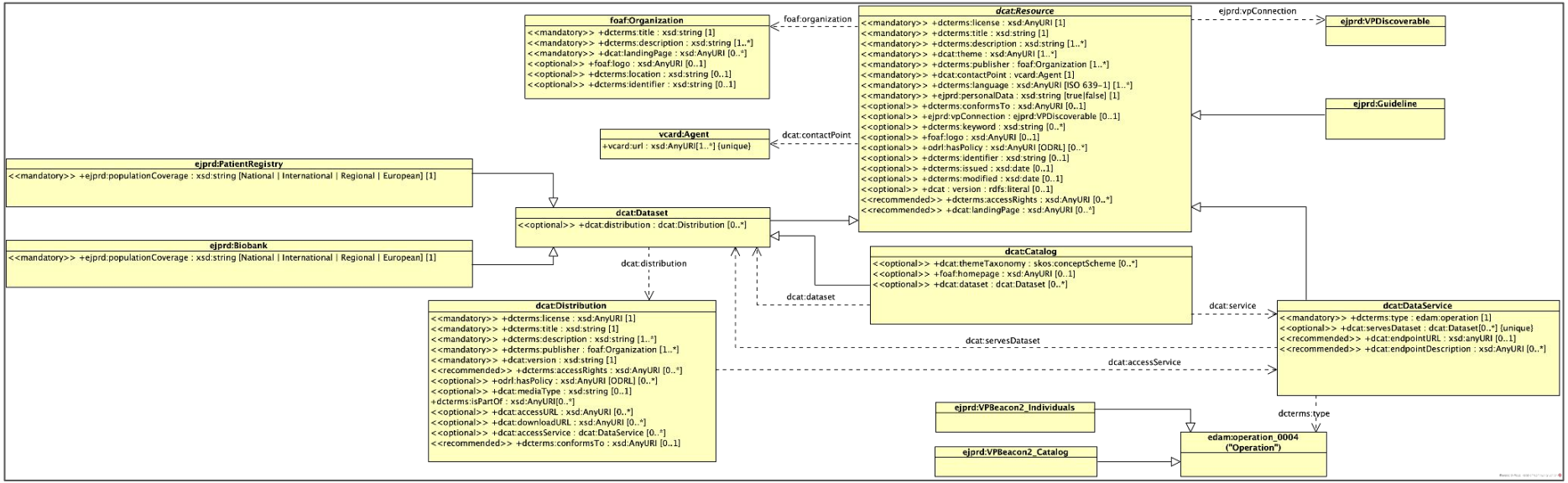
…but how?

EURORDIS RARE DISEASES EUROPE
eatris
elixir
INFRAFRONTIER mouse disease models
European Reference Networks
ECRIN EUROPEAN CLINICAL RESEARCH INFRASTRUCTURE NETWORK
BBMRI-ERIC gateway for health

Yanis Mimouni, 2021. https://vascern.eu/wp-content/uploads/2021/10/20211008_EJP-RD_presentation_VASCERN.pdf

Step 1 - Focus on Metadata that answers the question:

"What kind of data does this partner have"?



EJP-RD Metadata Schema:

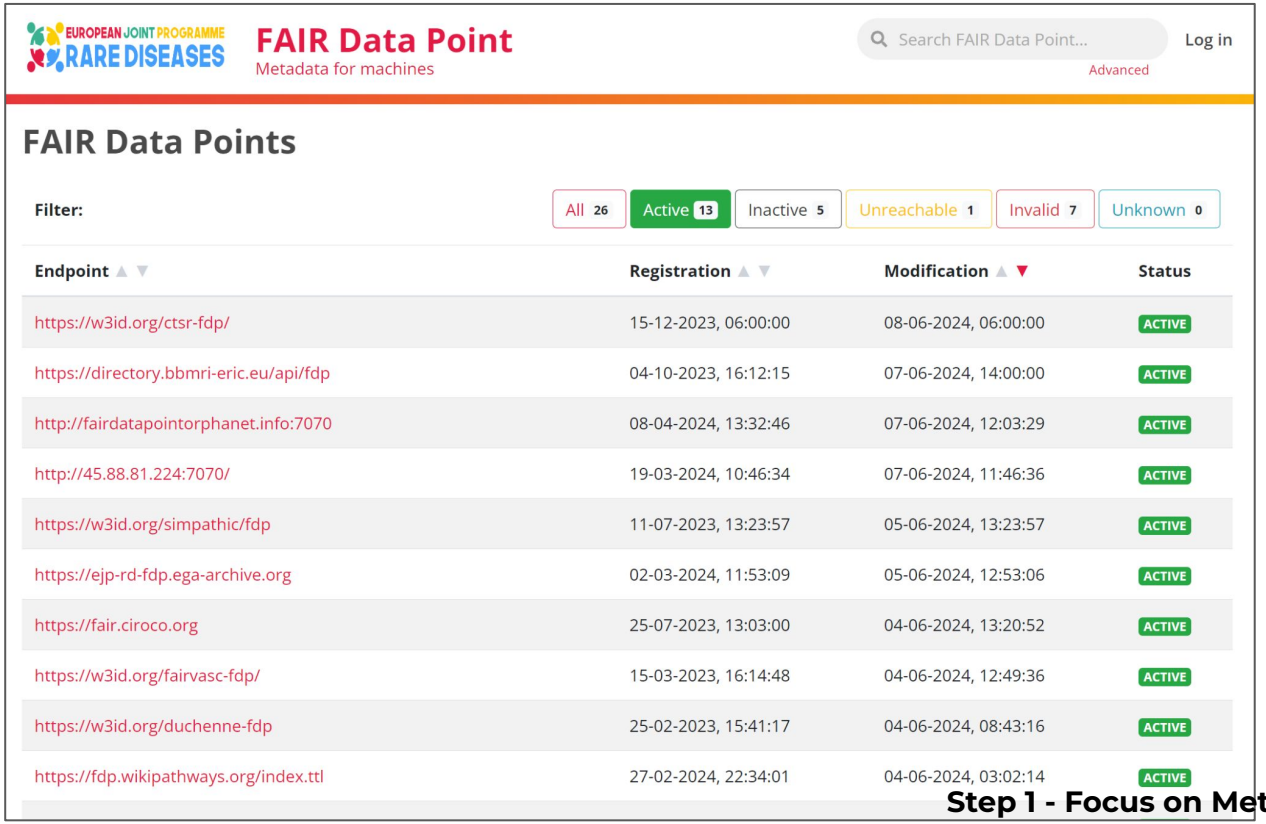
- Based on DCAT; attempts to follow the European DCAT-AP
- Published by all sites via semi-automated process using Excel spreadsheets

EJP-RD Metadata Publication Platform: FDP

- Open-source, currently distributed via Docker images
- Installs in seconds
- Publishes **DCAT metadata** for human exploration
- Publishes **DCAT metadata** for agent-based exploration
- Provides easy-to-use DCAT editing tools via Web pages
- Automatically registers the new FDP in a central index

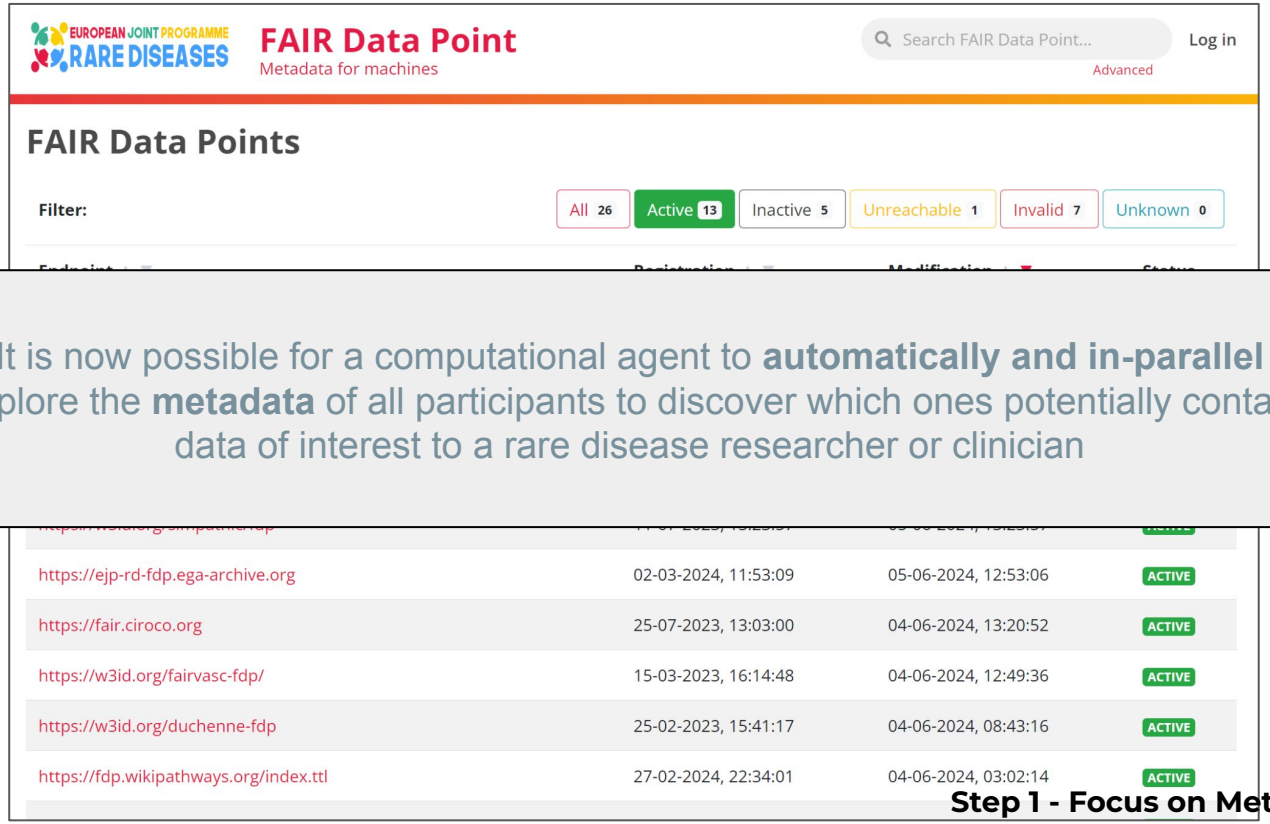
Result: a federated network of FDPs representing all biobanks and patient registries

FAIR Data Point Index



Step 1 - Focus on Metadata

Result: a federated network of FDPs representing all biobanks and patient registries



It is now possible for a computational agent to **automatically and in-parallel** explore the **metadata** of all participants to discover which ones potentially contain data of interest to a rare disease researcher or clinician

Step 1 - Focus on Metadata

FAIR Data

Metadata Metadata Metadata!

Am I suggesting that there's no point in making FAIR data?

No…

but there's little point in working on FAIR data until you get the discovery metadata right!

Some data will never be made FAIR!

FAIR Data

However…

EJP-RD also created FAIR Data

via an end-user-friendly reusable FAIRification pipeline

This led to some amazing examples of data-level interoperability!

Challenge

Need to make all data-focused network partner's (~50) resources work together

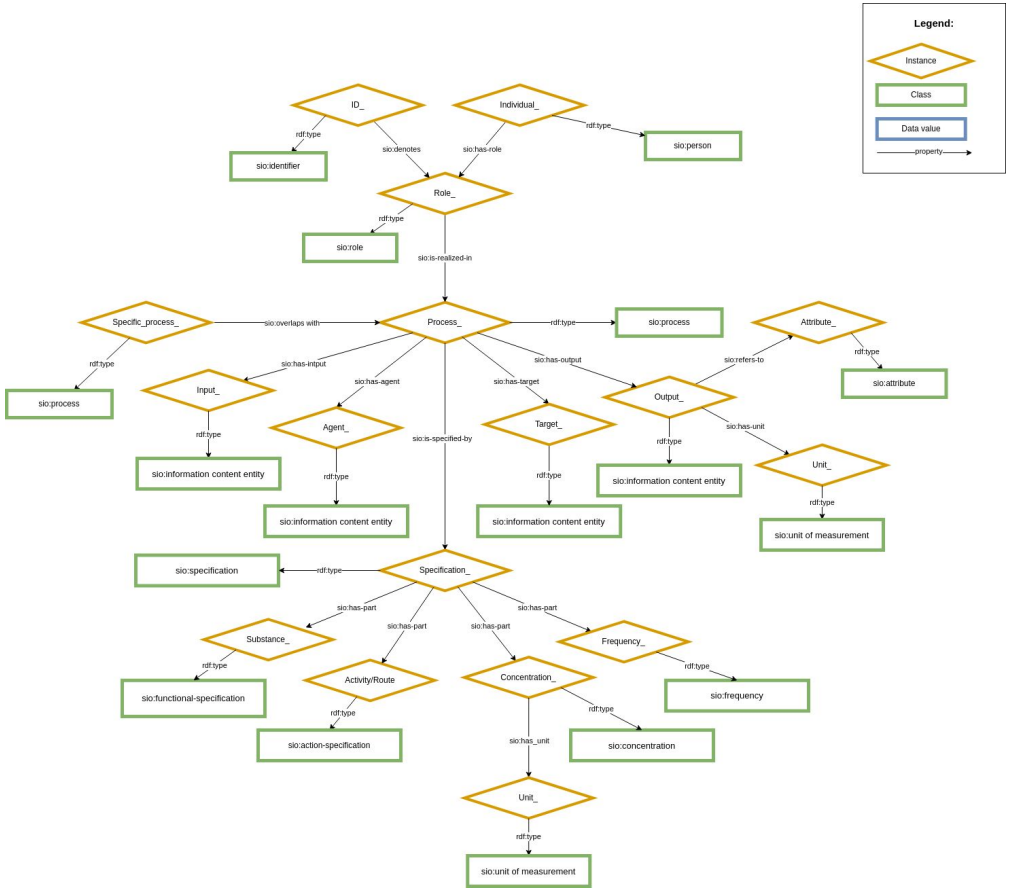
Partners have similar data (patient registry or biobank samples)

Partners have different starting formats

Generally, they are forbidden from sharing or moving their data

So the participants are going to have to do the FAIR transformation themselves, on-site, sometimes without even letting the FAIR experts see it!

