Applying FAIR Principles in Practice: From Semantic Models to Data Stewardship

Alberto Cámara

we expect that the audience of the public part of the meeting is going to consist mostly of the local UAM students and early career specialists from nuclear physics, so they don't know much about FAIR data management.

it could be good to talk about FAIR Data Principles (I'm sure all speakers before yo will use FAIR as a buzzword, but it would make sense to get more concrete with it), go over tenchologies/layers of semantic web architecture and then mention European level FAIR data initiatves, such as OSTrails or EOSC, where Wilkinson Lab is/was involved. Plus, you can share your personal experiences in EJP-RD/ERNs

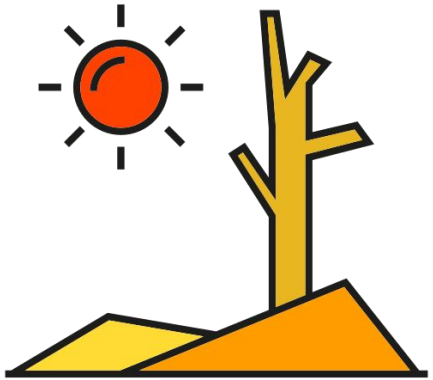
The aims are:

- Making sure audience knows about FAIRfication more than they knew before
- Give the ones, who would like to find some collaborators on European level (or in Spain?) idea, there to look

Biological question

How well represented is the *in situ* intra-specific diversity in Spanish germplasm banks?

- Different environments
- Different protected areas
- Different meteorological/climatological variables
- Co-occurring populations



Demo: germplasm bank intra-species representativeness

Species: ***Ilex aquifolium*** **L.**

- Legally protected species in Spain
- Collected from Islas Baleares, Castilla y León, Comunidad de Madrid, Comunidad Valenciana

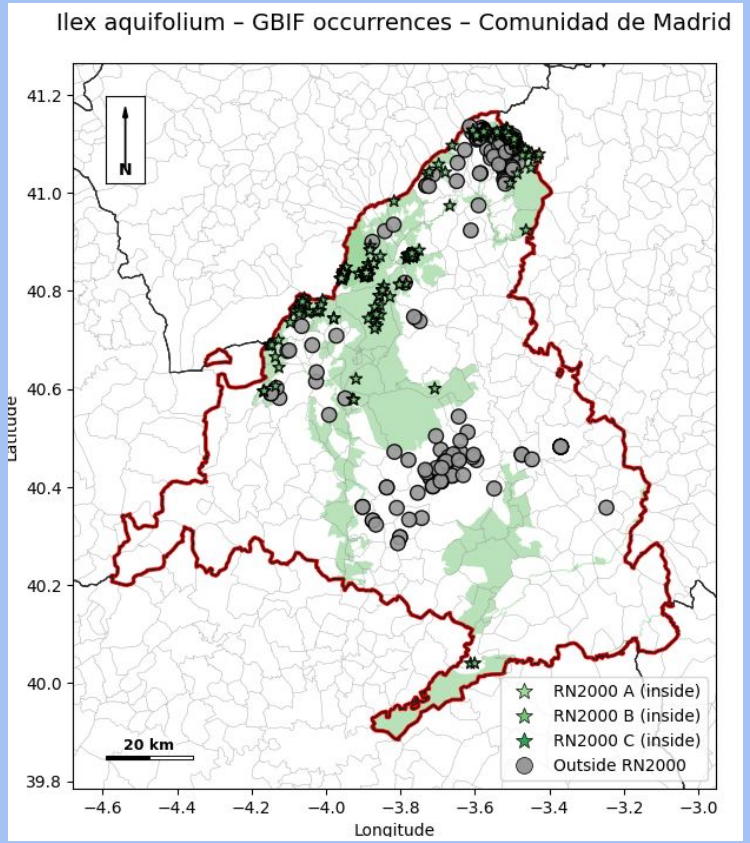
Data sources:

- *In situ* obs.: Global Biodiversity Information Facility (GBIF)
- *Ex situ* conservation: 6 spanish germplasm banks

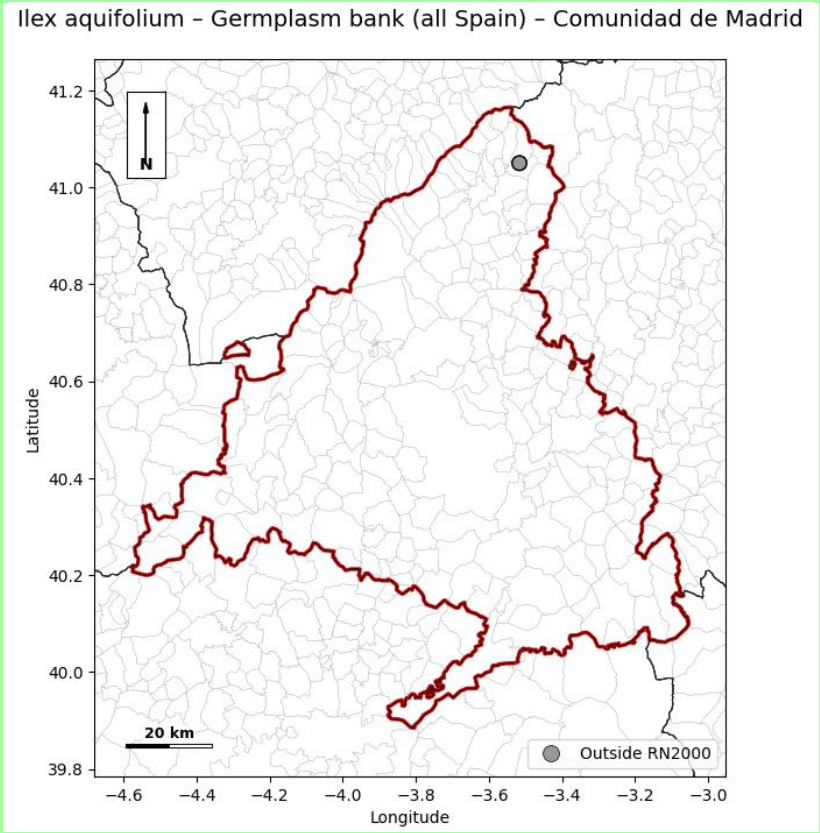
Demo: working with geographical data

Observation sites (GBIF) or collecting sites (banks) vs. RN2000

FLAIR-GG
Connecting Germplasm Resources



GBIF (*in situ*)

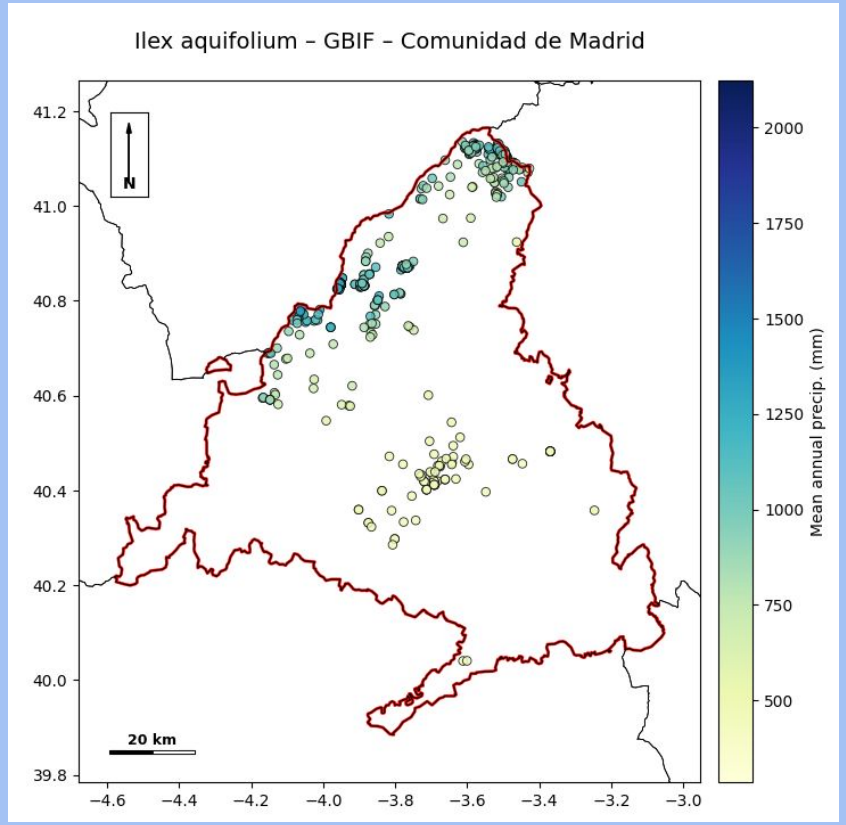


Bancos germoplasma (*ex situ*)

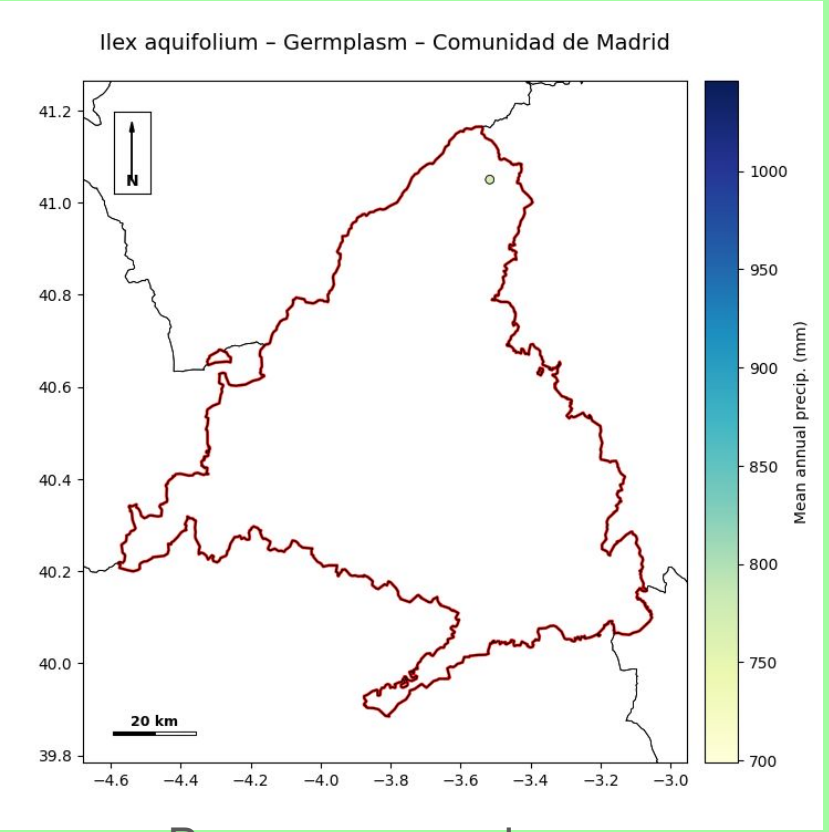
Demo: working with geographical data

Observation sites (GBIF) or collecting sites (banks) vs. average annual precipitation