Step 2: Build a shared, generic FAIR data model



CARE-SM
Clinical And Registry Entries Semantic Model

Dr. Michel Dumontier, UMaastricht, SemanticScience Integrated Ontology

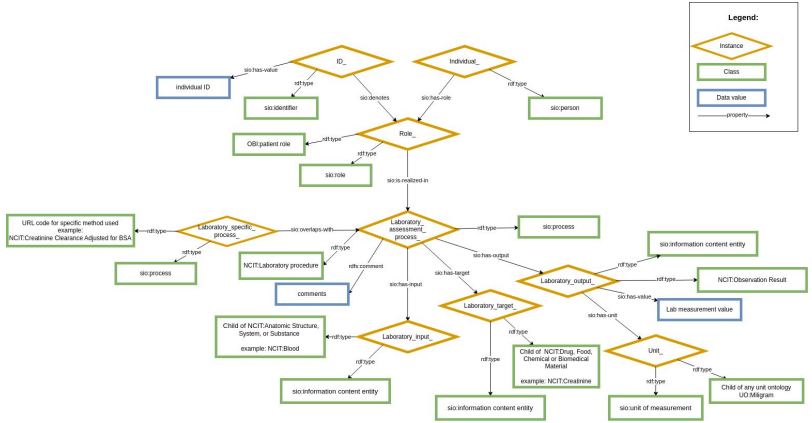
Pablo Alarcón, Clinical and Registry Entries Semantic Model (CARE-SM)

Step 3: Use CSV as a “lingua franca” for all partners

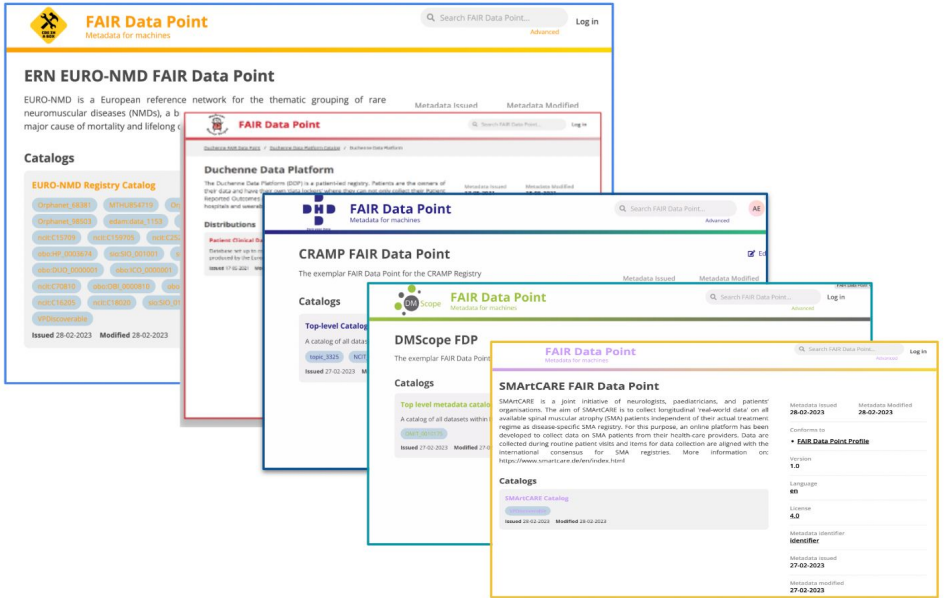
CSV Generated by the participants (easy!)

Property	Observation	Value	Date
Diagnosis	ORPHA:98896	-	2021-02-01
Body Measurement	NCIT_C25208	28kg	2020-04-05
Laboratory Measurement	NCIT_C399	10mg/L	2020-04-05

Fully automated transformation and “publication” of FAIR data into their FAIR Data Point



Does this work, in practice?



Five FAIR Data Points for partners representing NMDs

EURO-NMD Registry Hub
European Reference Network for neuromuscular diseases

WORLD DUCHENNE ORGANIZATION
UPPMD

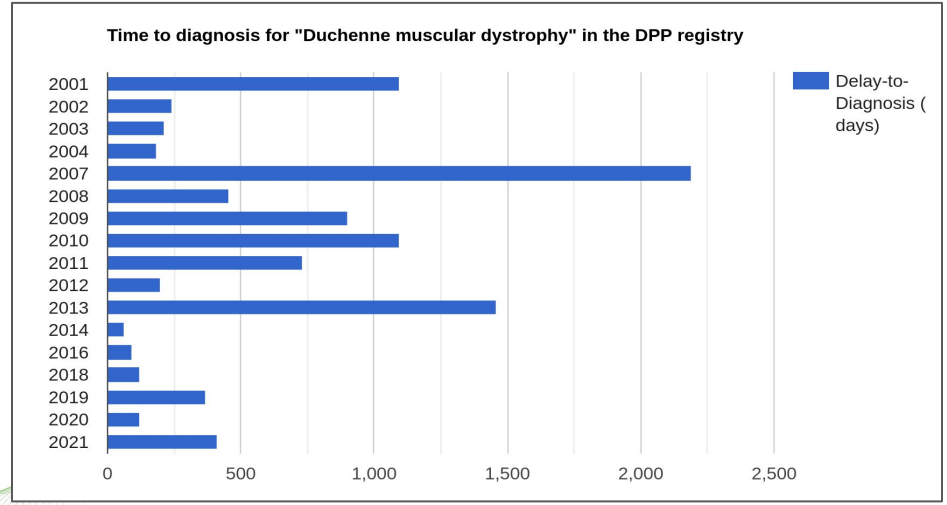
FAIR DATA SYSTEMS

“What is the delay between symptom onset and diagnosis?”

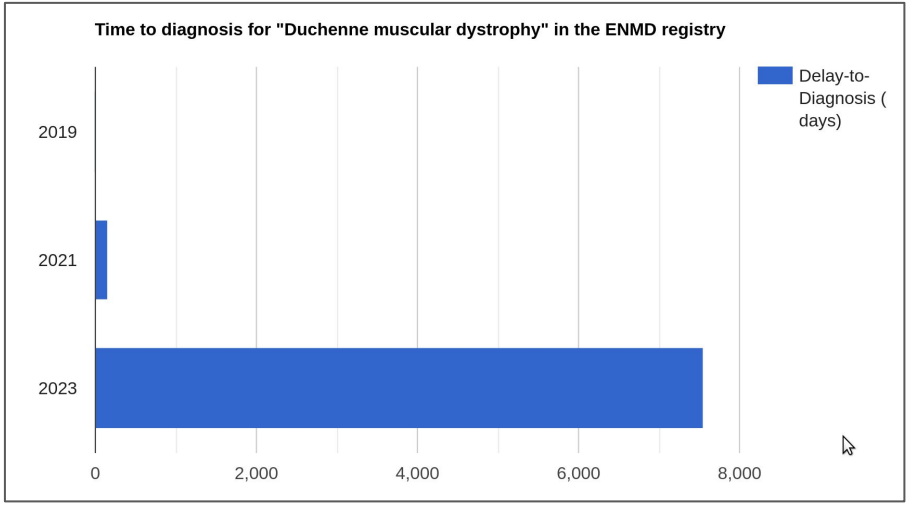
1) The same query* sent to all participants (shared model)

2) Integrate the output

Duchenne Parent Project (DPP)



EURO-NMD (mock data)



* we “cheat” a bit here, with our micro-services…

***Caveat emptor*!!**

FAIR alone is NOT sufficient to achieve interoperability!

This paper compares two independent FAIRification efforts (C-PATH and CARE-SM) over identical data items

Journal of the Society for Clinical Data Management

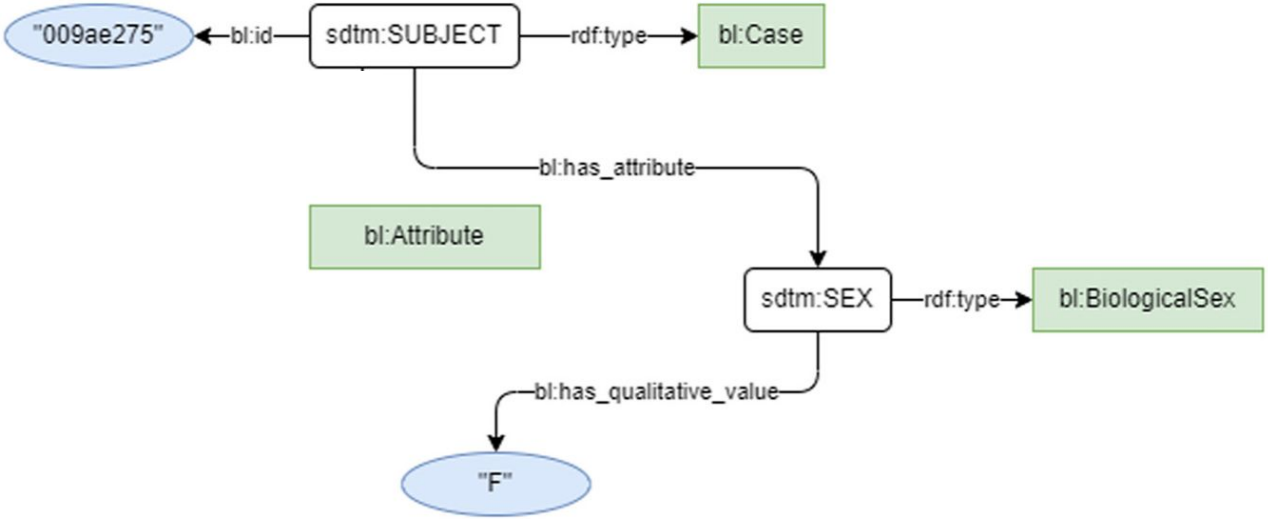
Alarcón-Moreno P, et al. Leveraging Biolink as a FAIR "Rosetta Stone" Between Clinical Semantic Models Provides Emergent Interoperability. *Journal of the Society for Clinical Data Management*. 2022; 2(3): 2, pp. 1–8. DOI: https://doi.org/10.47912/jscdm.130

ORIGINAL RESEARCH

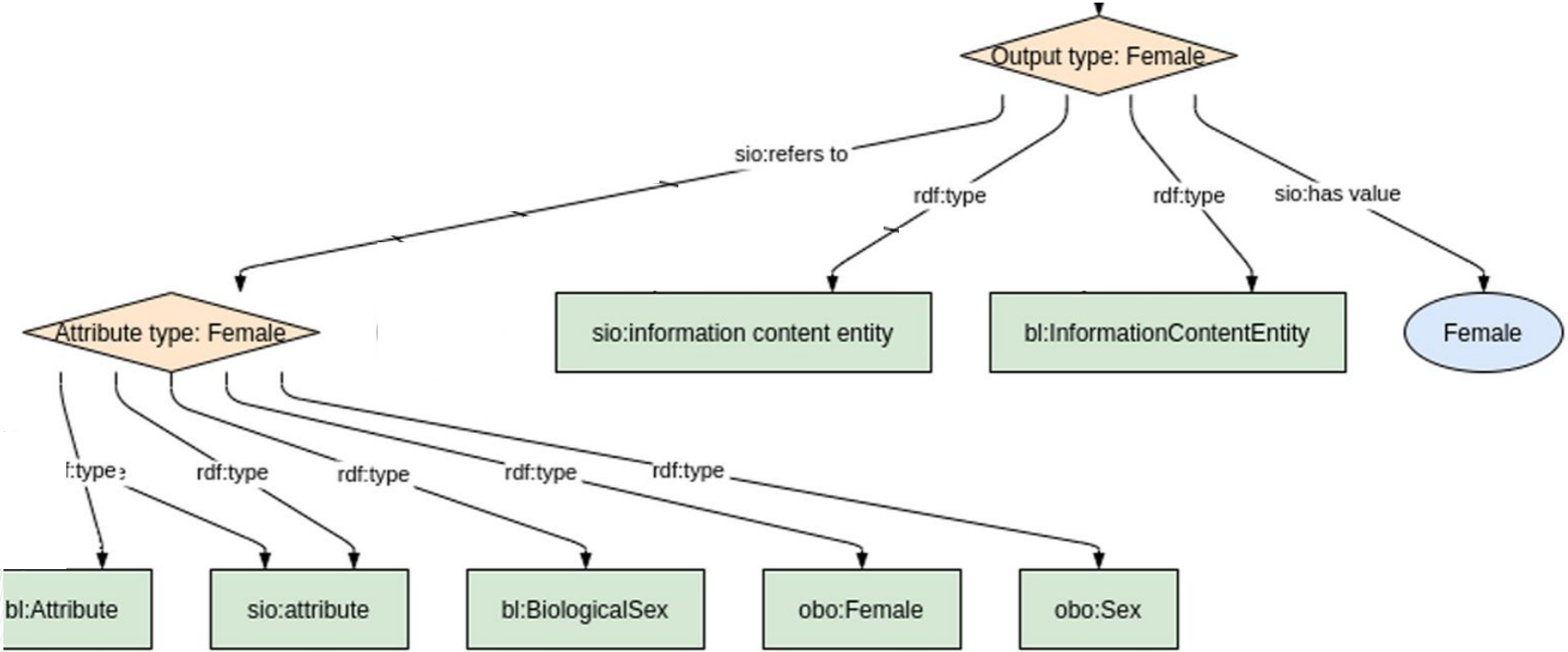
Leveraging Biolink as a FAIR "Rosetta Stone" Between Clinical Semantic Models Provides Emergent Interoperability

Pablo Alarcón-Moreno*, Ian Braun†, Emily Hartley†, Daniel Olson†, Nirupama Benis‡, Ronald Cornet‡, Mark D. Wilkinson* and Ramona L. Walls†

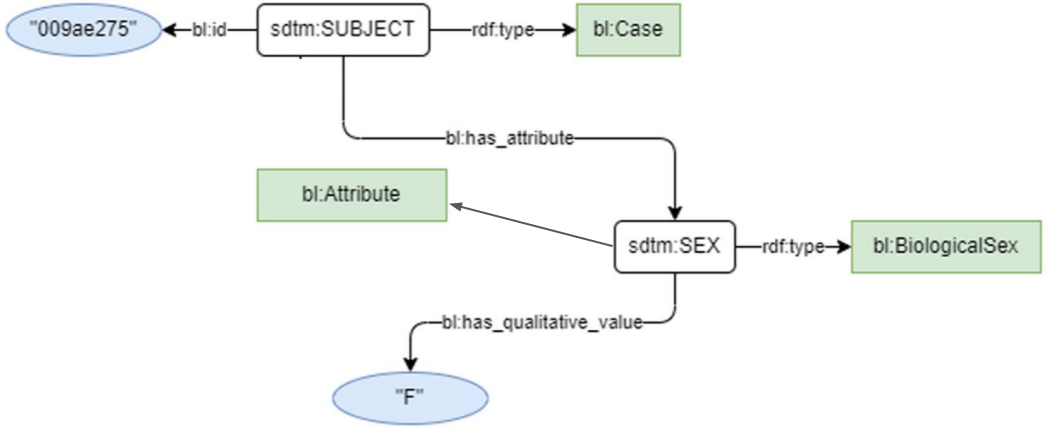
The C-PATH Semantic Model for Biological Sex