* cartografía ráster Gonzalo (2010)



GBIF (*in situ*)

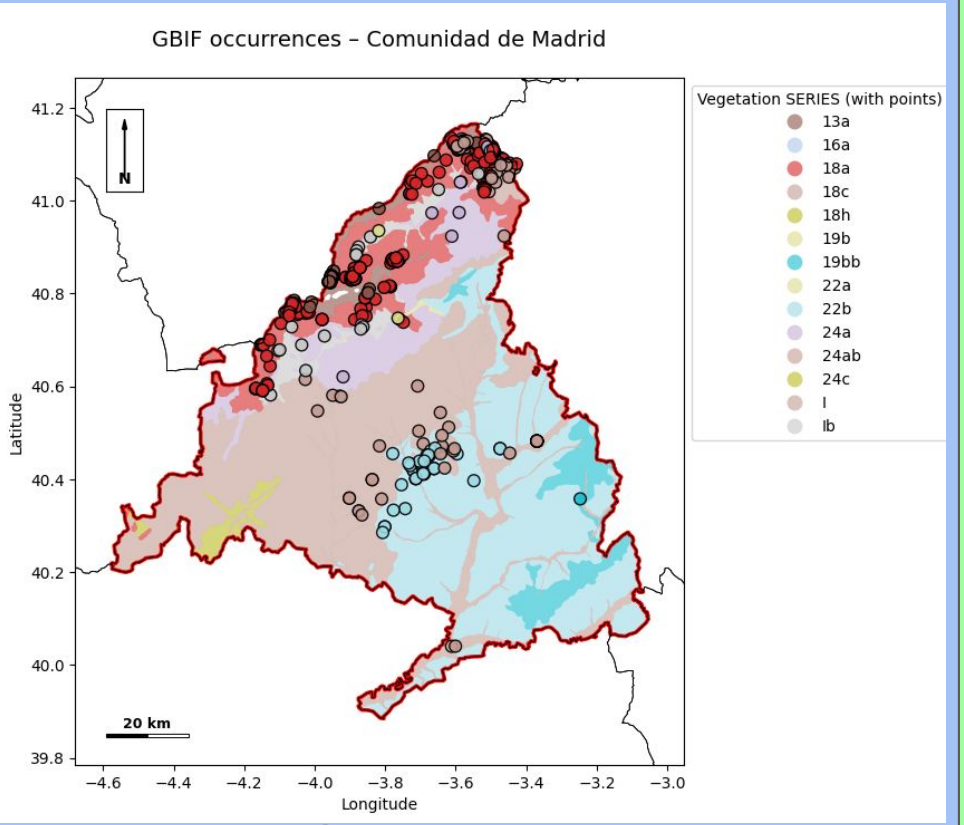


Bancos germoplasma

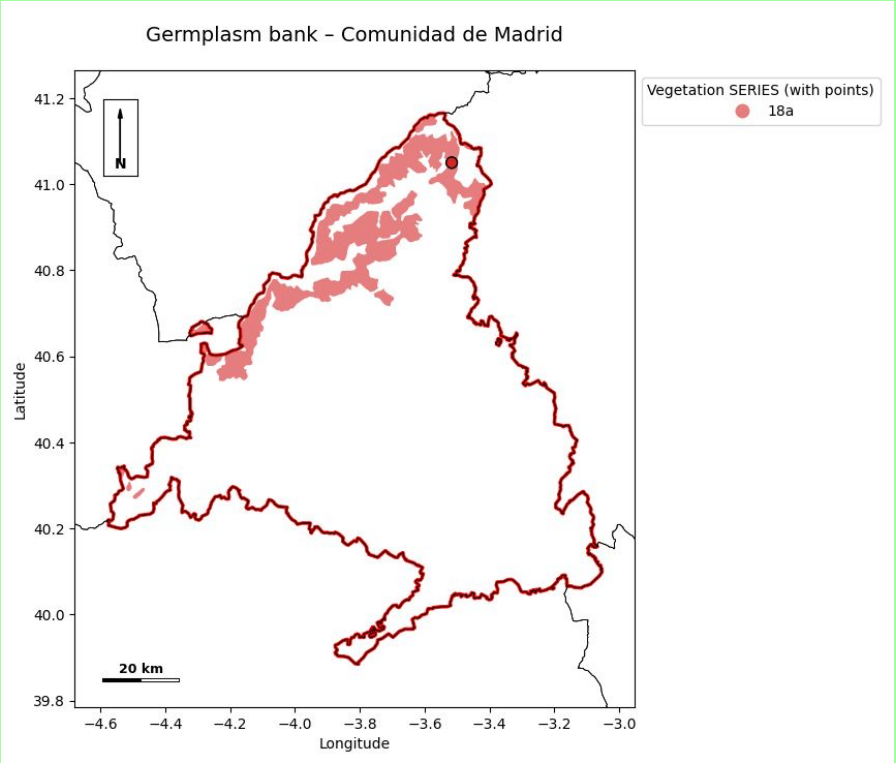
Demo: working with geographical data

Observation sites (GBIF) or collecting sites (banks) vs. vegetation series

FLAIR-GG
Connecting Germplasm Resources



GBIF (*in situ*)



Bancos germoplasma (*ex situ*)

Making the impossible difficult

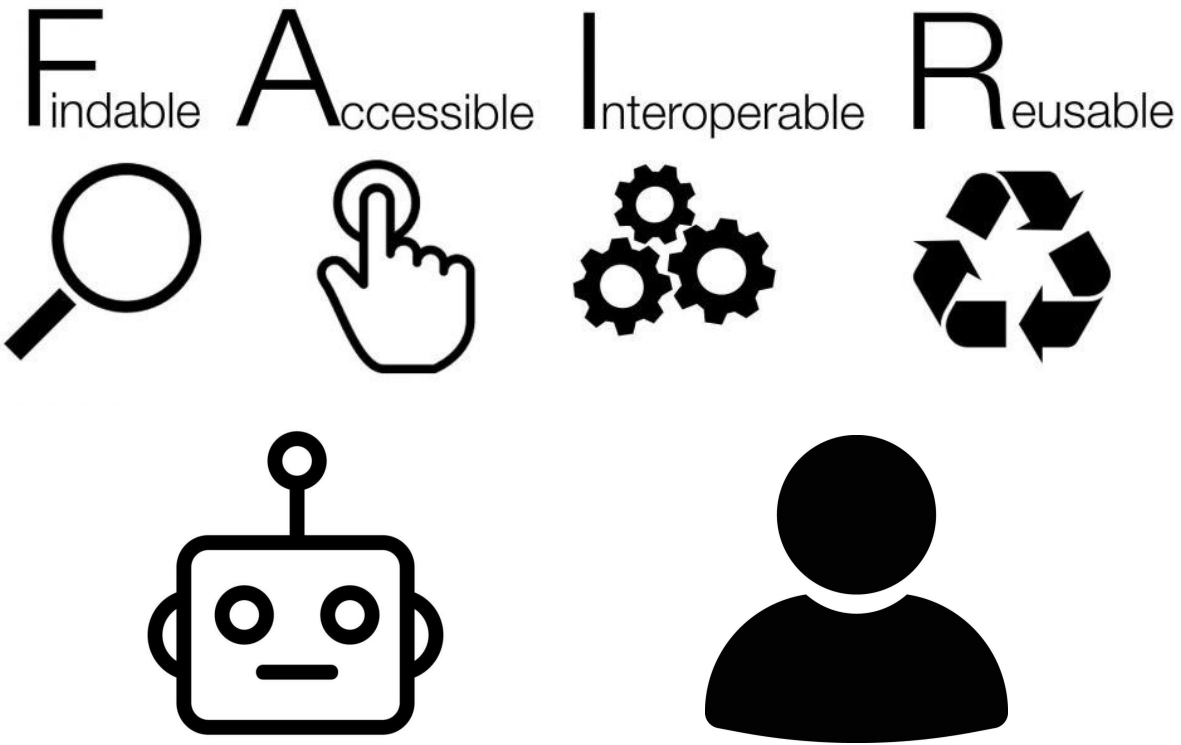
FAIR Makes the Impossible Easy Difficult

But it is worth the effort!

A ***simplified*** explanation of FAIR data

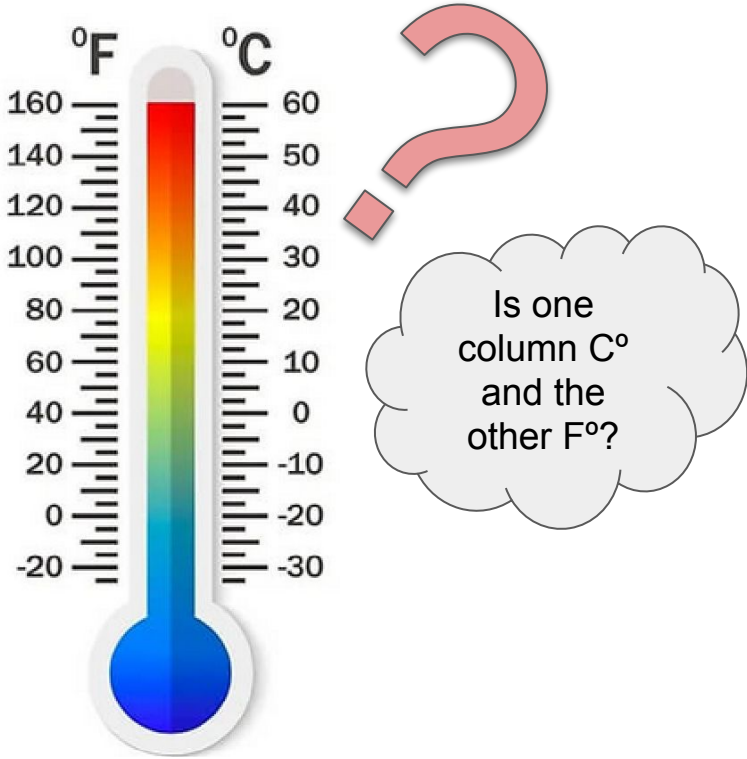
(you can ask me for details later!)