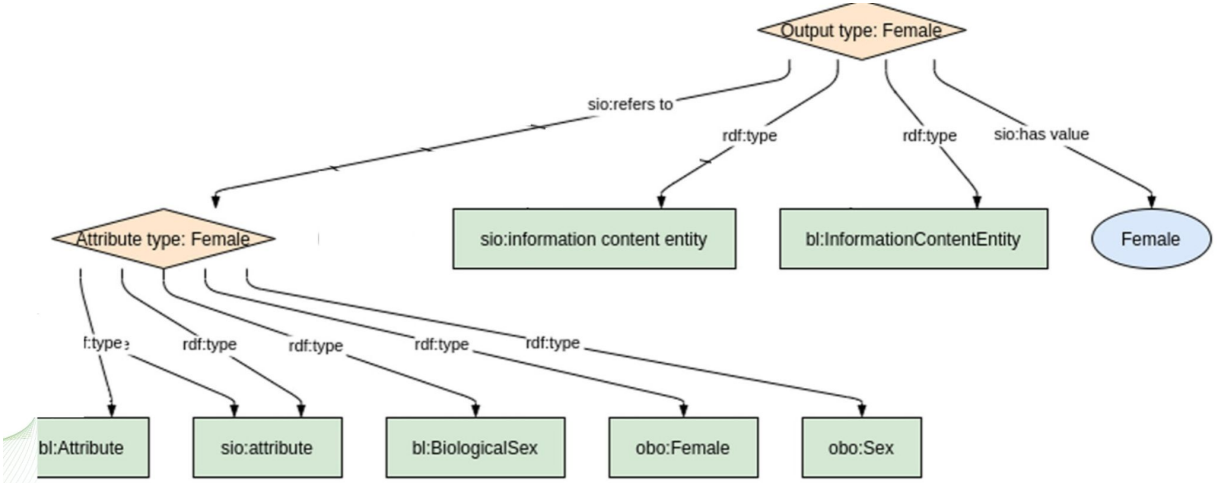


The CARE-SM Semantic Model for Biological Sex

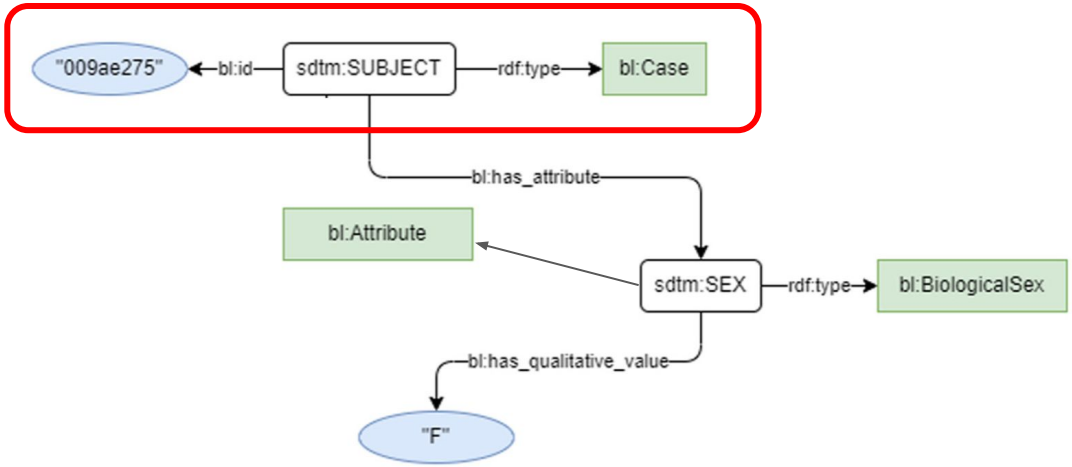


Comparison of the “FAIR” Models

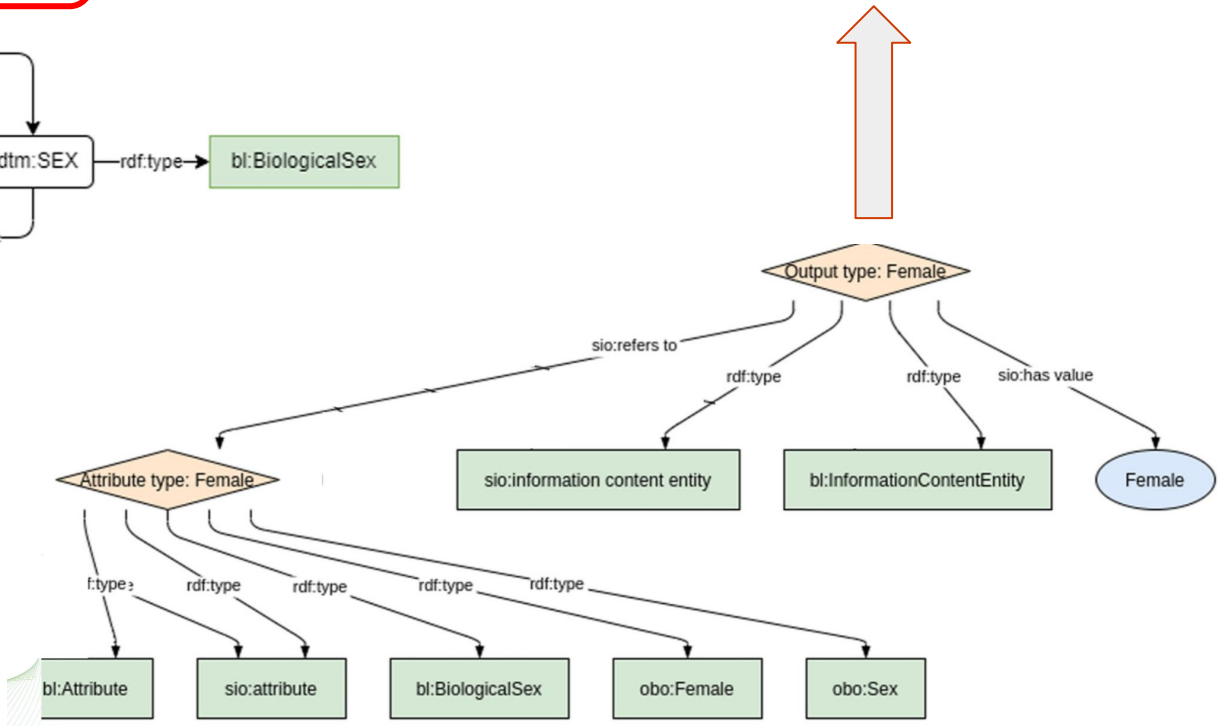




Comparison of the “FAIR” Models

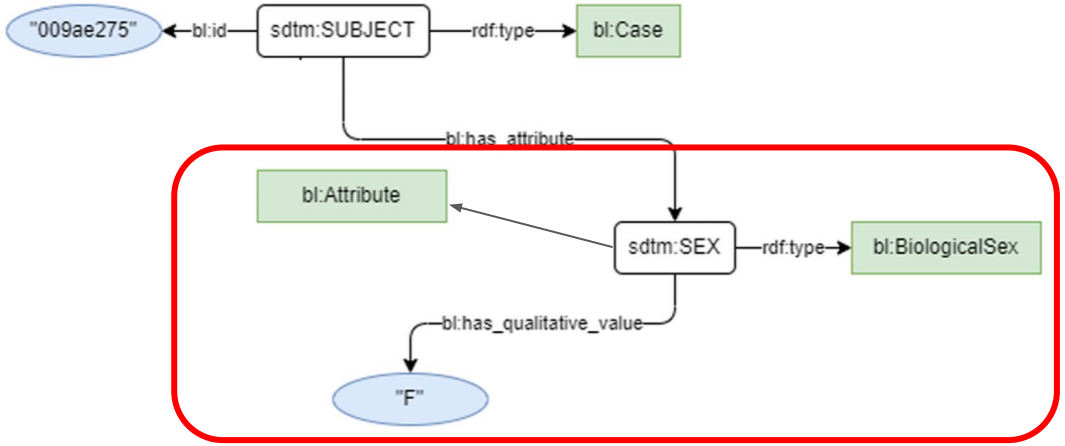


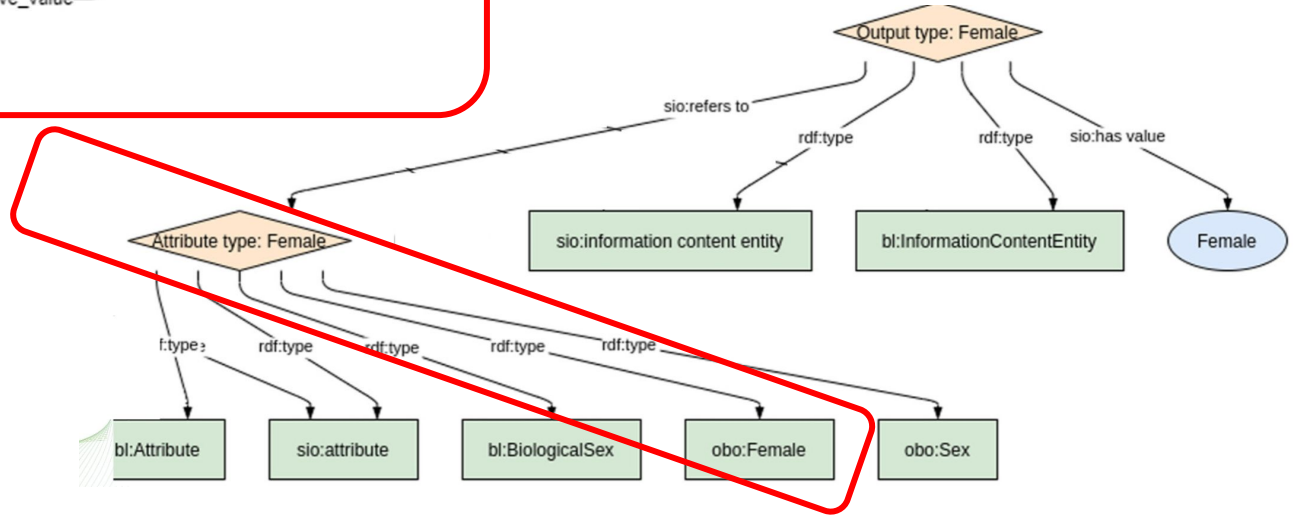
Our “patient” class is waaaay higher up in the model



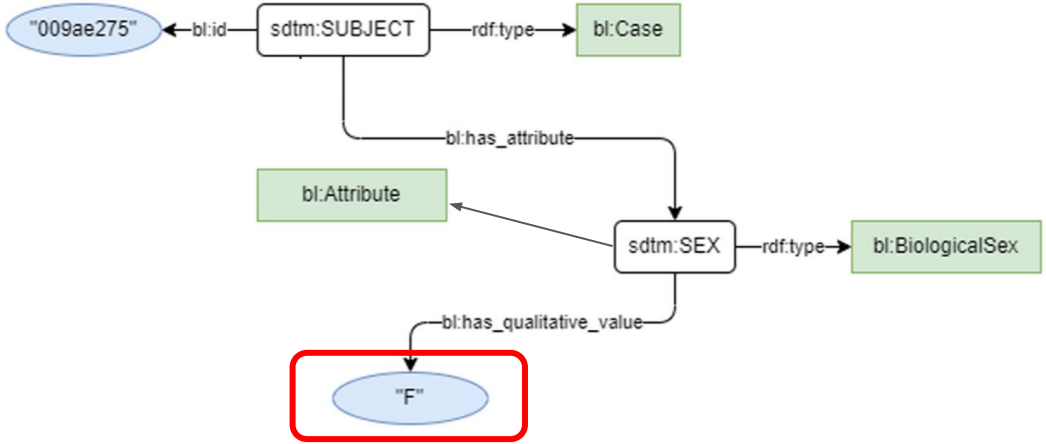
Comparison of the “FAIR” Models

We both have attributes that are “female”



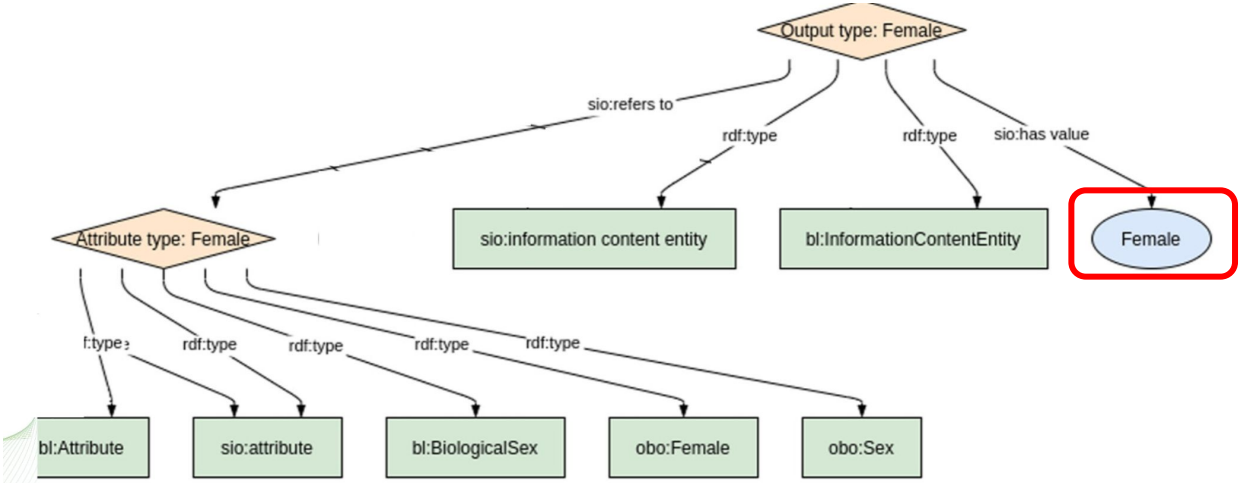


Comparison of the “FAIR” Models

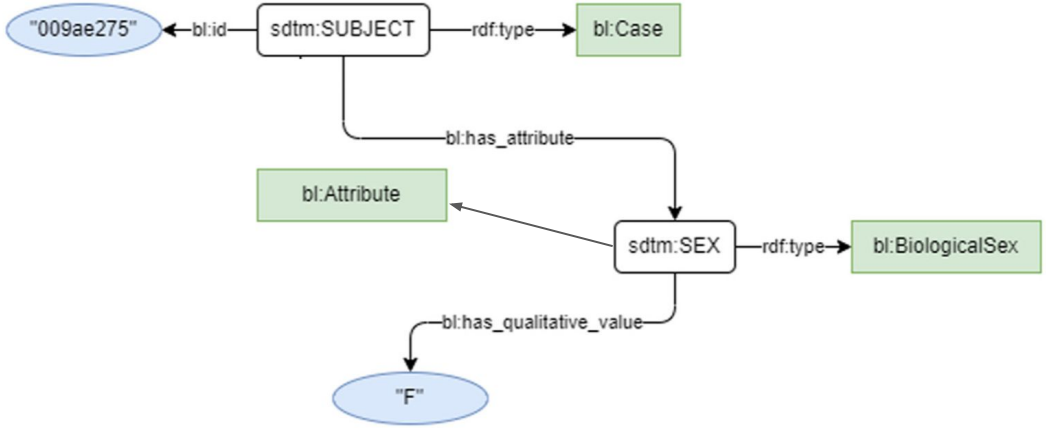


In their model, they use “F” where we use “Female”

…but…

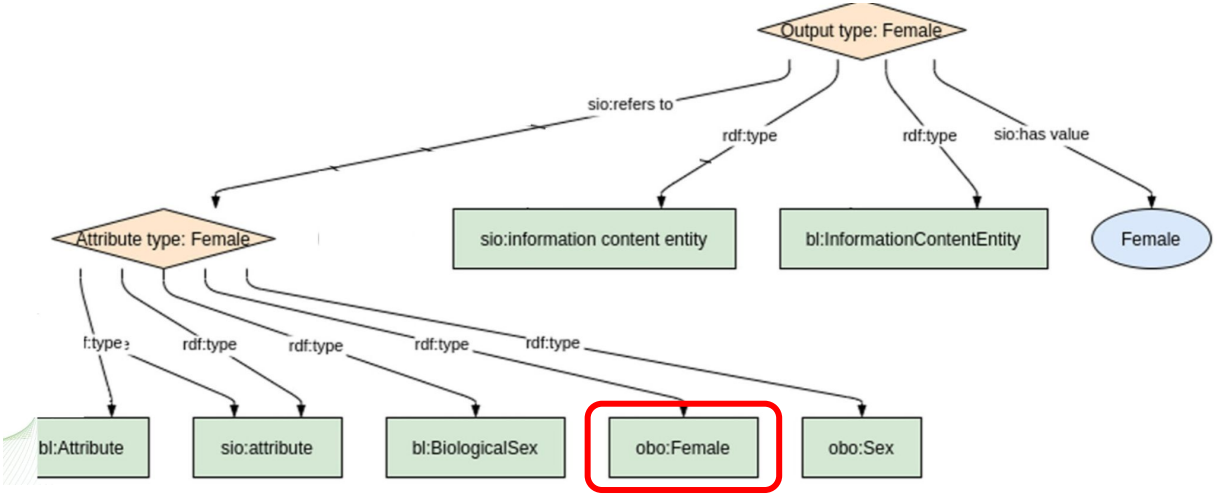


Comparison of the “FAIR” Models

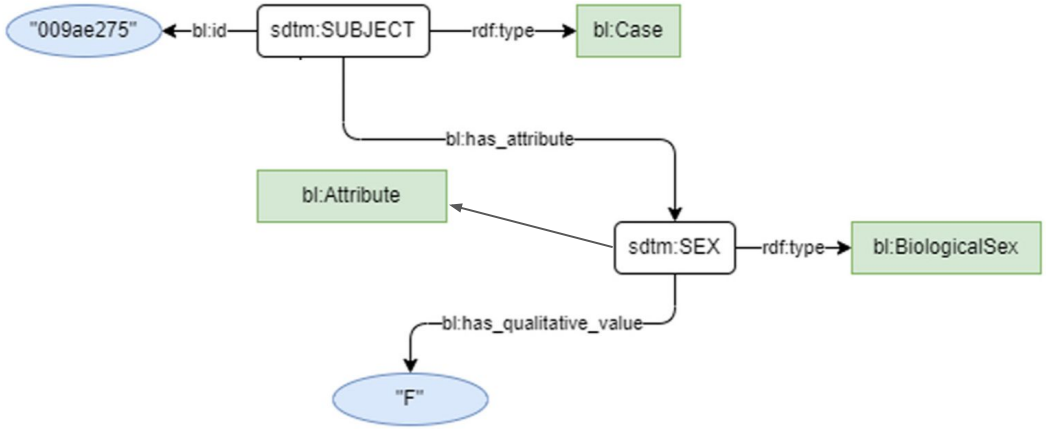


We don’t put any utility in the string “Female” anyway

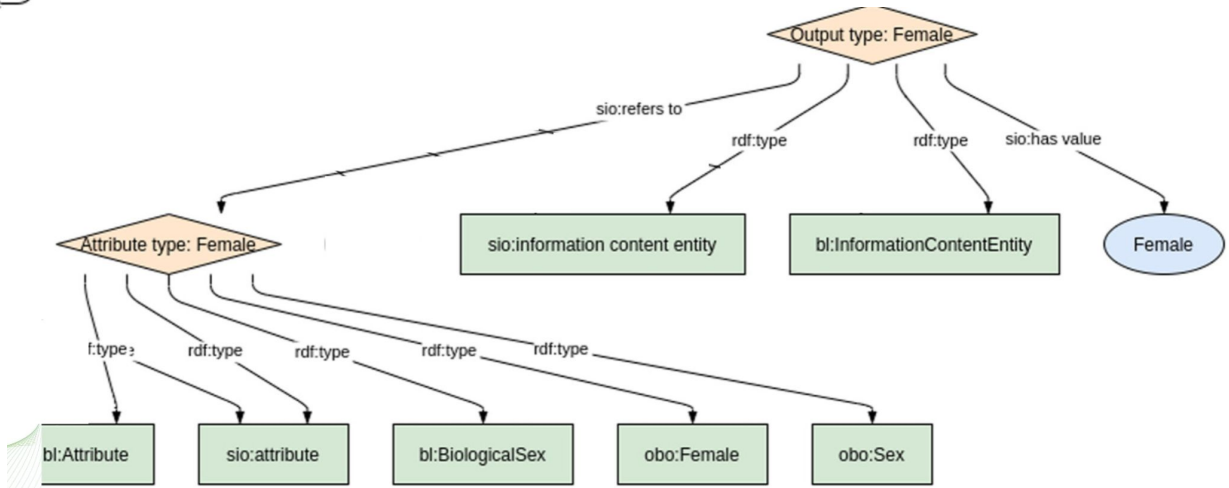
We use the NCIT Ontology term for “Female”



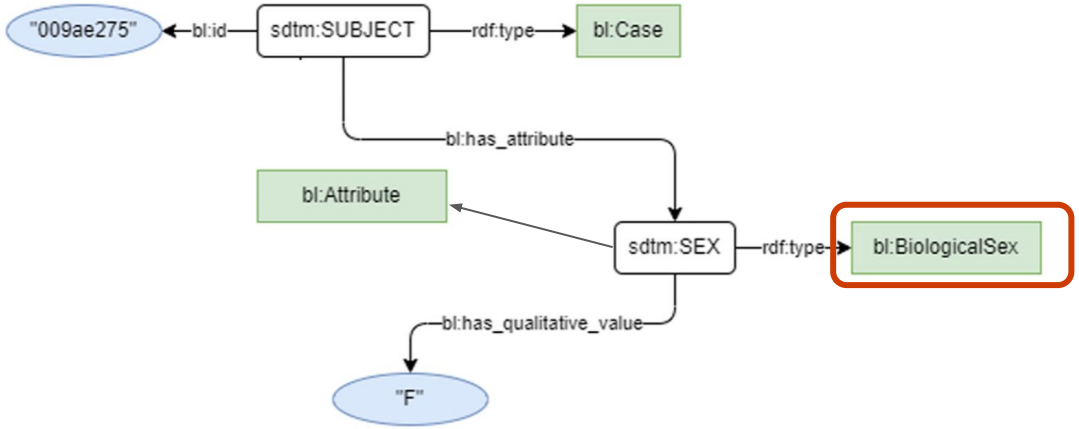
Comparison of the “FAIR” Models



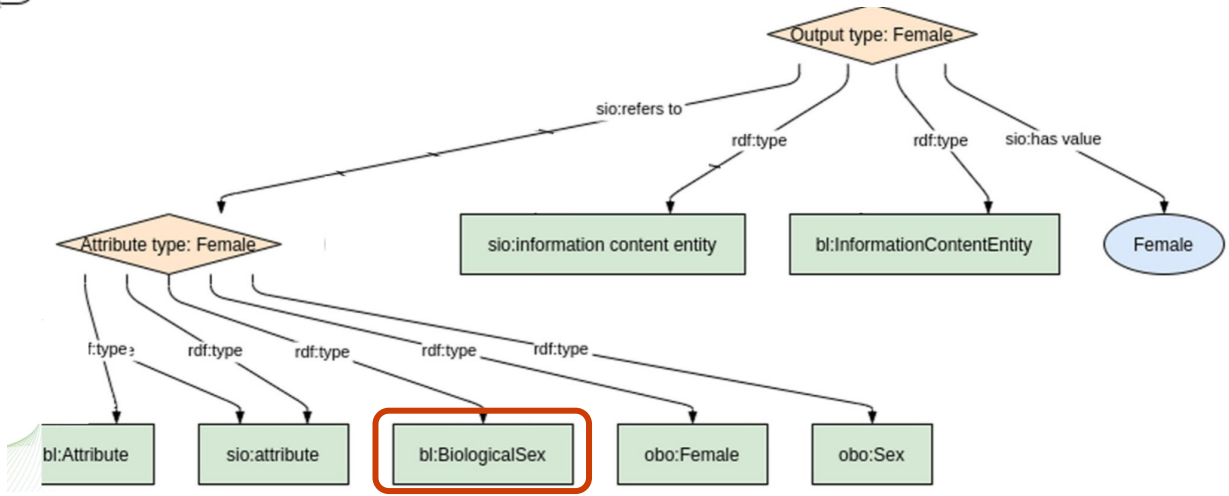
How do you approach integrating these two very different models?



Comparison of the “FAIR” Models



In this experiment, both participants CHANGED our data, to “anchor” our shared concepts in a third-party biomedical vocabulary (BioLink)



Did that work?

Summary:

We were both making semantic models of the same data type(s)

We used different vocabularies, and different models, but both would be reasonably considered “FAIR”

We mutually set “anchor” points throughout our data such that an agent/query could find the “semantic intersection-points” between the graphs (i.e. we gave the agent a bit of an advantage)

The Federated Query
(this is a query for Leukocyte count, not sex)

```
PREFIX ncit: <http://purl.obolibrary.org/obo/>
PREFIX obo: <http://purl.obolibrary.org/obo/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX sio: <http://semanticscience.org/resource/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX biolink: <https://w3id.org/biolink/vocab/>
PREFIX bl: <https://w3id.org/biolink/>

SELECT DISTINCT ?test ?value ?unit WHERE {
{SERVICE <http://fairdata.systems:8890/sparql>{
  {SELECT ?test ?value where {
     GRAPH <http://w3id.org/FAIR_Training_LDP/DAV/home/LDP/cpath/cpath_full> {
              ?test a biolink:Procedure, ncit:NCIT_C51948 .
              ?test ?has_output ?output .
              ?output a biolink:InformationContentEntity .
?output bl:has_attribute ?att .
?att bl:has_quantitative_value|bl:has_qualitative_value ?valnode .
?valnode bl:has_numeric_value ?value
    }
   }
  }
 }
}
UNION
{SERVICE <http://fairdata.systems:8890/sparql>{
  {SELECT ?test ?value ?unit where {
      GRAPH
<http://w3id.org/FAIR_Training_LDP/DAV/home/LDP/cpath/cbgp_leuk> {
              ?test a biolink:Procedure, ncit:NCIT_C51948 .
              ?test ?has_output ?output .
              ?output a biolink:InformationContentEntity .
?output sio:SIO_000300 ?value .
?output sio:SIO_000221 ?unitnode .
              ?unitnode rdfs:label ?unit
    }
    }
   }
  }
 }
}
```

The Federated Query (relevant bits)

```
        ?test a biolink:Procedure, ncit:NCIT_C51948 .
        ?test ?has_output ?output .
        ?output a biolink:InformationContentEntity .

?output bl:has_attribute ?att .
?att bl:has_quantitative_value|bl:has_qualitative_value ?valnode .
?valnode bl:has_numeric_value ?value
```

```
        ?test a biolink:Procedure, ncit:NCIT_C51948 .
        ?test ?has_output ?output .
        ?output a biolink:InformationContentEntity .

?output sio:SIO_000300 ?value .
?output sio:SIO_000221 ?unitnode .
?unitnode rdfs:label ?unit
```

The Federated Query (relevant bits)

Anchor Points used to set the “frame” of the query over the two models

```
?test a biolink:Procedure, ncit:NCIT_C51948 .
?test ?has_output ?output .
?output a biolink:InformationContentEntity .

?output bl:has_attribute ?att .
?att bl:has_quantitative_value|bl:has_qualitative_value ?valnode .
?valnode bl:has_numeric_value ?value
```

```
?test a biolink:Procedure, ncit:NCIT_C51948 .
?test ?has_output ?output .
?output a biolink:InformationContentEntity .

?output sio:SIO_000300 ?value .
?output sio:SIO_000221 ?unitnode .
?unitnode rdfs:label ?unit
```

The Federated Query (relevant bits)

The remainder of the query was, nevertheless, distinct

```
?test a biolink:Procedure, ncit:NCIT_C51948 .
?test ?has_output ?output .
?output a biolink:InformationContentEntity .

?output bl:has_attribute ?att .
?att bl:has_quantitative_value|bl:has_qualitative_value ?valnode .
?valnode bl:has_numeric_value ?value
```

```
?test a biolink:Procedure, ncit:NCIT_C51948 .
?test ?has_output ?output .
?output a biolink:InformationContentEntity .

?output sio:SIO_000300 ?value .
?output sio:SIO_000221 ?unitnode .
?unitnode rdfs:label ?unit
```

The Federated Query (relevant bits)

One team had the measurement modeled **as an attribute** of the output of the procedure

```
?test a biolink:Procedure, ncit:NCIT_C51948 .
?test ?has_output ?output .
?output a biolink:InformationContentEntity .

?output bl:has_attribute ?att .
?att bl:has_quantitative_value|bl:has_qualitative_value ?valnode .
?valnode bl:has_numeric_value ?value
```

The other team modeled the measurement **as the value** of the output

```
?test a biolink:Procedure, ncit:NCIT_C51948 .
?test ?has_output ?output .
?output a biolink:InformationContentEntity .

?output sio:SIO_000300 ?value .      # SIO_000300 → "has value"
?output sio:SIO_000221 ?unitnode .
?unitnode rdfs:label ?unit
```

The Federated Query (relevant bits)

And things like measurement units did not exist in one of the models

```
?test a biolink:Procedure, ncit:NCIT_C51948 .
?test ?has_output ?output .
?output a biolink:InformationContentEntity .

?output bl:has_attribute ?att .
?att bl:has_quantitative_value|bl:has_qualitative_value ?valnode .
?valnode bl:has_numeric_value ?value
```

```
?test a biolink:Procedure, ncit:NCIT_C51948 .
?test ?has_output ?output .
?output a biolink:InformationContentEntity .

?output sio:SIO_000300 ?value .
?output sio:SIO_000221 ?unitnode .
?unitnode rdfs:label ?unit
```

We tried!

We tried many ways of solving the problem:

Pre-configured blocks of query fragments that could be dynamically assembled

Query clauses that were “OPTIONAL” between one model and another

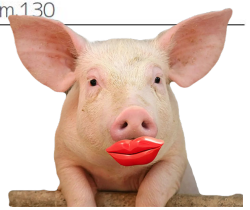
Additional semantic anchors

(abortive) attempt to bring an ontological mapping service into the query as a third participant

This paper compares two independent FAIRification efforts (C-PATH and CARE-SM) over identical data items

Journal of the Society for Clinical Data Management

Alarcón-Moreno P, et al. Leveraging Biolink as a FAIR "Rosetta Stone" Between Clinical Semantic Models Provides Emergent Interoperability. *Journal of the Society for Clinical Data Management*. 2022; 2(3): 2, pp. 1–8. DOI: https://doi.org/10.47912/jscdm.130



ORIGINAL RESEARCH

Leveraging Biolink as a FAIR "Rosetta Stone" Between Clinical Semantic Models Provides Emergent Interoperability

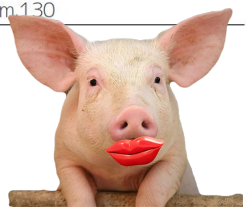
Pablo Alarcón-Moreno*, Ian Braun†, Emily Hartley†, Daniel Olson†, Nirupama Benis‡, Ronald Cornet‡, Mark D. Wilkinson* and Ramona L. Walls†

Interoperability was **almost zero!**

This paper compares two independent FAIRification efforts (C-PATH and CARE-SM) over identical data items

Journal of the Society for Clinical Data Management

Alarcón-Moreno P, et al. Leveraging Biolink as a FAIR "Rosetta Stone" Between Clinical Semantic Models Provides Emergent Interoperability. *Journal of the Society for Clinical Data Management*. 2022; 2(3): 2, pp. 1–8. DOI: https://doi.org/10.47912/jscdm.130



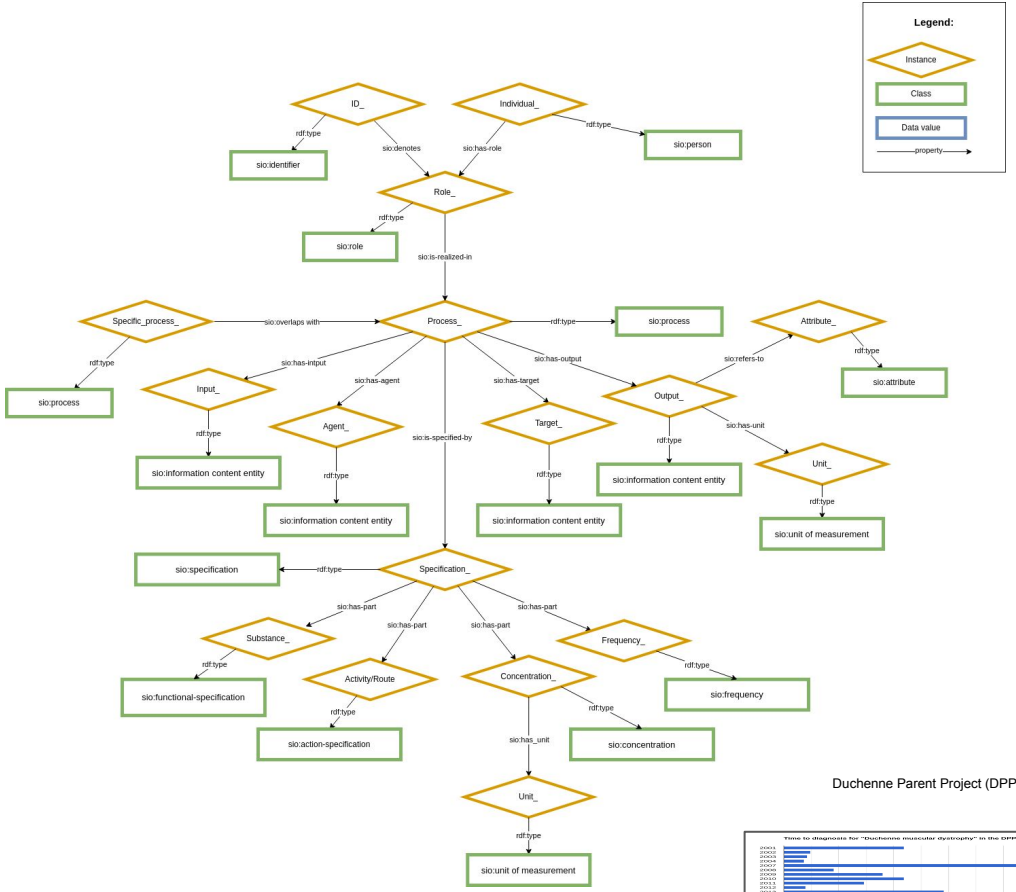
ORIGINAL RESEARCH

Leveraging Biolink as a FAIR "Rosetta Stone" Between Clinical Semantic Models Provides Emergent Interoperability

Pablo Alarcón-Moreno*, Ian Braun†, Emily Hartley†, Daniel Olson†, Nirupama Benis‡, Ronald Cornet‡, Mark D. Wilkinson* and Ramona L. Walls†

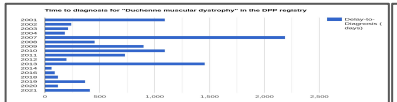
And this was even after ***we cheated*** by manually modifying the starting data by adding "anchors"!