FAIR Principles: Introduction



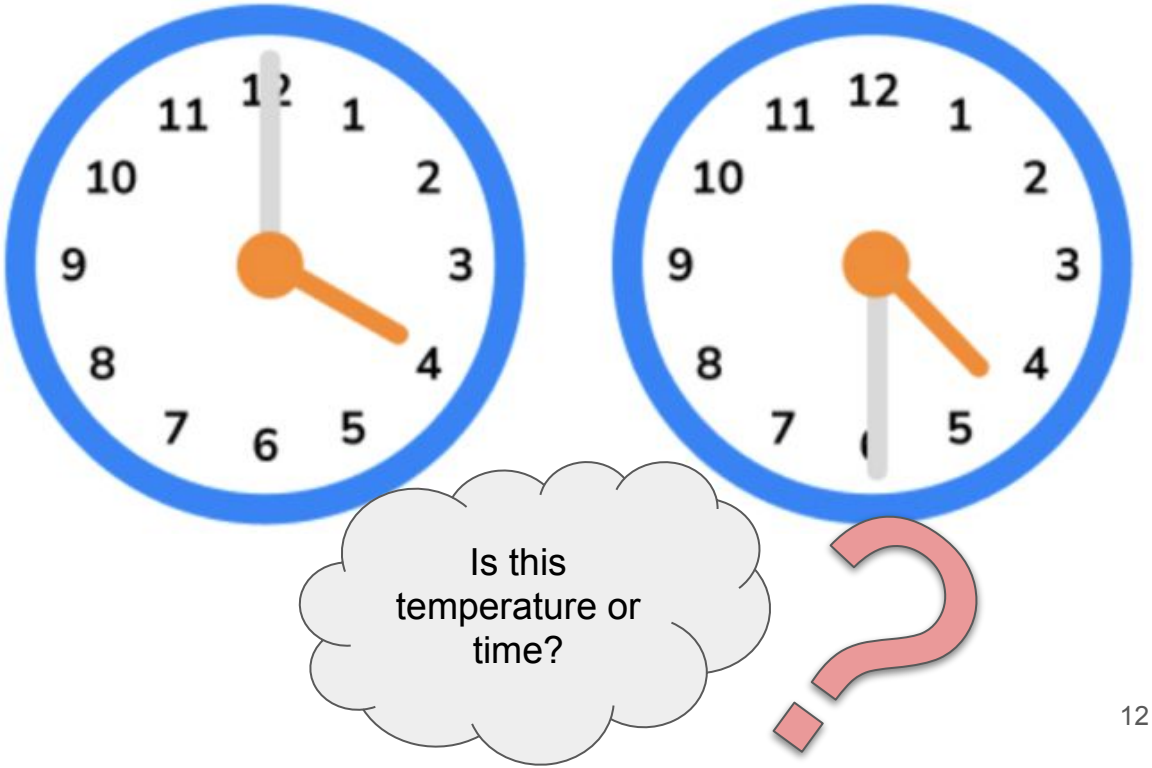
Struggles with non-FAIR data

T1	T2
36.5	97.7
37.0	105
36.8	98.2



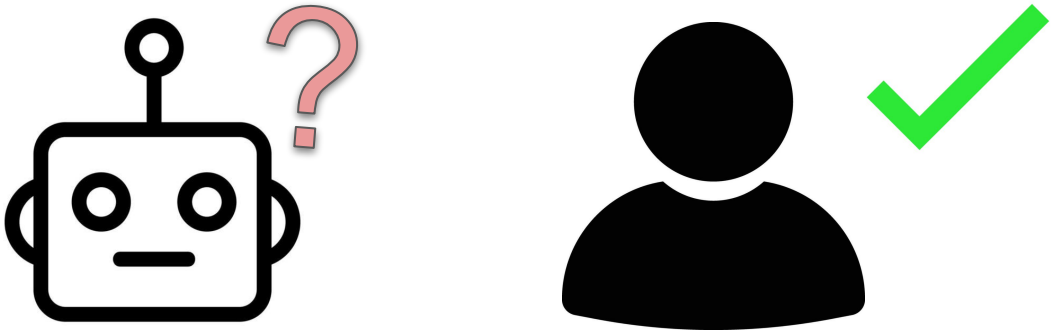
Struggles with non-FAIR data

T1	T2
36.5	97.7
37.0	105
36.8	98.2



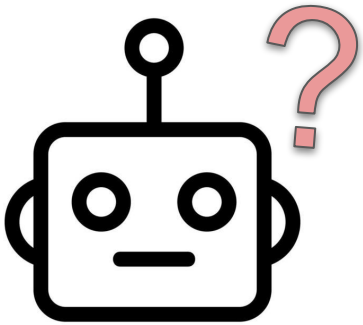
Struggles with non-FAIR data

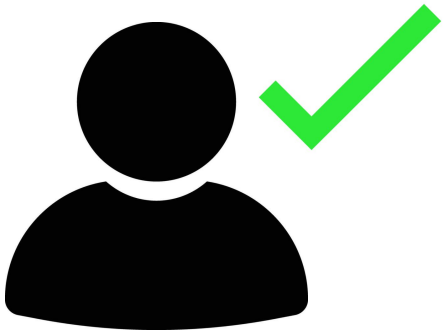
Alberto’s temperature is 36.5ºC



Struggles with non-FAIR data

Alberto was born in Madrid

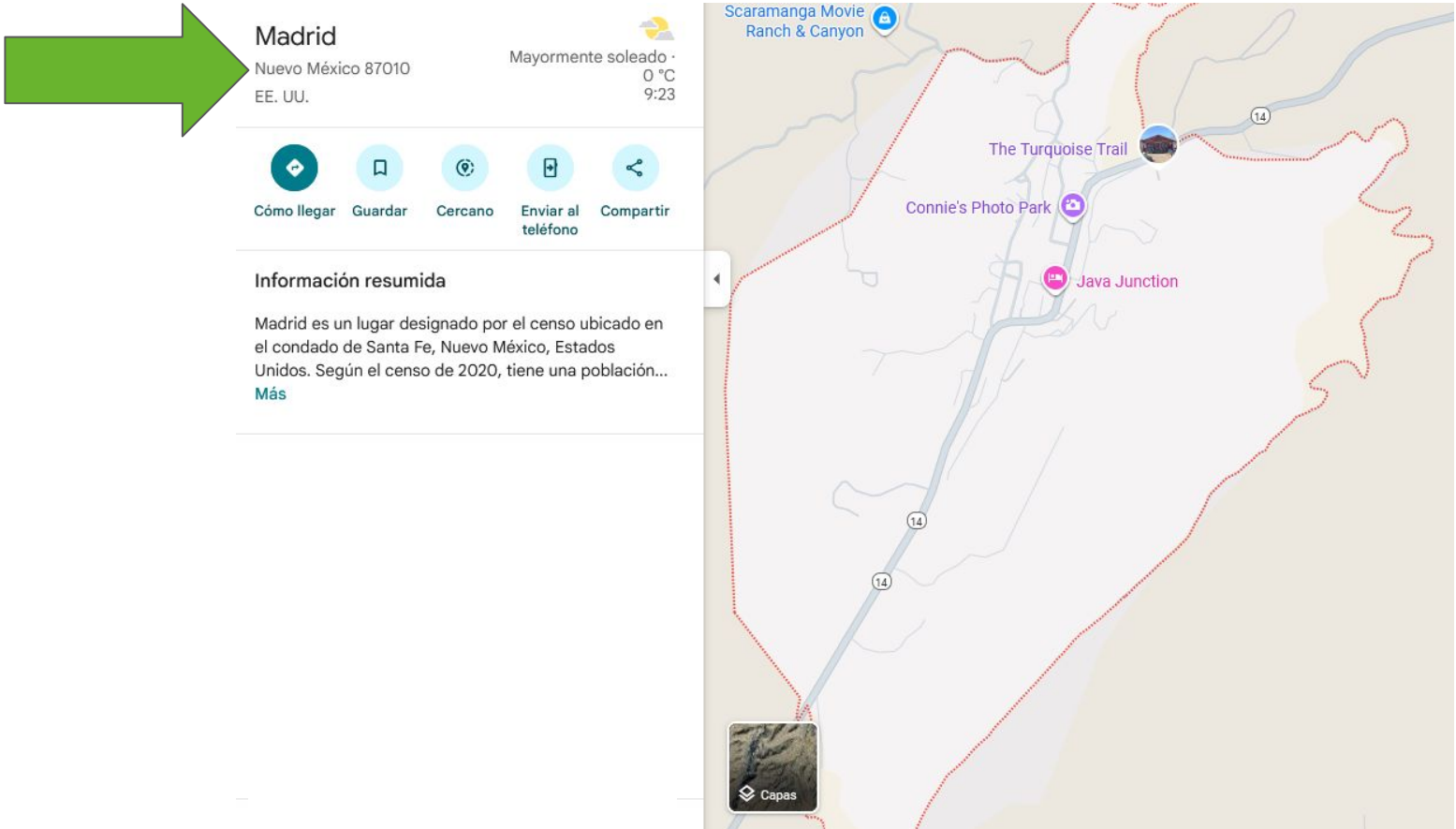




Struggles with non-FAIR data

Alberto was born in Madrid

Alberto was born in Madrid

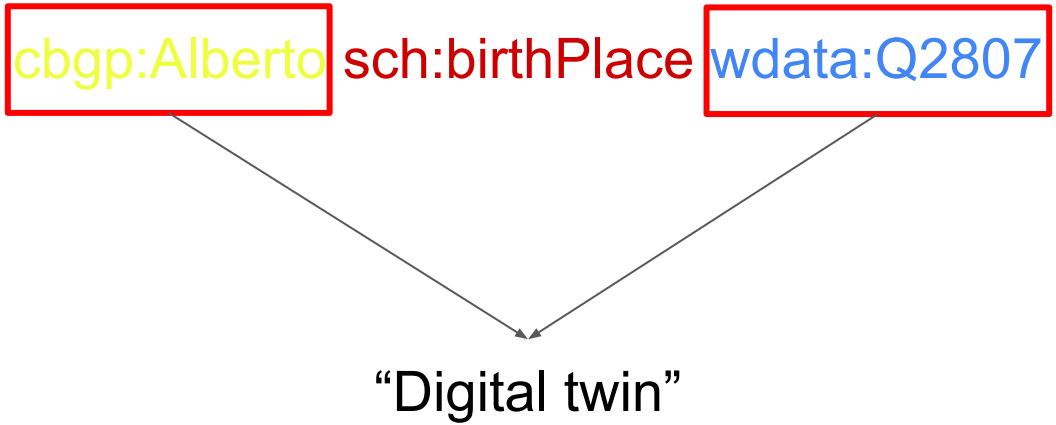


From natural language to triples

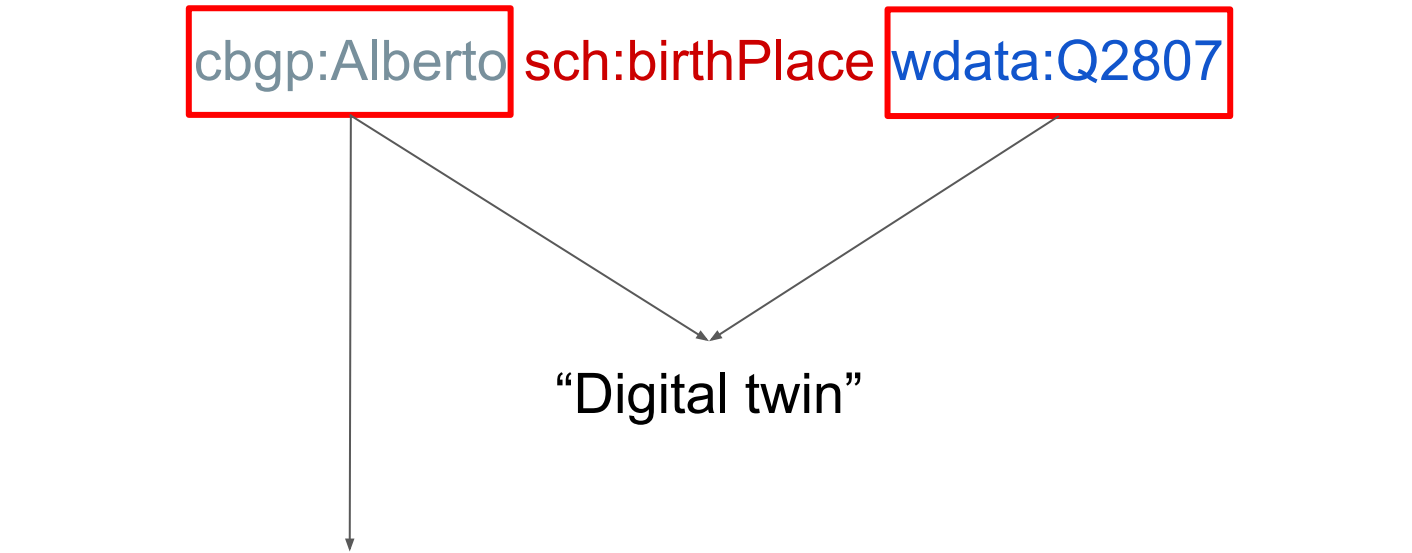
Alberto was born in Madrid

cbgp:Alberto	sch:birthPlace	wdata:Q2807
Subject	Predicate	Object

From natural language to triples

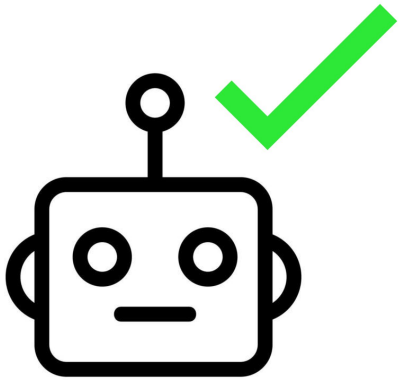


From natural language to triples



From natural language to triples

cbgp:Alberto sch:birthPlace wdata:Q2807



birthPlace

A Schema.org Property

Thing › Property :: birthPlace

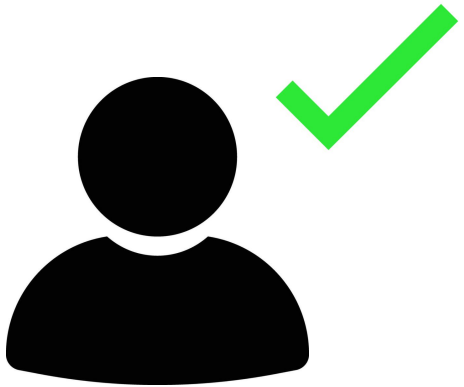
The place where the person was born.

Values expected to be one of these types

- Place

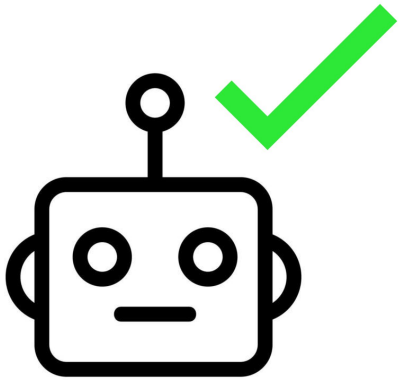
Used on these types

- Person

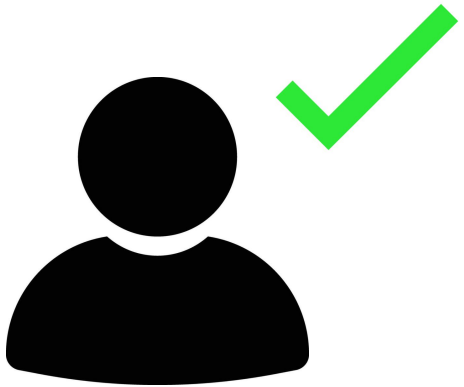


From natural language to triples

cbgp:Alberto sch:birthPlace wdata:Q2807



country	Spain
	1 reference
capital of	Spain
	0 references
	Community of Madrid
	start time
	0 references