A shared data model is <u>required</u> for data-level reusability*

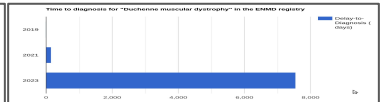


CARE-SM
Clinical And Registry Entries Semantic Model

Duchenne Parent Project (DPP)



EURO-NMD (mock data)



* Note Principle R1.3

Latest Initiative

Duplicate and expand the Rare Disease success story in the domain of **Seed (Germplasm) Banking**

Santiago Moreno-Vázquez
Universidad Politécnica de Madrid
- Biodiversidad y conservación de recursos fitogenéticos
- Professor BIOTECNOLOGÍA - BIOLOGÍA VEGETAL
- Lead of the Banco de Germoplasma Vegetal de la UPM "César Gómez Campo"

TED2021-130788B-I00

FLAIR-GG

FAIRification, Linking And Integrated Reuse of Global *ex situ* plant Germplasm resources

Dr. Santiago Moreno Vázquez
Dr. Mark D. Wilkinson
Oussama Mohammed Benhamed, PhD Candidate
Alberto Cámara Ballesteros, PhD Candidate

El Banco de Germoplasma Vegetal UPM "César Gómez-Campo"

One of the most complete collections of wild crucifers in the world (1,027 taxa with 4,863 accessions); Seeds of Iberian and Macaronesian endemic species; **currently preserves 24% of the threatened flora in Spain**

FLAIR-GG Objectives

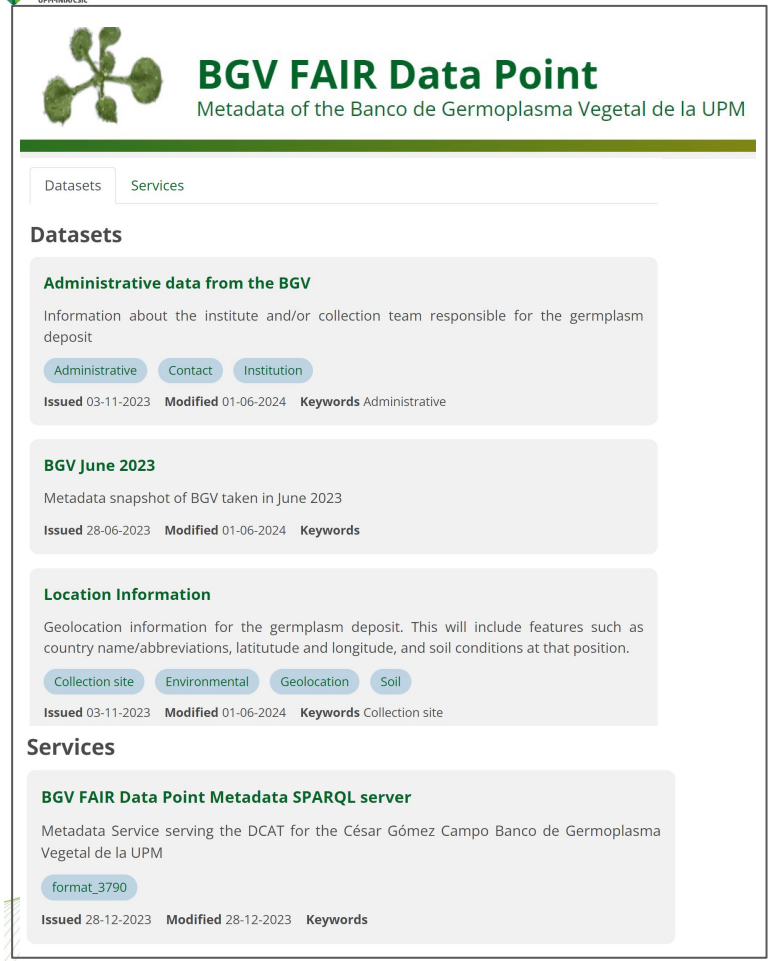
EUROPEAN JOINT PROGRAMME RARE DISEASES

BGV-UPM
"César Gómez Campo"

Replicate EJP-RDs success with
BGV as our first target

…then expand!

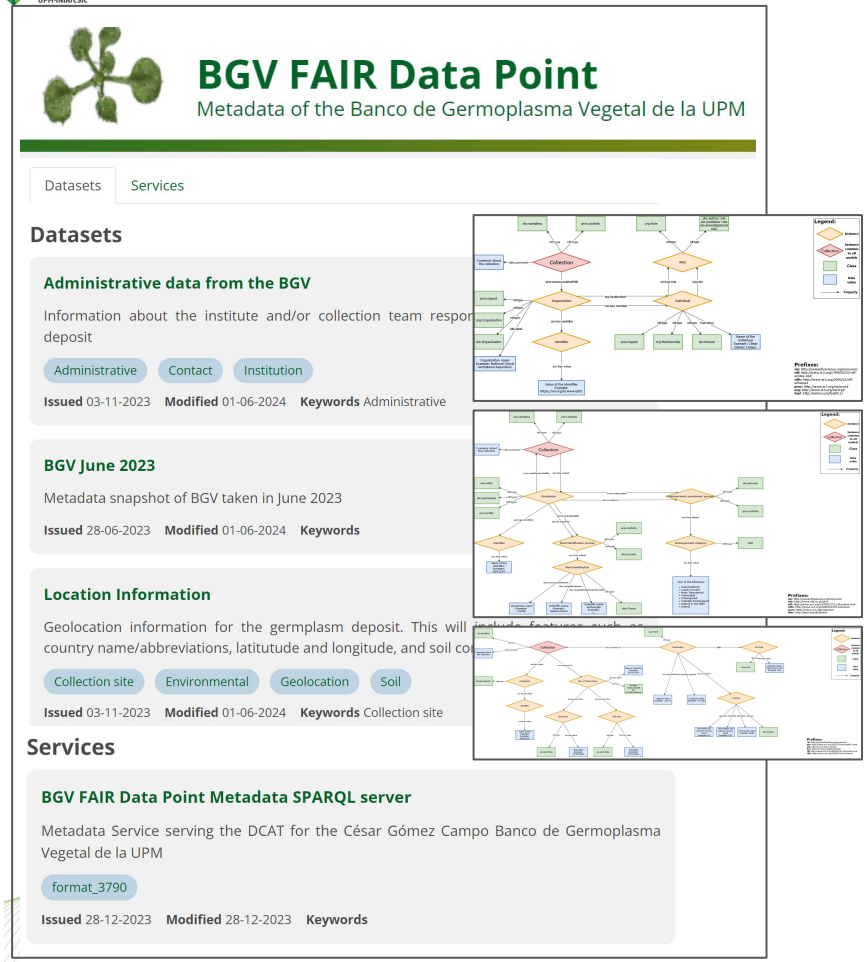
FLAIR-GG Status



FAIR Data Point:

- Customized to the Germplasm case
- Automated installer available
- Excel metadata capture templates
- We offer to host the FDP for any new network partner to reduce cost-of-entry

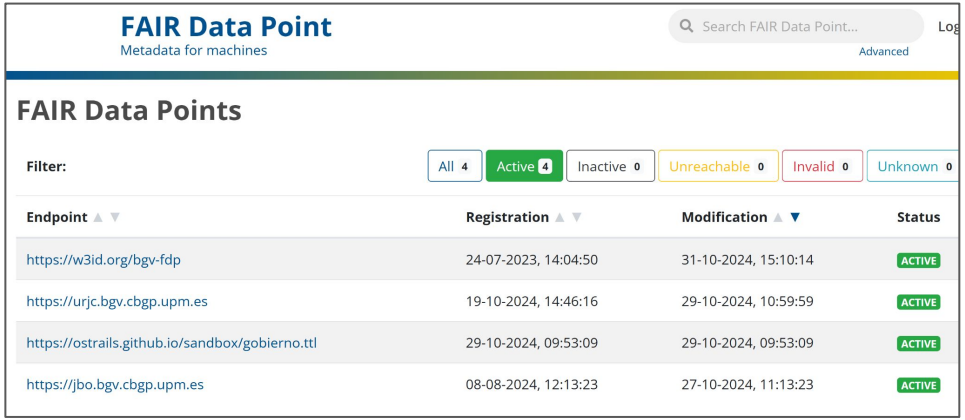
FLAIR-GG Status



Models and Templates

- Shared data models have been created
- CSV Templates are completed
- Transformation pipeline tests confirm success
- FLAIR-GG infrastructure can be replicated by any partner within minutes.

FLAIR-GG Status



FAIR Data Point Federated Partner Index:

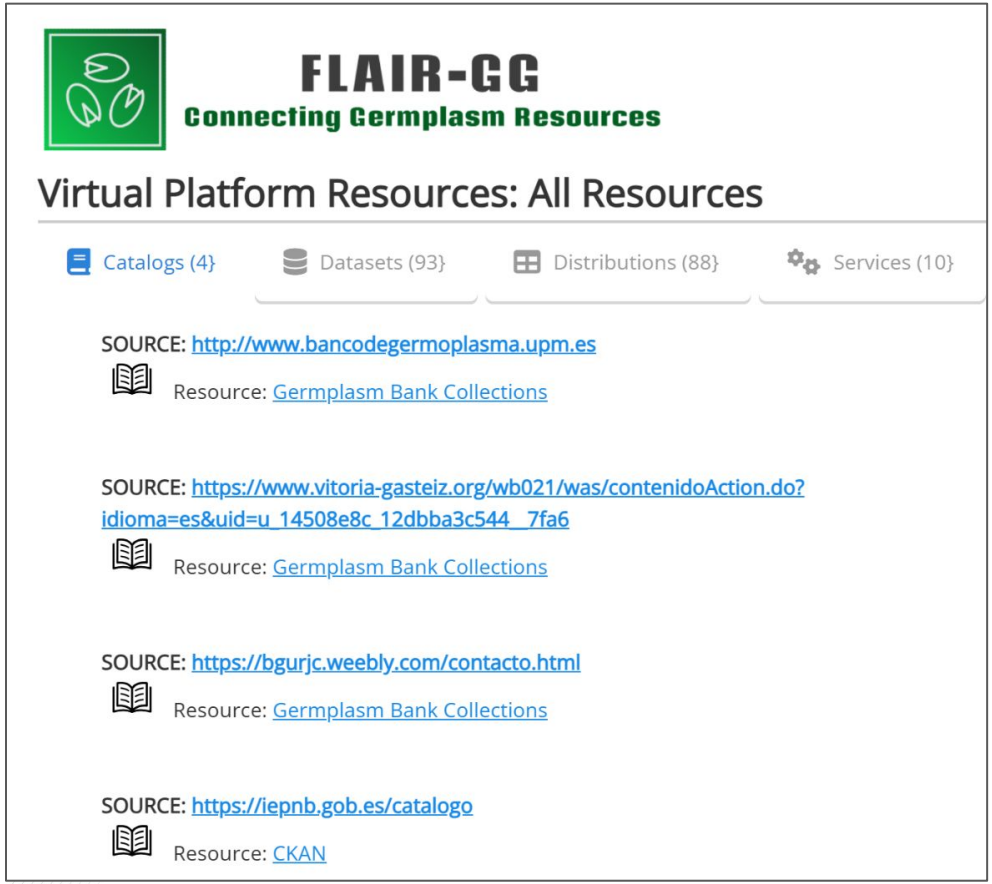
- So far, only 4 of us
- Several other national and international germplasm bank partners identified who should be onboard within a few weeks

Advances over EJP-RD

The FLAIR-GG Virtual Platform

The FLAIR-GG FDP Proxy

FLAIR-GG “Virtual Platform” (VP)



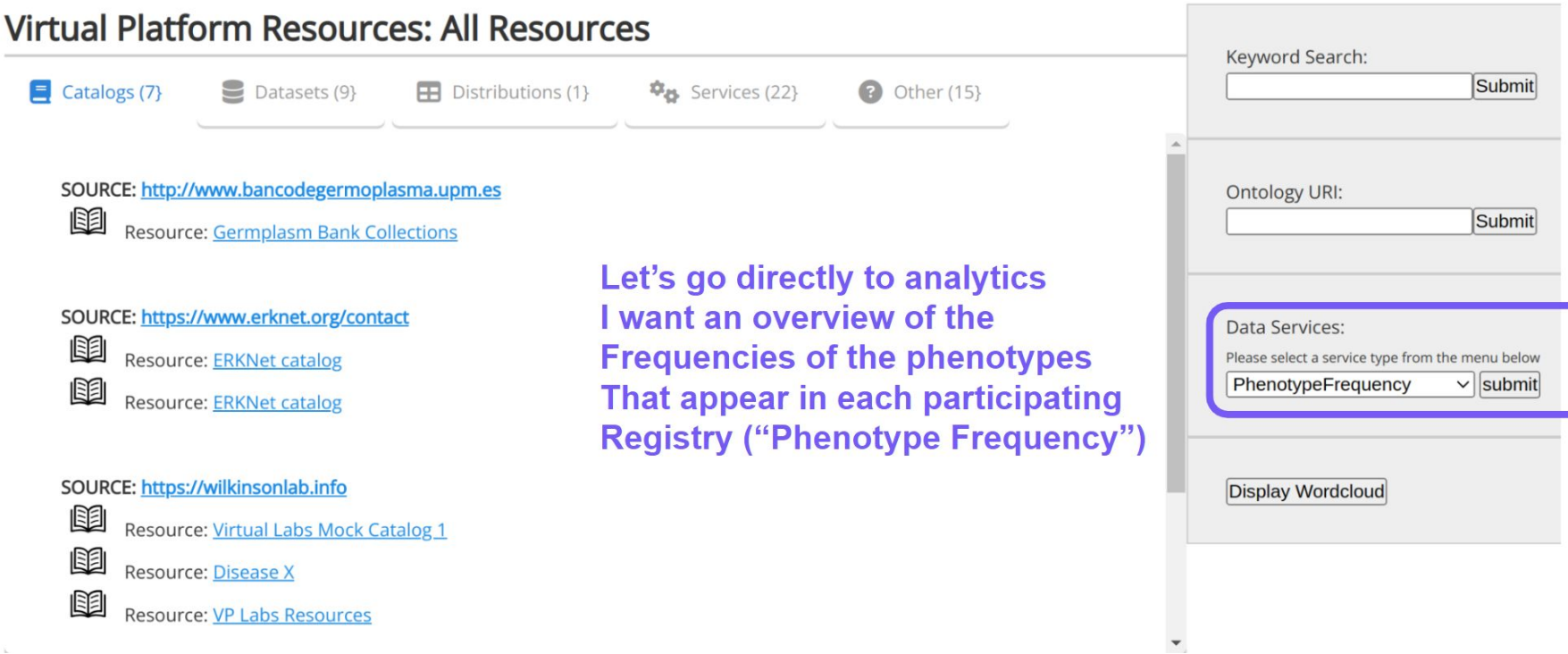
- Entrypoint for federated exploration of the partner network
- Drives traffic to partner websites
- One-click launching of **question-specific analytics environments** such as map-integration (possible because of FAIR annotations of data services)