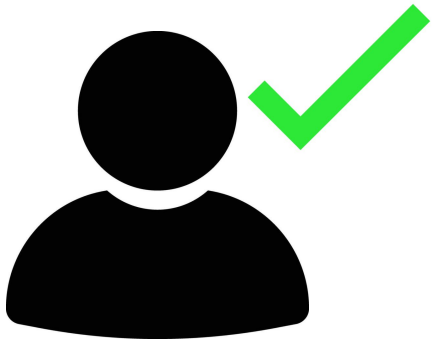


Triple linking

cbgp:Alberto sch:birthPlace wdata:Q2807

Triple linking

Alberto birthPlace Madrid

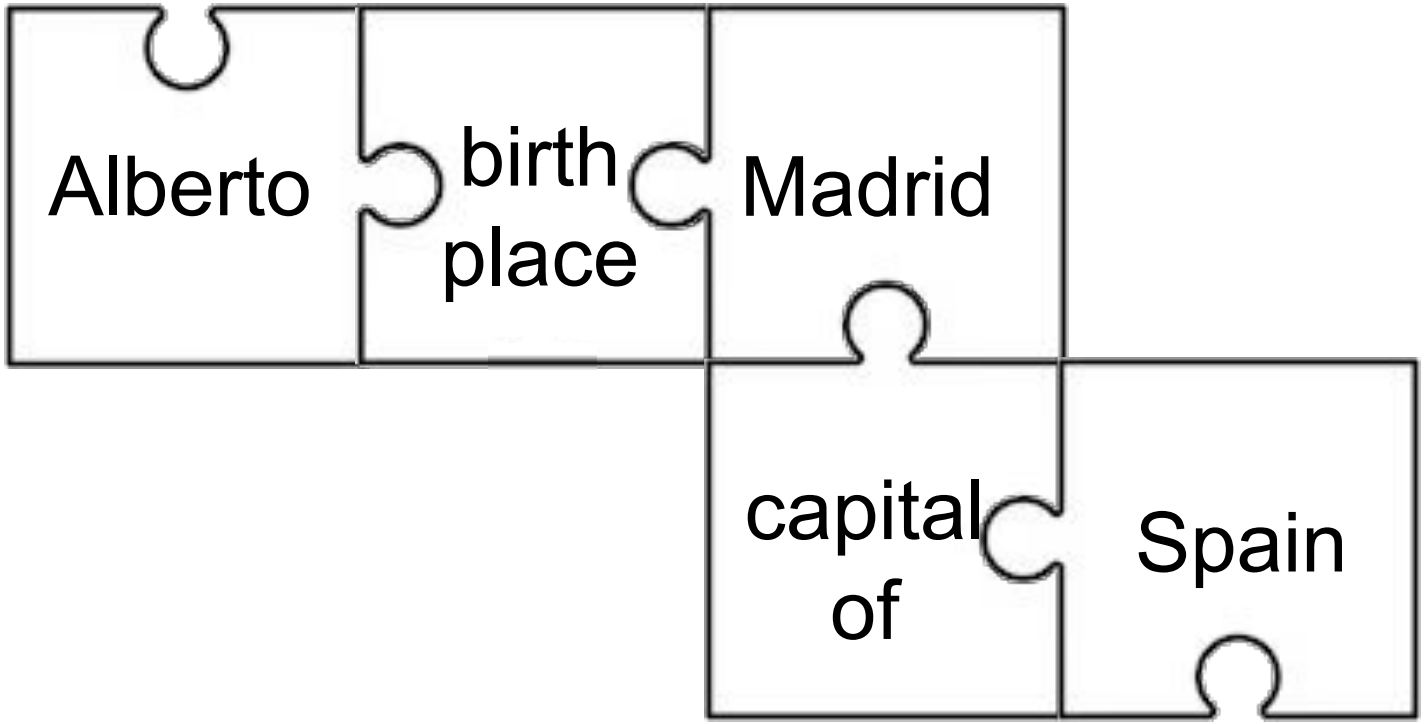


Triple linking

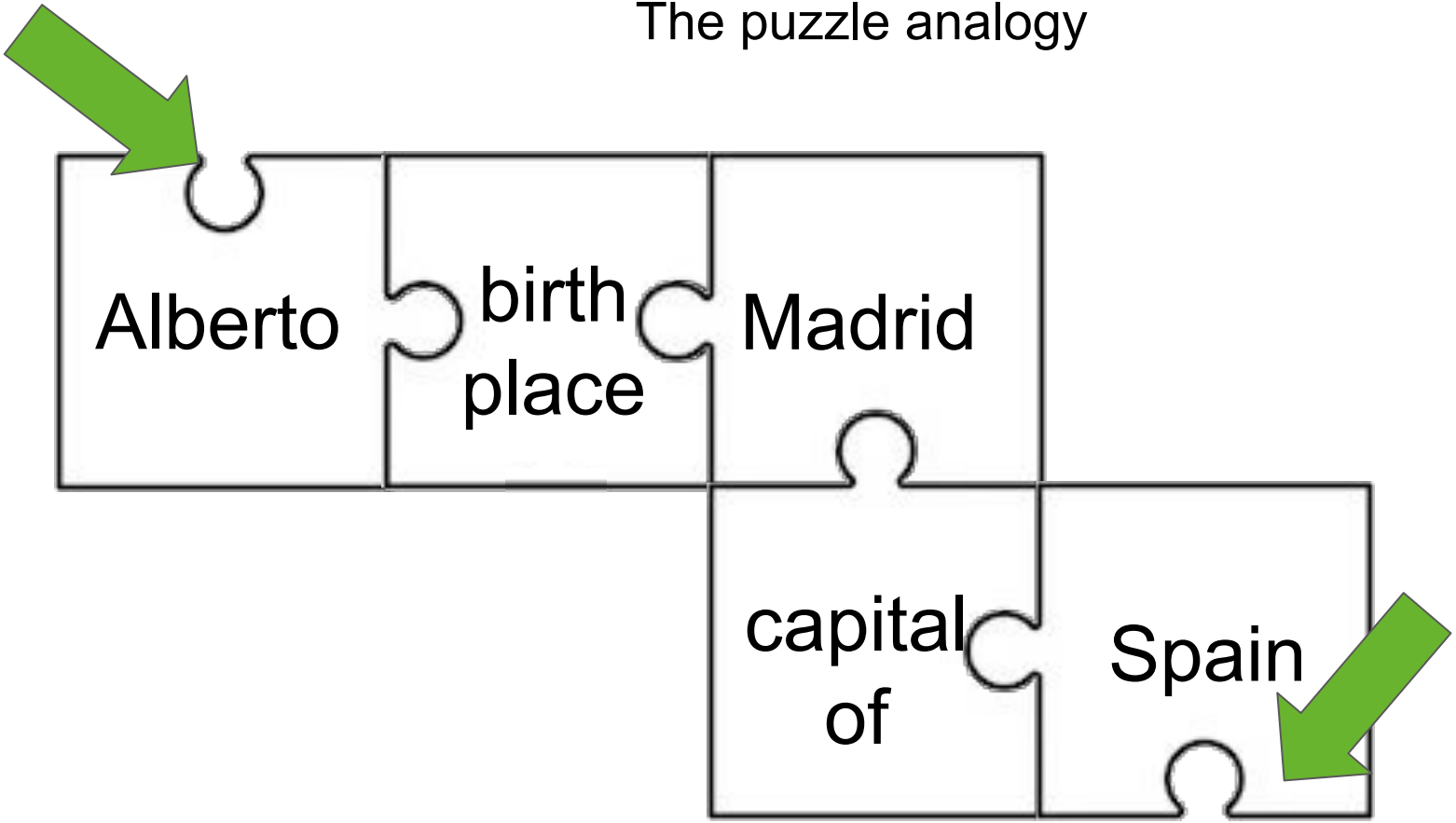
Alberto birthPlace Madrid

Madrid capitalOf Spain

The puzzle analogy



The puzzle analogy

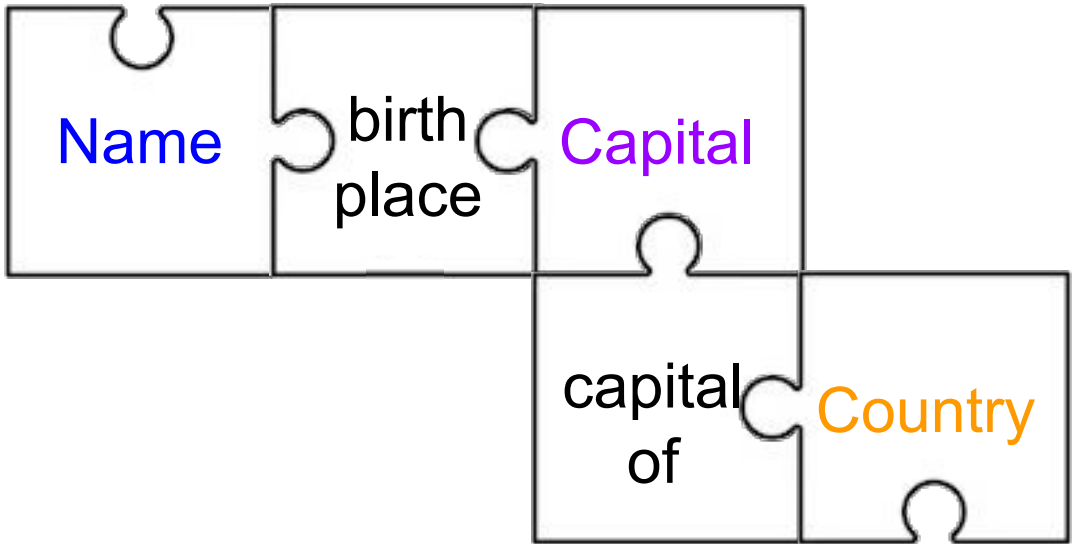


Comparison: Excel vs FAIR Data

Name	Capital	Country
Alberto	Madrid	Spain
Pancho	Madrid	Spain
…	…	…

Comparison: Excel vs FAIR Data

Name	Capital	Country
Alberto	Madrid	Spain
Pancho	Madrid	Spain
…	…	…

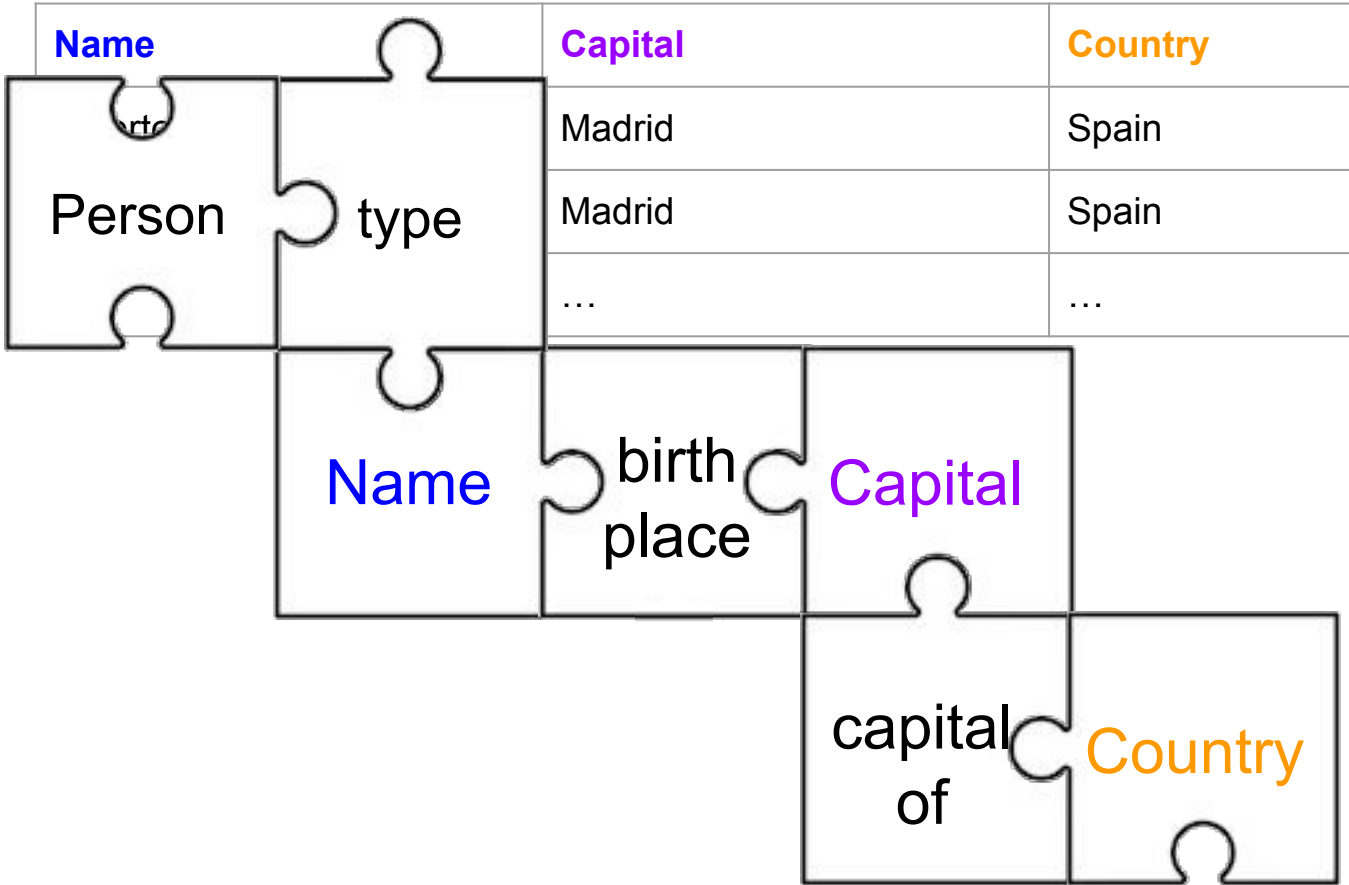


Comparison: Excel vs FAIR Data

Name	Capital	Country
Alberto	Madrid	Spain
Pancho	Madrid	Spain
…		…

Comparison: Excel vs FAIR Data

Name	Capital	Country
	Madrid	Spain
	Madrid	Spain
	…	…



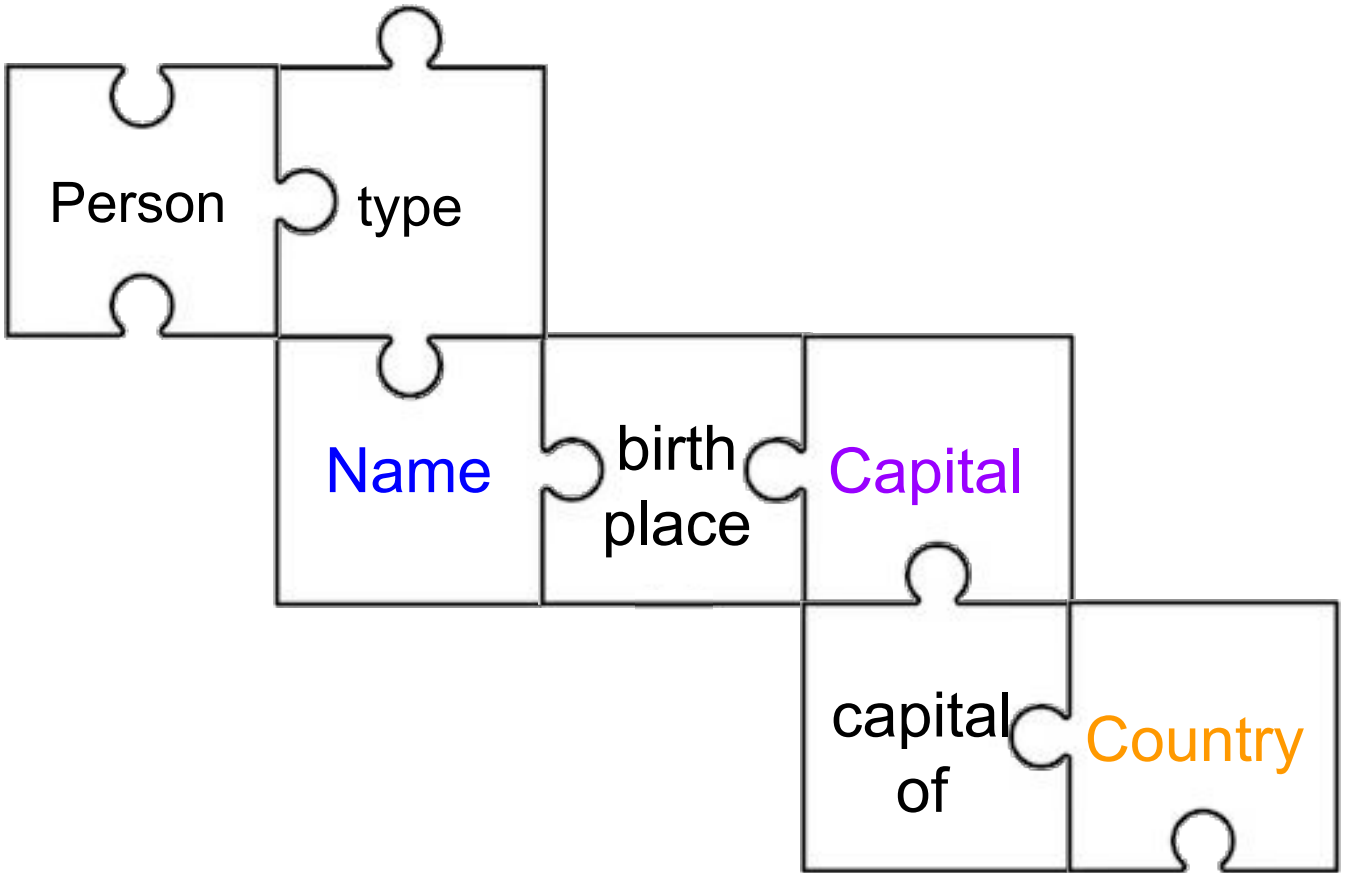
Comparison: Excel vs FAIR Data

Name	Capital	Country
Alberto	Madrid	Spain
Pancho	Madrid	Spain
…	…	…

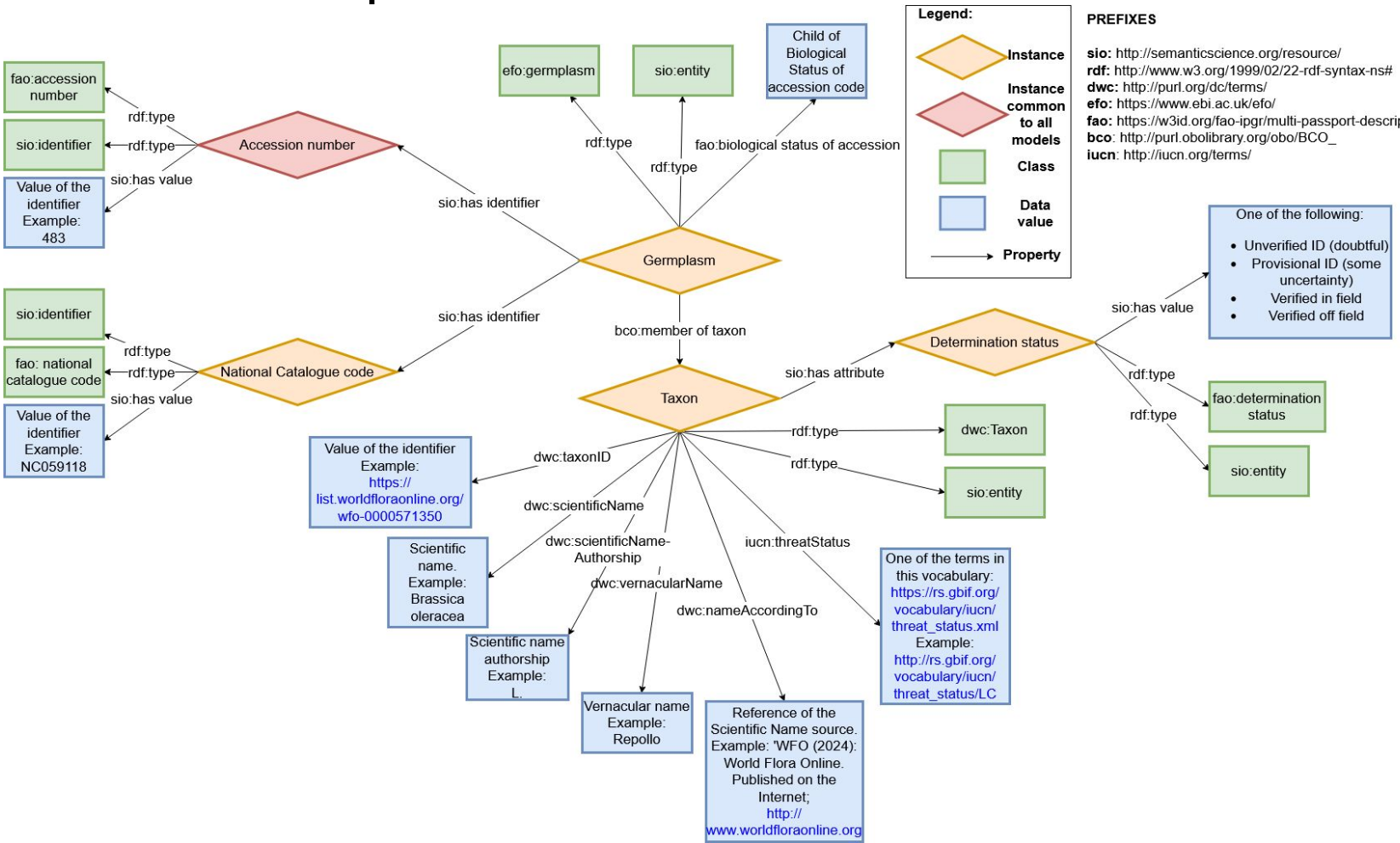
Alberto birthPlace Madrid
Alberto type Person
Madrid capitalOf Spain