Screenshots of the FLAIR-GG Virtual Platform Running over the EJP-RD FAIR Network

We do not yet have enough participants with data services for me to show you a demo of the VP doing federated analytics over FAIR Data

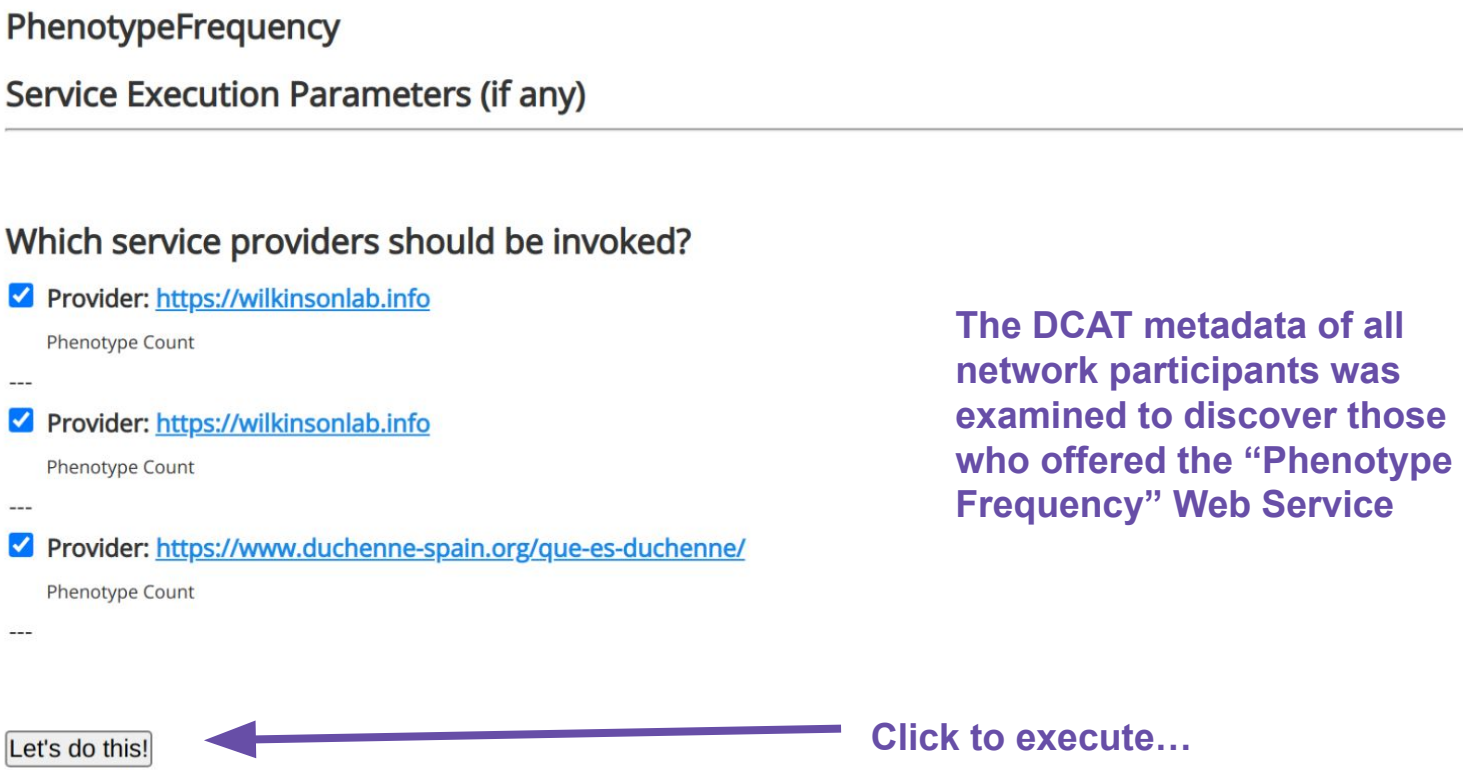
FLAIR-GG VP utilizes the same standard (DCAT) that is used by the Rare Disease network

therefore I will point my VP to look at that network to do this demonstration



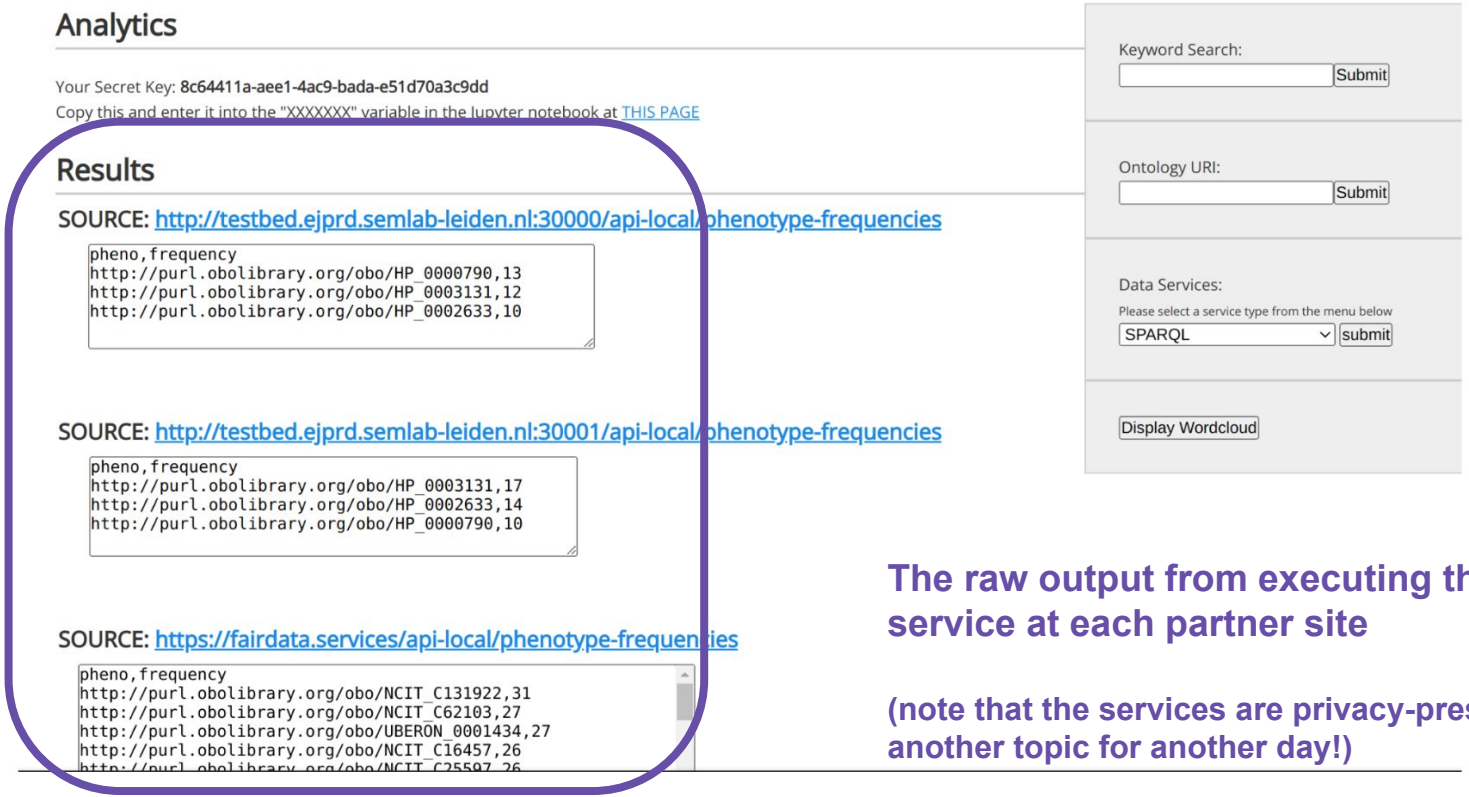
Let's go directly to analytics
I want an overview of the Frequencies of the phenotypes That appear in each participating Registry ("Phenotype Frequency")

* this is DCAT DataService record metadata



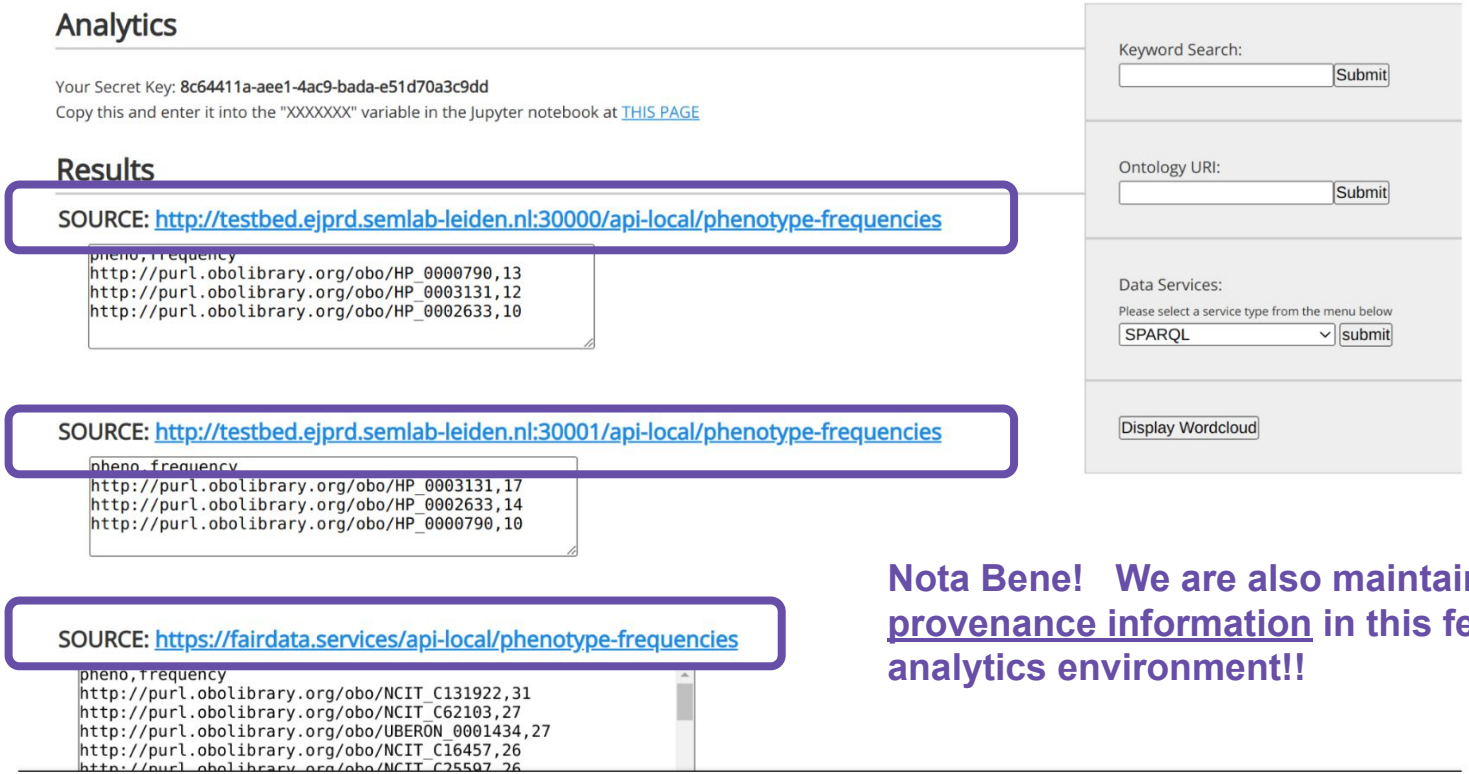
The DCAT metadata of all network participants was examined to discover those who offered the "Phenotype Frequency" Web Service

Click to execute…



The raw output from executing the service at each partner site

(note that the services are privacy-preserving… another topic for another day!)



Nota Bene! We are also maintaining provenance information in this federated analytics environment!!

What's this?

Analytics

Your Secret Key: **8c64411a-aee1-4ac9-bada-e51d70a3c9dd**
Copy this and enter it into the "XXXXXXX" variable in the Jupyter notebook at THIS PAGE

Results

SOURCE: http://testbed.ejprd.semlab-leiden.nl:30000/api-local/phenotype-frequencies

```
pheno,frequency
http://purl.obolibrary.org/obo/HP_0000790,13
http://purl.obolibrary.org/obo/HP_0003131,12
http://purl.obolibrary.org/obo/HP_0002633,10
```

SOURCE: http://testbed.ejprd.semlab-leiden.nl:30001/api-local/phenotype-frequencies

```
pheno,frequency
http://purl.obolibrary.org/obo/HP_0003131,17
http://purl.obolibrary.org/obo/HP_0002633,14
http://purl.obolibrary.org/obo/HP_0000790,10
```

SOURCE: https://fairdata.services/api-local/phenotype-frequencies

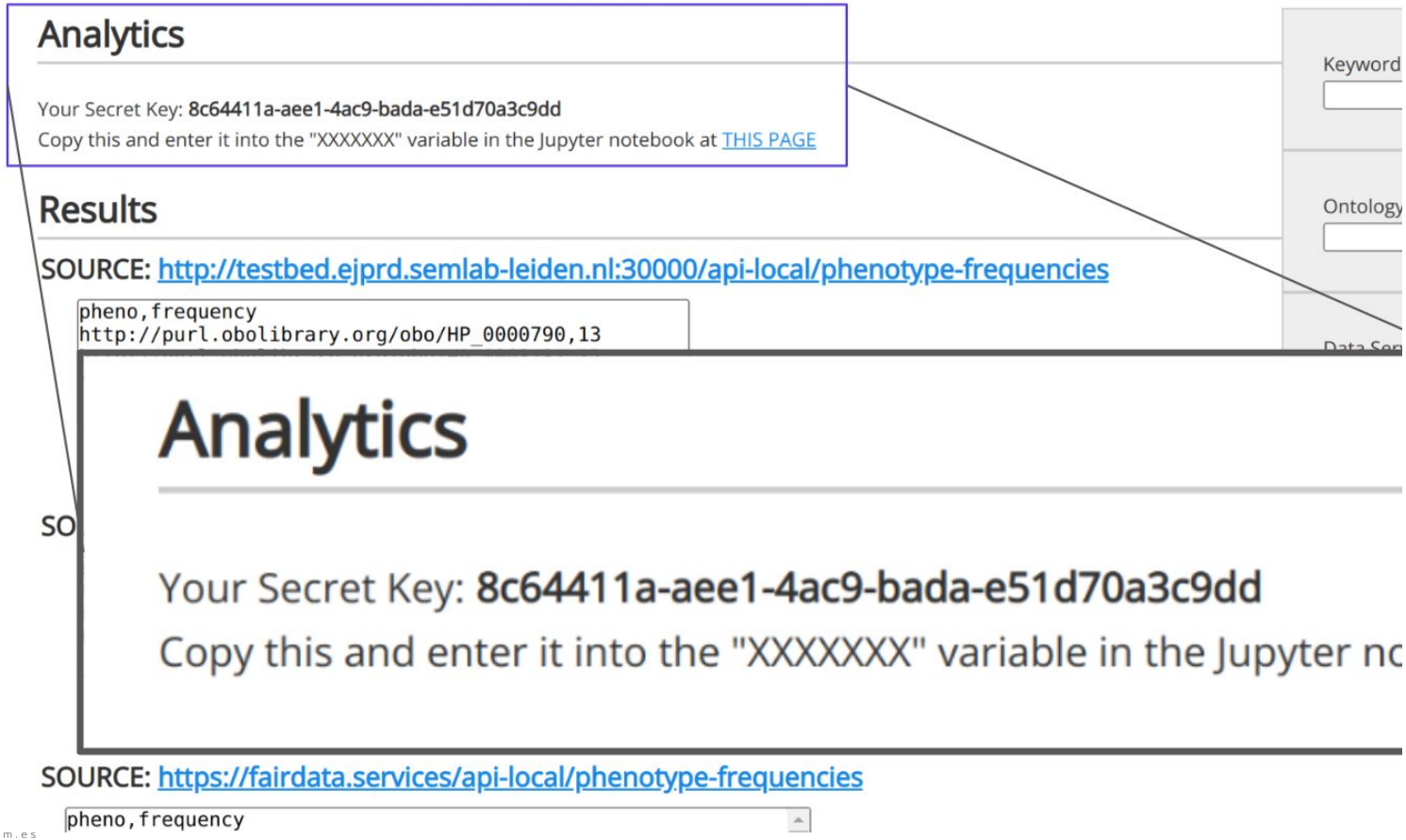
```
pheno,frequency
http://purl.obolibrary.org/obo/NCIT_C131922,31
http://purl.obolibrary.org/obo/NCIT_C62103,27
http://purl.obolibrary.org/obo/UBERON_0001434,27
http://purl.obolibrary.org/obo/NCIT_C16457,26
http://purl.obolibrary.org/obo/NCIT_C25597,26
```

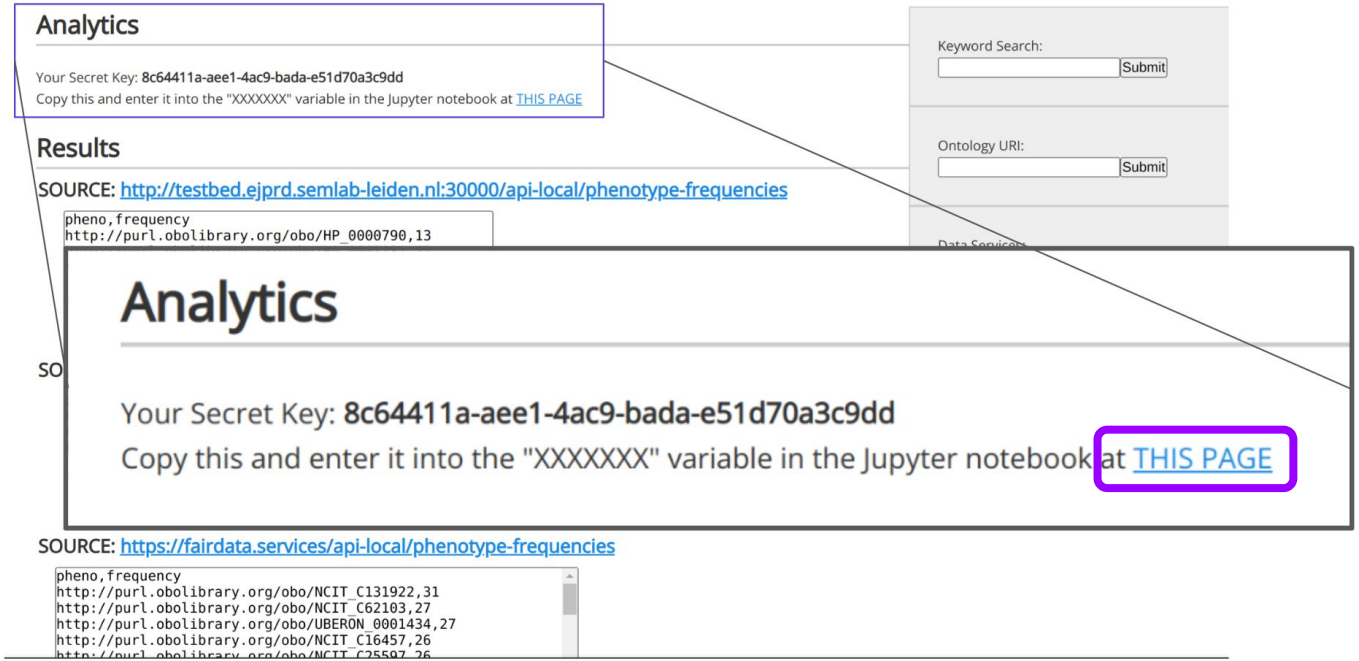
Keyword Search: Submit

Ontology URI: Submit

Data Services:
Please select a service type from the menu below
SPARQL submit

Display Wordcloud


Analytics
Your Secret Key: 8c64411a-aee1-4ac9-bada-e51d70a3c9dd
Copy this and enter it into the "XXXXXXX" variable in the Jupyter notebook at THIS PAGE
Results
SOURCE: http://testbed.ejprd.semlab-leiden.nl:30000/api-local/phenotype-frequencies
pheno,frequency
http://purl.obolibrary.org/obo/HP_0000790,13
Keyword
Ontology
Analytics
Your Secret Key: 8c64411a-aee1-4ac9-bada-e51d70a3c9dd
Copy this and enter it into the "XXXXXXX" variable in the Jupyter nc
SOURCE: https://fairdata.services/api-local/phenotype-frequencies
pheno,frequency


Analytics
Your Secret Key: 8c64411a-aee1-4ac9-bada-e51d70a3c9dd
Copy this and enter it into the "XXXXXXX" variable in the Jupyter notebook at THIS PAGE
Results
SOURCE: http://testbed.ejprd.semlab-leiden.nl:30000/api-local/phenotype-frequencies
pheno,frequency
http://purl.obolibrary.org/obo/HP_0000790,13
Keyword Search:
Submit
Ontology URI:
Submit
Analytics
Your Secret Key: 8c64411a-aee1-4ac9-bada-e51d70a3c9dd
Copy this and enter it into the "XXXXXXX" variable in the Jupyter notebook at THIS PAGE
SOURCE: https://fairdata.services/api-local/phenotype-frequencies
pheno,frequency
http://purl.obolibrary.org/obo/NCIT_C131922,31
http://purl.obolibrary.org/obo/NCIT_C62103,27
http://purl.obolibrary.org/obo/UBERON_0001434,27
http://purl.obolibrary.org/obo/NCIT_C16457,26

Welcome to the EURO-NMD Hub Phenotype Frequency Analytics Notebook

The Phenotype Frequency Data Service will count the number of times a phenotype appears in a registry (on a per-patient basis - only counts once per patient)

Please run the first cell to set-up the analytics environment

In the second cell, we have pre-filled a basic analytics, to show you how to access and manipulate the data that was retrieved from the FLAIR-GG Virtual Platform.

Fill in the "key = 'XXXXXXX' with the secret key for your federated exploration output, and then... go!

]:
```
import pyodide_kernel

%pip install altair
%pip install pandas
%pip install requests

import altair as alt
import pandas as pd
import pyodide_http
import ssl
import json
import requests
import urllib3
import urllib.parse
urllib3.disable_warnings(urllib3.exceptions.InsecureRequestWarning)
print("DONE!  Proceed to next cell and replace the XXXXX with your secret key")
```

Pre-configured JupyterLite Notebook

]:
```
import io
import array
import sys

key = "8c64411a-aee1-4ac9-bada-e51d70a3c9dd"   # your secret key from the EURO-NMD Hub Virtual Platform
```

Welcome to the EURO-NMD Hub Phenotype Frequency Analytics Notebook

The Phenotype Frequency Data Service will count the number of times a phenotype appears in a registry (on a per-patient basis - only counts once per patient)

Please run the first cell to set-up the analytics environment

In the second cell, we have pre-filled a basic analytics, to show you how to access and manipulate the data that was retrieved from the FLAIR-GG Virtual Platform.

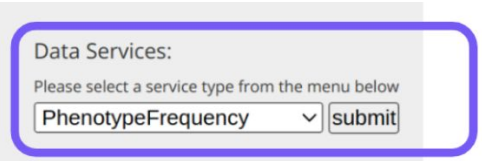
Fill in the "key = 'XXXXXXX' with the secret key for your federated exploration output, and then... go!

]:
```
import pyodide_kernel

%pip install altair
%pip install pandas
%pip install requests

import altair as alt
import pandas as pd
import pyodide_http
import ssl
import json
import requests
import urllib3
import urllib.parse
urllib3.disable_warnings(urllib3.exceptions.InsecureRequestWarning)
print("DONE!  Proceed to next cell and replace the XXXXX with your secret key")
```

Correct notebook selected based on this metadata element from the participant's DCAT record



]:
```
import io
import array
import sys

key = "8c64411a-aee1-4ac9-bada-e51d70a3c9dd"   # your secret key from the EURO-NMD Hub Virtual Platform
```

Welcome to the EURO-NMD Hub Phenotype Frequency Analytics Notebook

The Phenotype Frequency Data Service will count the number of times a phenotype appears in a registry (on a per-patient basis - only counts once per patient)

Please run the first cell to set-up the analytics environment

In the second cell, we have pre-filled a basic analytics, to show you how to access and manipulate the data that was retrieved from the FLAIR-GG Virtual Platform.

Fill in the "key = 'XXXXXXX' with the secret key for your federated exploration output, and then... go!

]:
```
import pyodide_kernel

%pip install altair
%pip install pandas
%pip install requests

import altair as alt
import pandas as pd
import pyodide_http
import ssl
import json
import requests
import urllib3
import urllib.parse
urllib3.disable_warnings(urllib3.exceptions.InsecureRequestWarning)
print("DONE!  Proceed to next cell and replace the XXXXX with your secret key")
```

]:
```
import io
import array
import sys

key = "8c64411a-aee1-4ac9-bada-e51d70a3c9dd"   # your secret key from the EURO-NMD Hub Virtual Platform
```

Copy/paste your key here

phenotypefrequency.ipynb

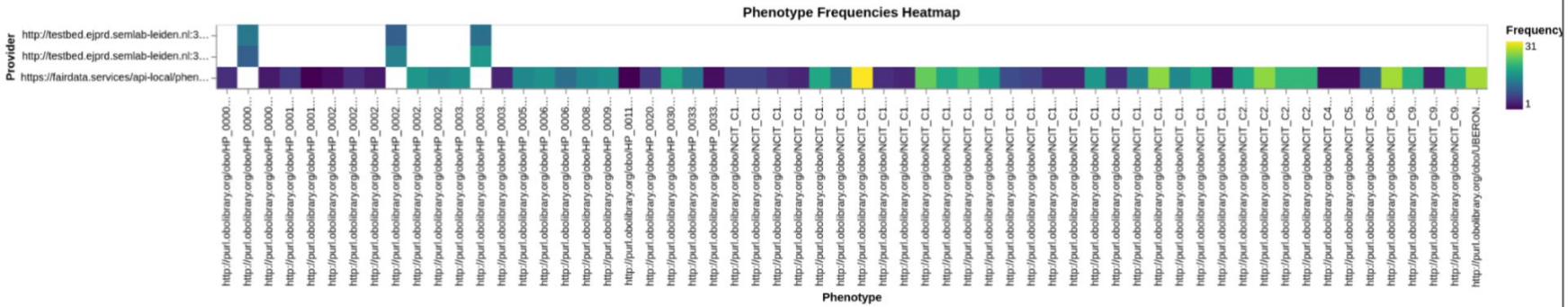
Code

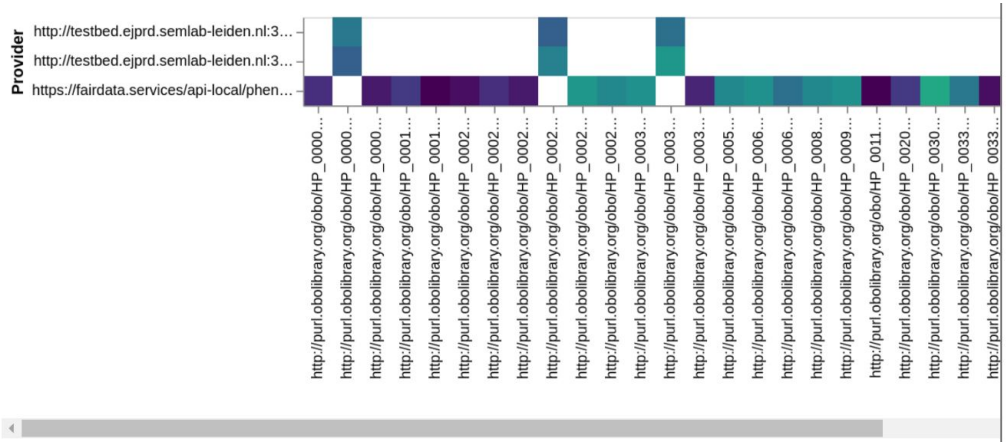
Notebook

```
[7]:
# Heatmap
heatmap = alt.Chart(df).mark_rect().encode(
    x='Phenotype:N',
    y='Provider:N',
    color=alt.Color('Frequency:Q', scale=alt.Scale(scheme='viridis')),
    tooltip=['Provider', 'Phenotype', 'Frequency']
).properties(
    title='Phenotype Frequencies Heatmap'
)

heatmap
```

[7]:






```
[ ]: # Put

     # Your

     # Own

     # Code

     # Here!
```

Of course, your data scientists can do whatever analysis they wish using the data that is pre-extracted for them into the most popular analytics environment (Python Pandas)

FLAIR-GG Next Steps

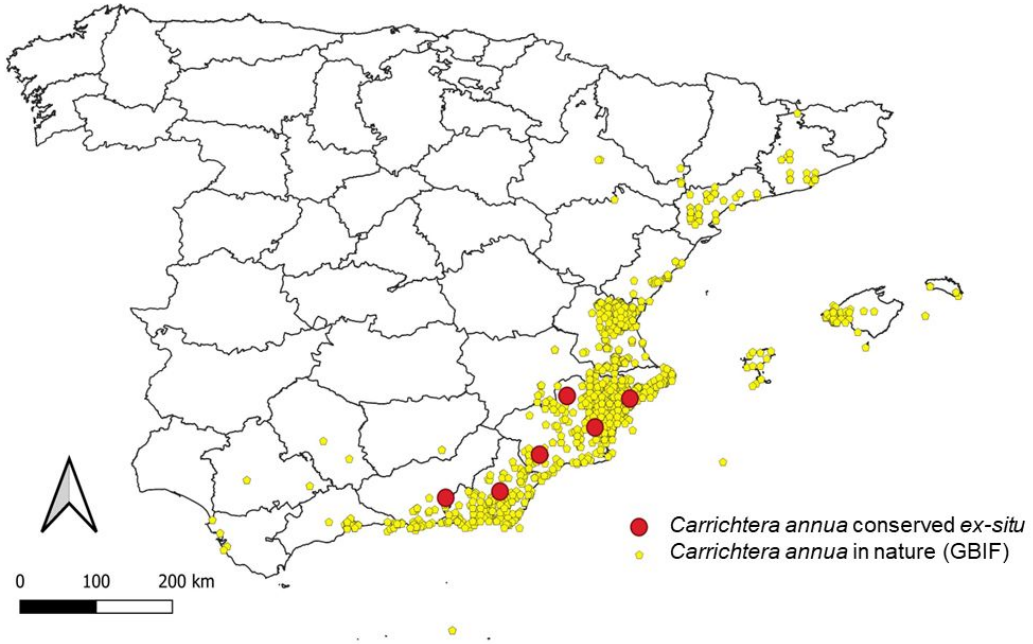
Begin constructing **shared queries** to help build data-driven conservation strategies

FLAIR-GG Next Steps

Dynamic integration of partner seedbank collection records with **GBIF** species observations

Conservation Strategy - Breadth:

"What geographic locations have not been sampled by any collection expeditions from our network partners?"

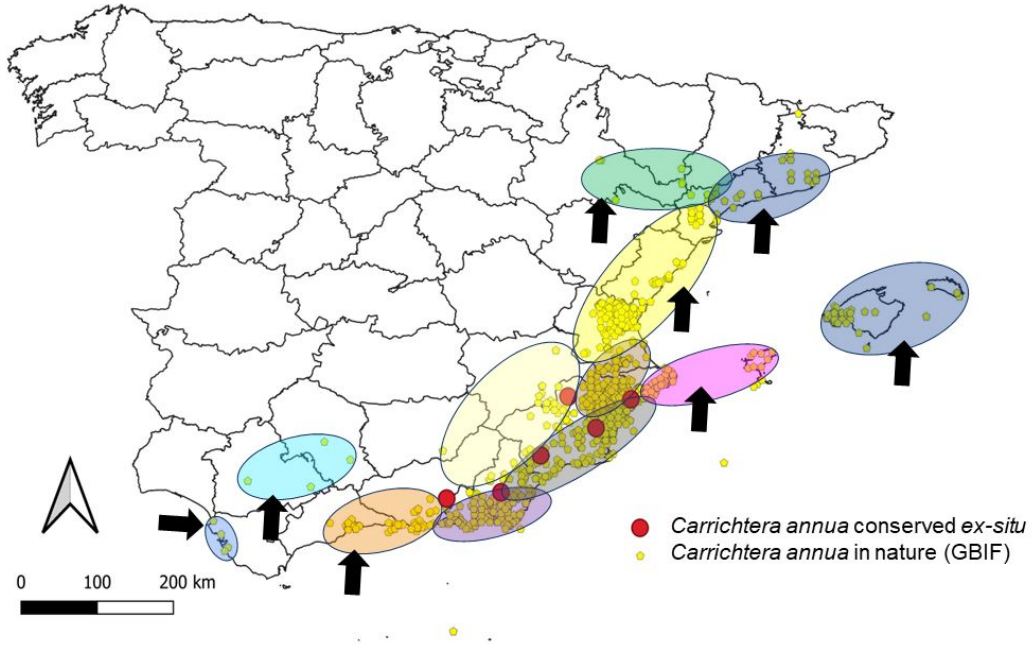


FLAIR-GG Next Steps

Environment drives intra-species diversity - add resources such as AEMET (Spanish Meteorological Agency), IGME (Geological and Mining Institute), CNIG (Spanish Geographical Agency) to capture environmental information associated to territories where the species lives.

Conservation Strategy - Depth:

"Are there occurrence locations of species X that are within soil types or microclimates for which we lack samples in our germplasm banks?"



Which leads me to the final advance

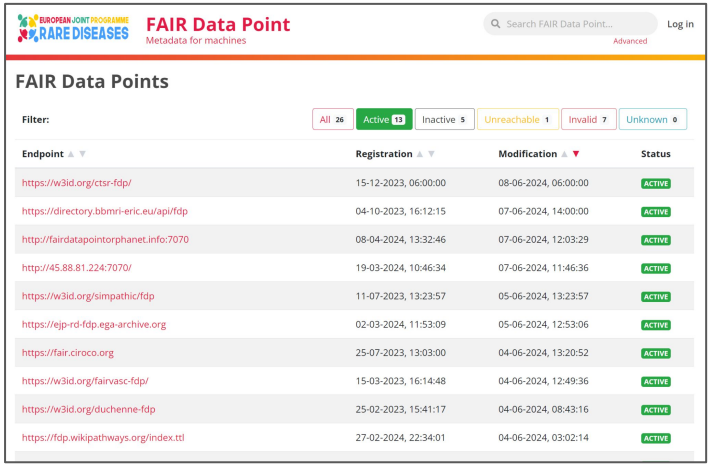
Bringing third-party data (e.g. from the government) into the VP via a “Proxy”

FLAIR-GG FDP Proxy

"Vanilla" DCAT Record

FDP Proxy

FDP Compliant DCAT Record



FLAIR-GG “Virtual Platform” (VP)

FLAIR-GG
Connecting Germplasm Resources

Virtual Platform Resources: All Resources

Catalogs (4) | Datasets (93) | Distributions (88) | Services (10)

SOURCE: http://www.bancodegermoplasma.upm.es
Resource: Germplasm Bank Collections

SOURCE: https://www.vitoria-gasteiz.org/wb021/was/contenidoAction.do?idioma=es&uid=u_14508e8c_12dbba3c544__7fa6
Resource: Germplasm Bank Collections

SOURCE: https://bgurjc.weebly.com/contacto.html
Resource: Germplasm Bank Collections

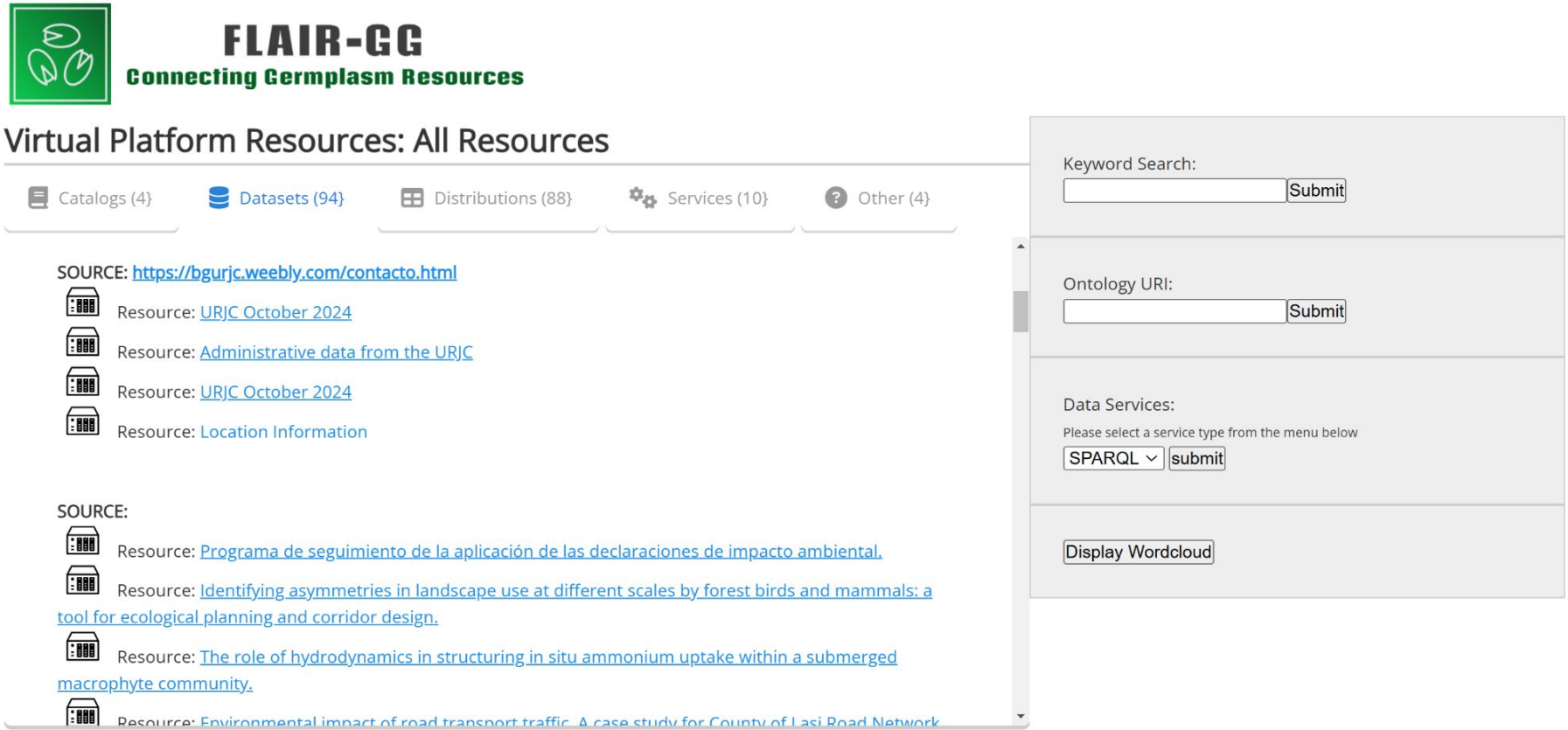
SOURCE: https://iepnb.gob.es/catalogo
Resource: CKAN

GOBIERNO DE ESPAÑA
MINISTERIO PARA LA TRANSICIÓN ECOLÓGICA Y EL RETO DEMOGRÁFICO

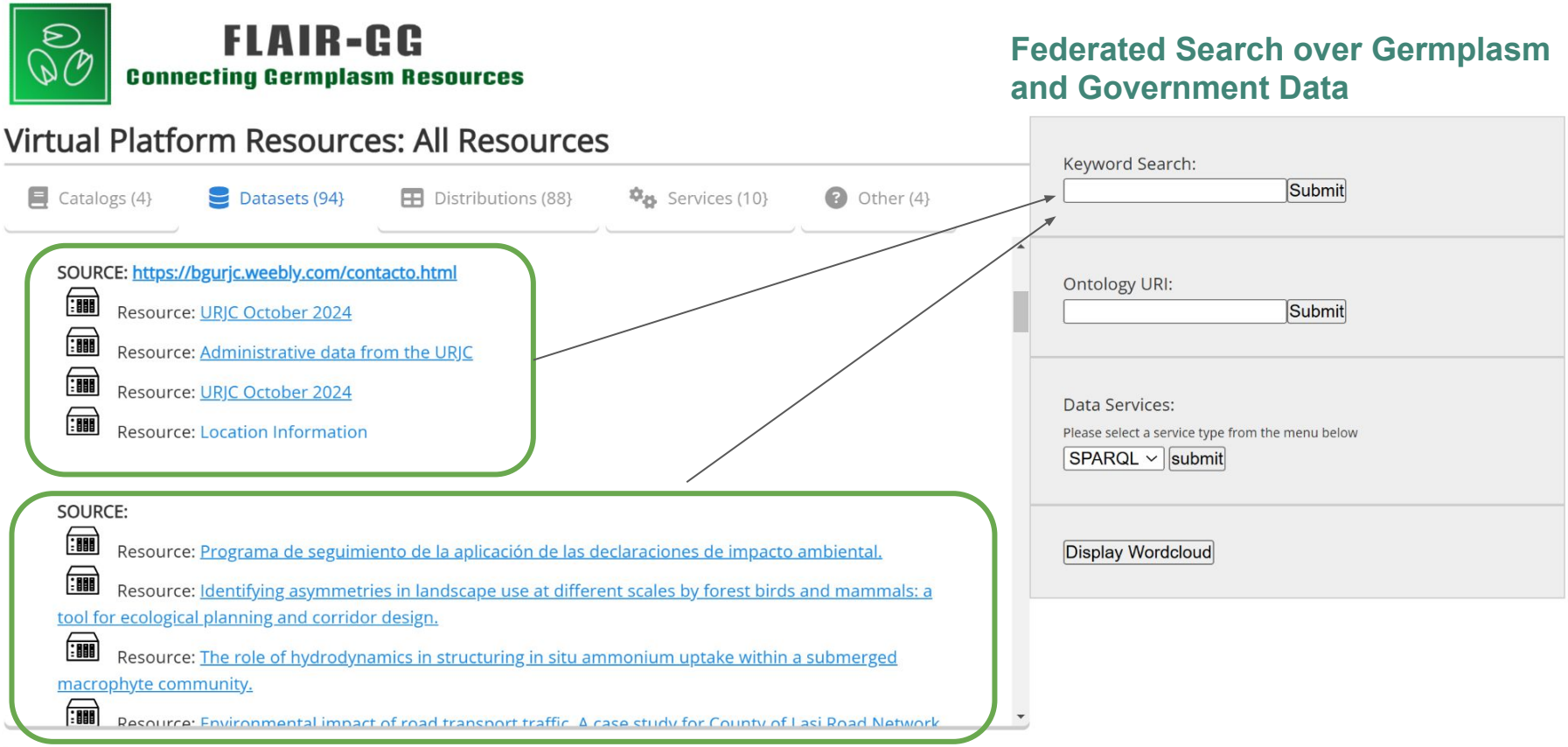
iepnb Inventario Español de Patrimonio Natural y de la Biodiversidad

“Proxy” the government data into the VP

FLAIR-GG “Virtual Platform” (VP)



FLAIR-GG “Virtual Platform” (VP)



FLAIR-GG “Virtual Platform” (VP)

Federated Analytics over Germplasm and Government Data (coming soon!)

Virtual Platform Resources: All Resources

Catalogs (4) | Datasets (94) | Distributions (88) | Services (10) | Other (4)

SOURCE: https://bgurjc.weebly.com/contacto.html

- Resource: URJC October 2024
- Resource: Administrative data from the URJC
- Resource: URJC October 2024
- Resource: Location Information

SOURCE:

- Resource: Programa de seguimiento de la aplicación de las declaraciones de impacto ambiental.
- Resource: Identifying asymmetries in landscape use at different scales by forest birds and mammals: a tool for ecological planning and corridor design.
- Resource: The role of hydrodynamics in structuring in situ ammonium uptake within a submerged macrophyte community.
- Resource: Environmental impact of road transport traffic. A case study for County of Lasi Road Network

Keyword Search: [ ] Submit

Ontology URI: [ ] Submit

Data Services:

Please select a service type from the menu below

SPARQL submit

Display Wordcloud

Take Home Messages

Repeated examples of “interoperability” @ both metadata and data layers shows
FAIR is necessary, but not sufficient!

FAIR is metadata first! Almost all useful behavior derived from rich metadata

FAIRness is measurable, but we need global governance of testing
before anyone can trust FAIR quality assessment;
we are almost never able to measure the FAIRness of data

Technologies/strategies allow FAIR experts to assist domain-expert data owners
in creating FAIR (meta)data themselves

Within a defined community, a high degree of interoperability can be achieved
through harmonization of semantics and models
(which, of course, is unsurprising!)

Acknowledgements

https://tinyurl.com/WEBIST-FAIR

Wilkinson Lab:

- *Alberto Cámara Ballesteros*
- *Dr. Oussama Mohammed Benhamed*
- *Elena Aguayo Jara*
- *Dr. Pablo Alarcón*

ETSIAAB:

- *Dr. Elena Torres*
- *German Pastor*
- *Evrykleia Sofia Verykaki*

CBGP CENTRO DE BIOTECNOLOGÍA Y GENÓMICA DE PLANTAS
EXCELENCIA SEVERO OCHOA 2022 - 2025 2017 - 2021
POLITÉCNICA
INIA Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria
CSIC CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS
ETSIAAB ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA ALIMENTARIA Y DE BIOSISTEMAS
BGV-UPM "César Gómez Campo"
eosc
EUROPEAN JOINT PROGRAMME RARE DISEASES
European Union's Horizon 2020 grant agreement N°825575
FAIR DATA SYSTEMS
TED2021-130788B-I00
EURO-NMD Registry Hub
European Reference Network for neuromuscular diseases
MINISTERIO DE CIENCIA E INNOVACIÓN
Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
AGENCIA ESTATAL DE INVESTIGACIÓN
ostrails
European Union's Horizon Europe fgrant agreement No. 101130187
WORLD DUCHENNE ORGANIZATION UPPMD