Pancho birthPlace Madrid
Pancho type Dog
Madrid capitalOf Spain

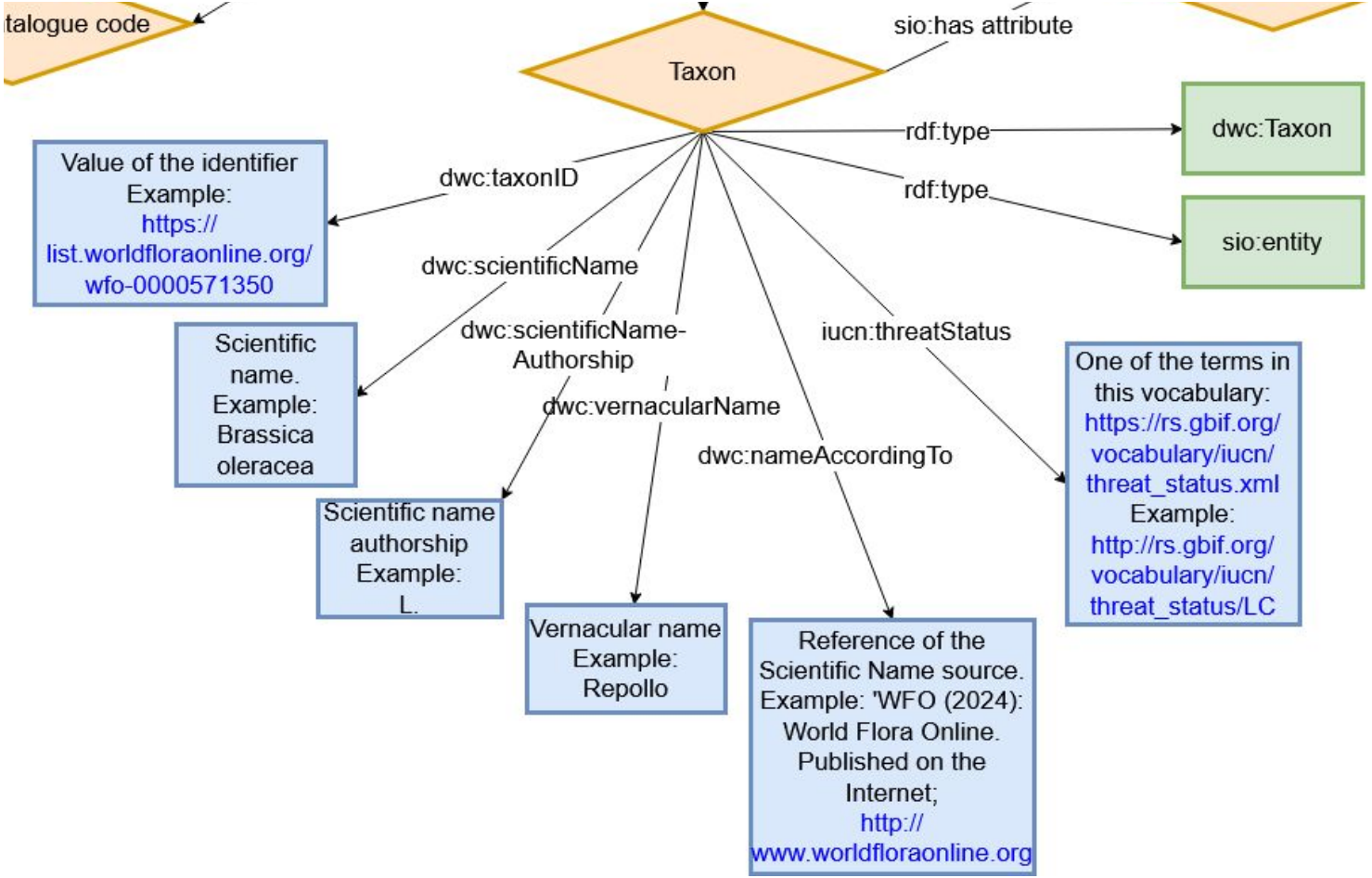
Semantic models



FLAIR-GG Germplasm model

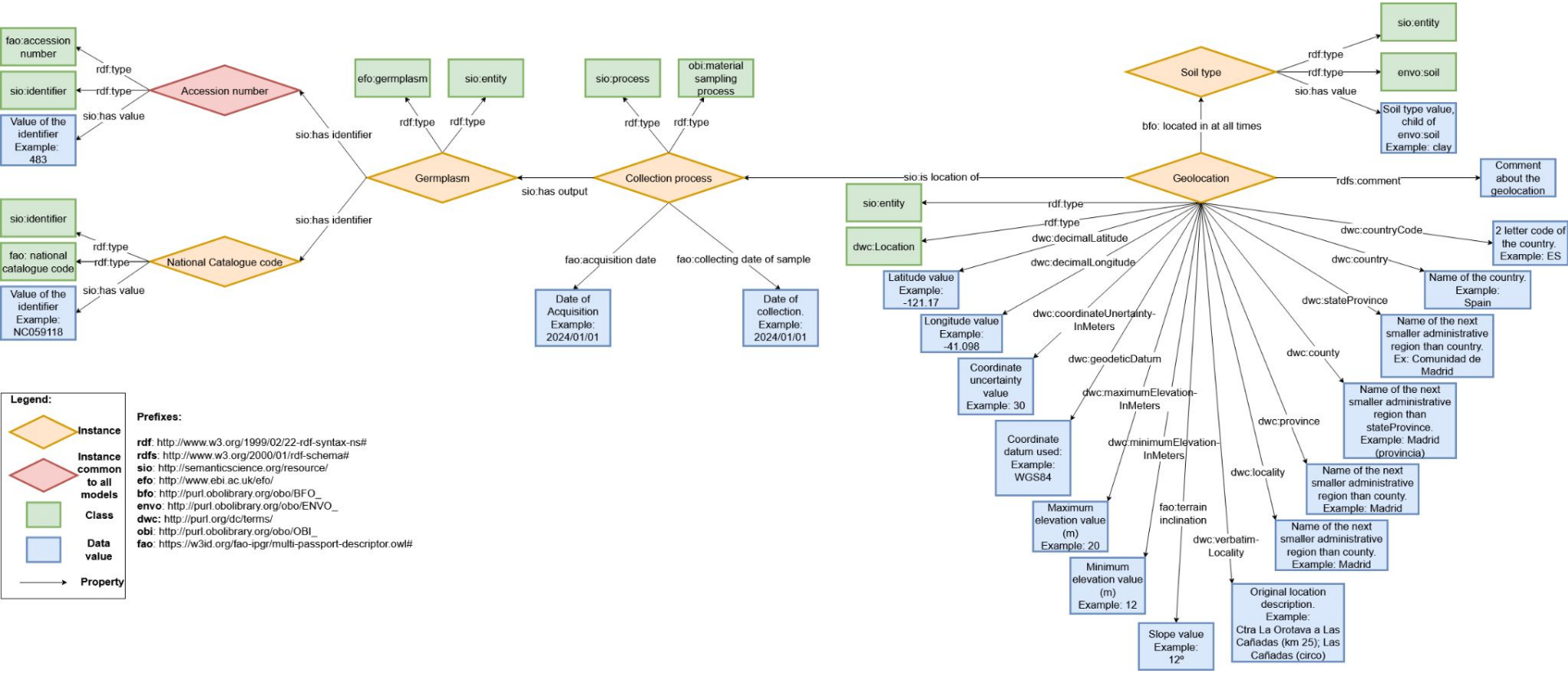


FLAIR-GG Germplasm model

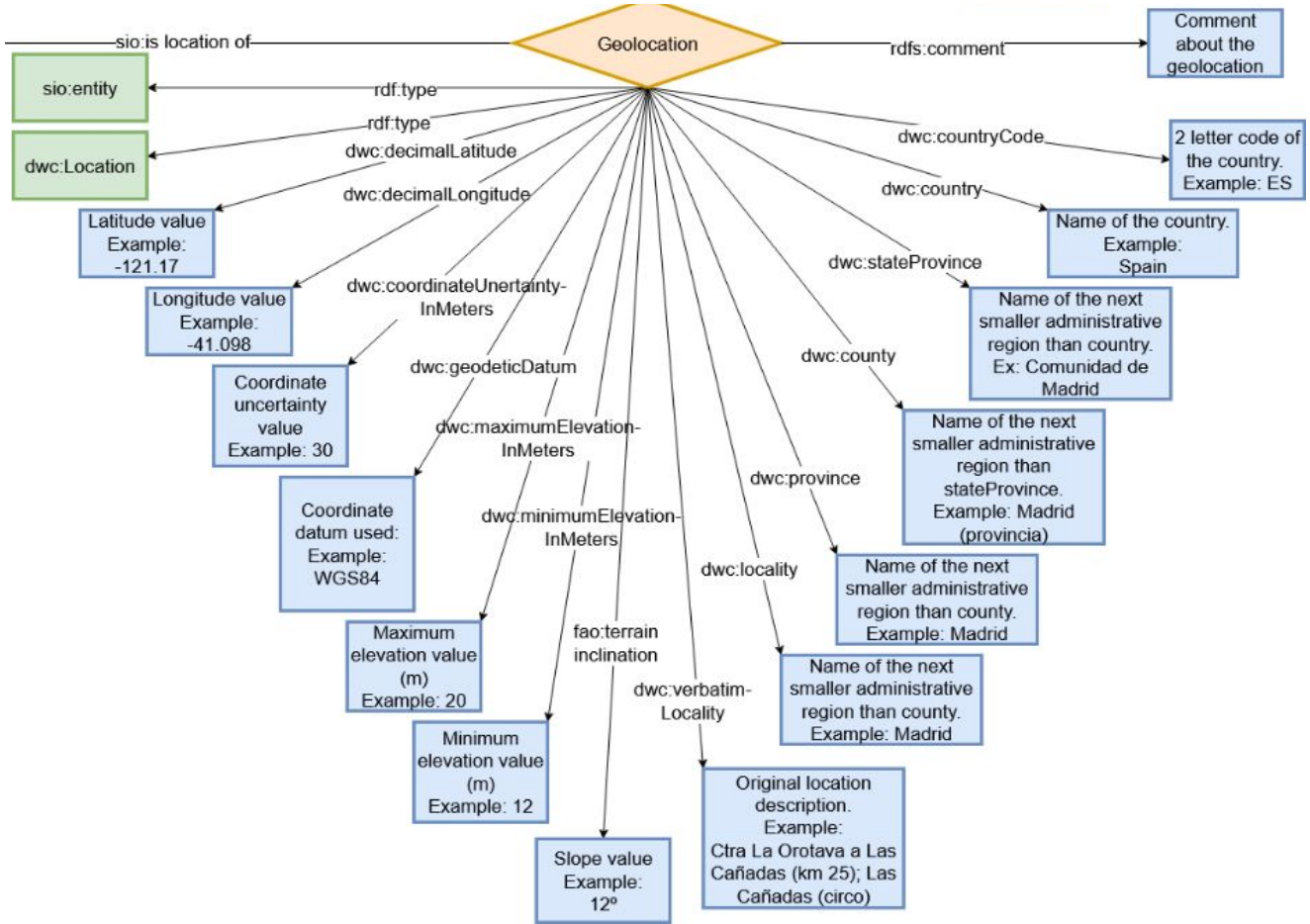


FLAIR-GG Germplasm model

FLAIR-GG Location model



FLAIR-GG Location model

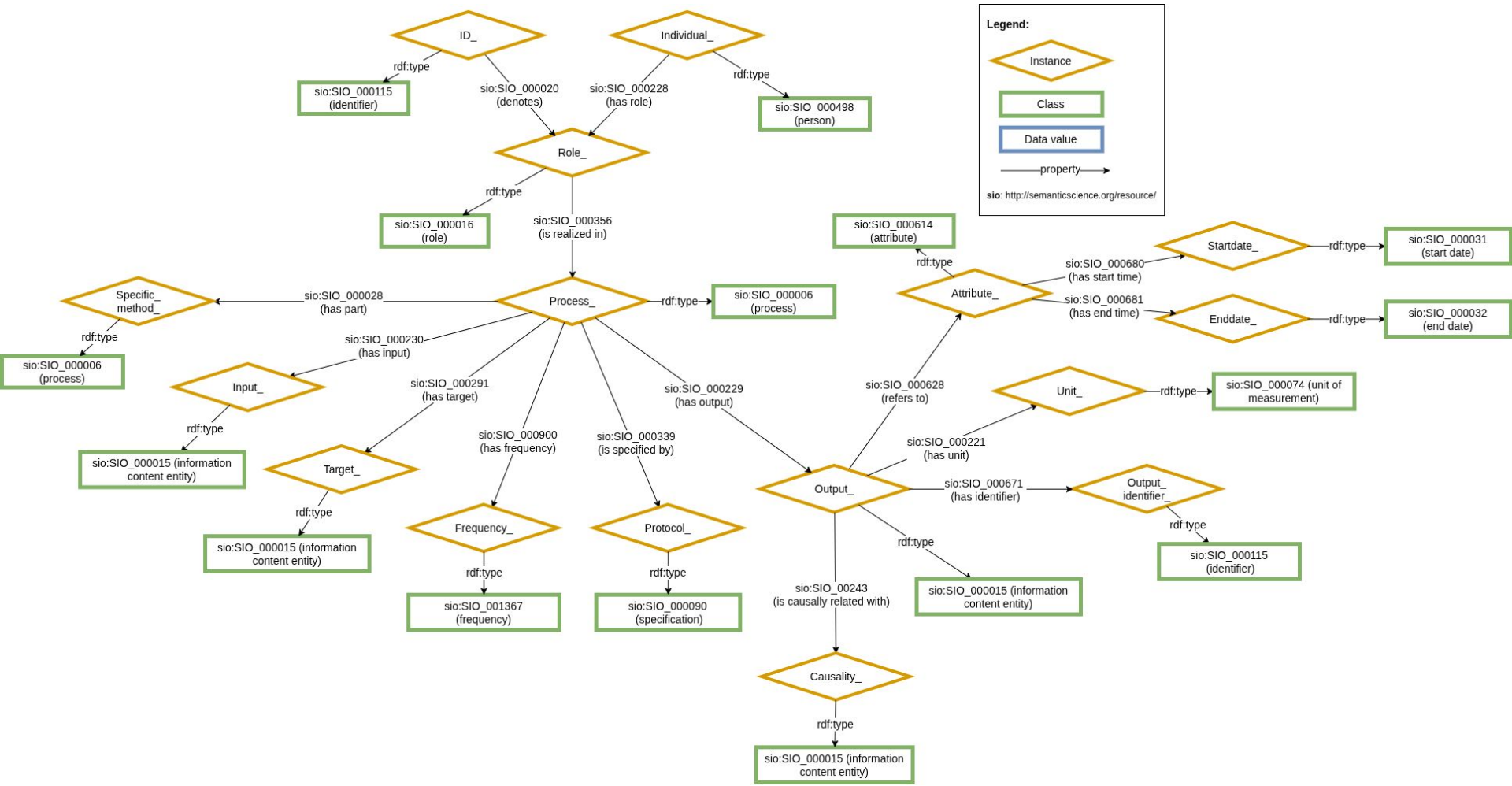


CARE-SM

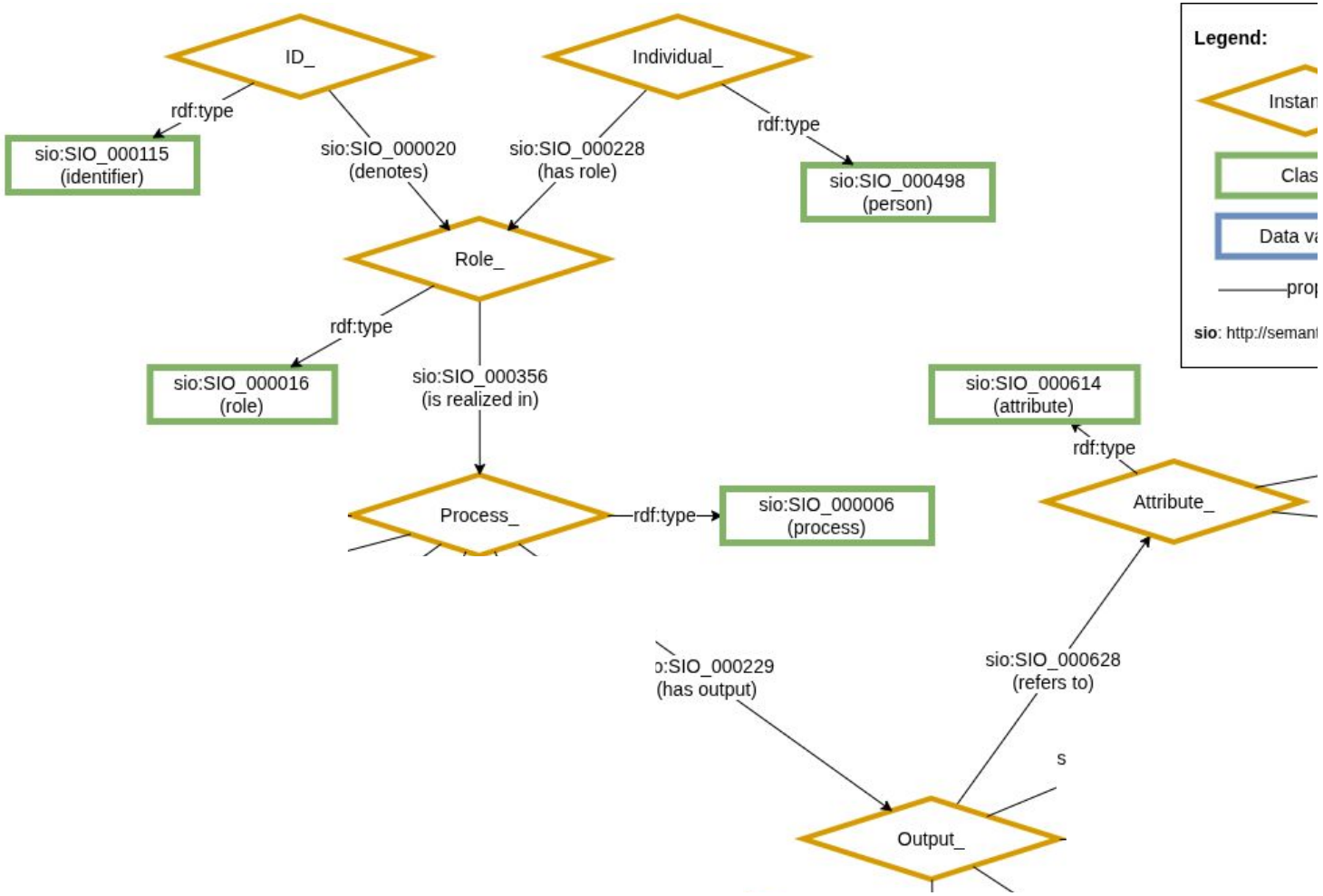
CARE-SM

Clinical And Registry Entries
Semantic Model

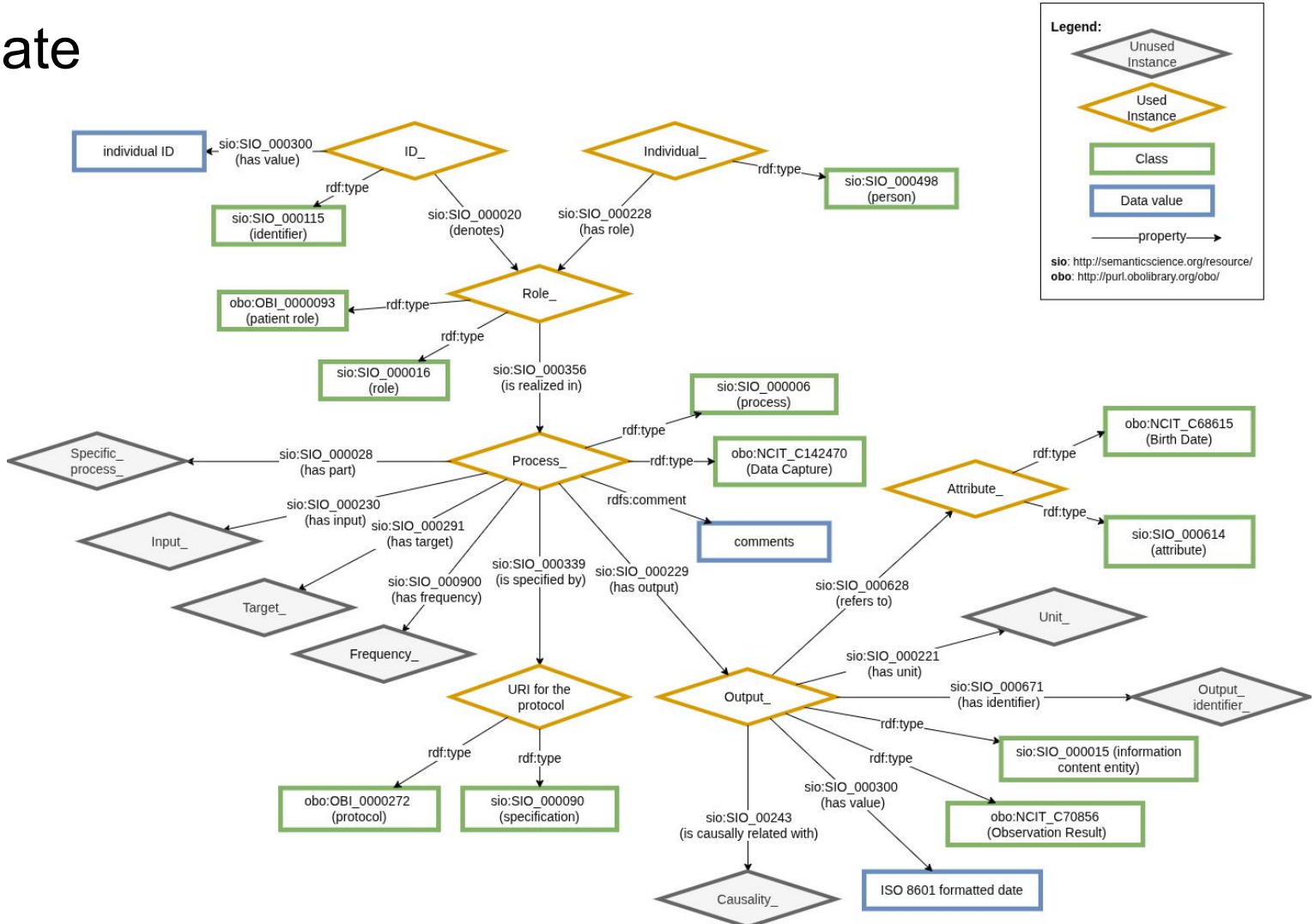
Dr. Pablo Alarcón


Legend:
Instance
Class
Data value
property
sio: http://semanticscience.org/resource/
ID_
rdf:type
sio:SIO_000115 (identifier)
sio:SIO_000020 (denotes)
Individual_
sio:SIO_000228 (has role)
rdf:type
sio:SIO_000498 (person)
Role_
rdf:type
sio:SIO_000016 (role)
sio:SIO_000356 (is realized in)
Specific_ method_
sio:SIO_000028 (has part)
Process_
rdf:type
sio:SIO_000006 (process)
rdf:type
sio:SIO_000006 (process)
sio:SIO_000230 (has input)
Input_
rdf:type
sio:SIO_000015 (information content entity)
sio:SIO_000291 (has target)
Target_
rdf:type
sio:SIO_000015 (information content entity)
sio:SIO_000900 (has frequency)
Frequency_
rdf:type
sio:SIO_001367 (frequency)
sio:SIO_000339 (is specified by)
Protocol_
rdf:type
sio:SIO_000090 (specification)
sio:SIO_000229 (has output)
Output_
sio:SIO_00243 (is causally related with)
Causality_
rdf:type
sio:SIO_000015 (information content entity)
rdf:type
sio:SIO_000015 (information content entity)
sio:SIO_000628 (refers to)
sio:SIO_000614 (attribute)
rdf:type
Attribute_
sio:SIO_000680 (has start time)
Startdate_
rdf:type
sio:SIO_000031 (start date)
sio:SIO_000681 (has end time)
Enddate_
rdf:type
sio:SIO_000032 (end date)
sio:SIO_000221 (has unit)
Unit_
rdf:type
sio:SIO_000074 (unit of measurement)
sio:SIO_000671 (has identifier)
Output_ identifier_
rdf:type
sio:SIO_000115 (identifier)

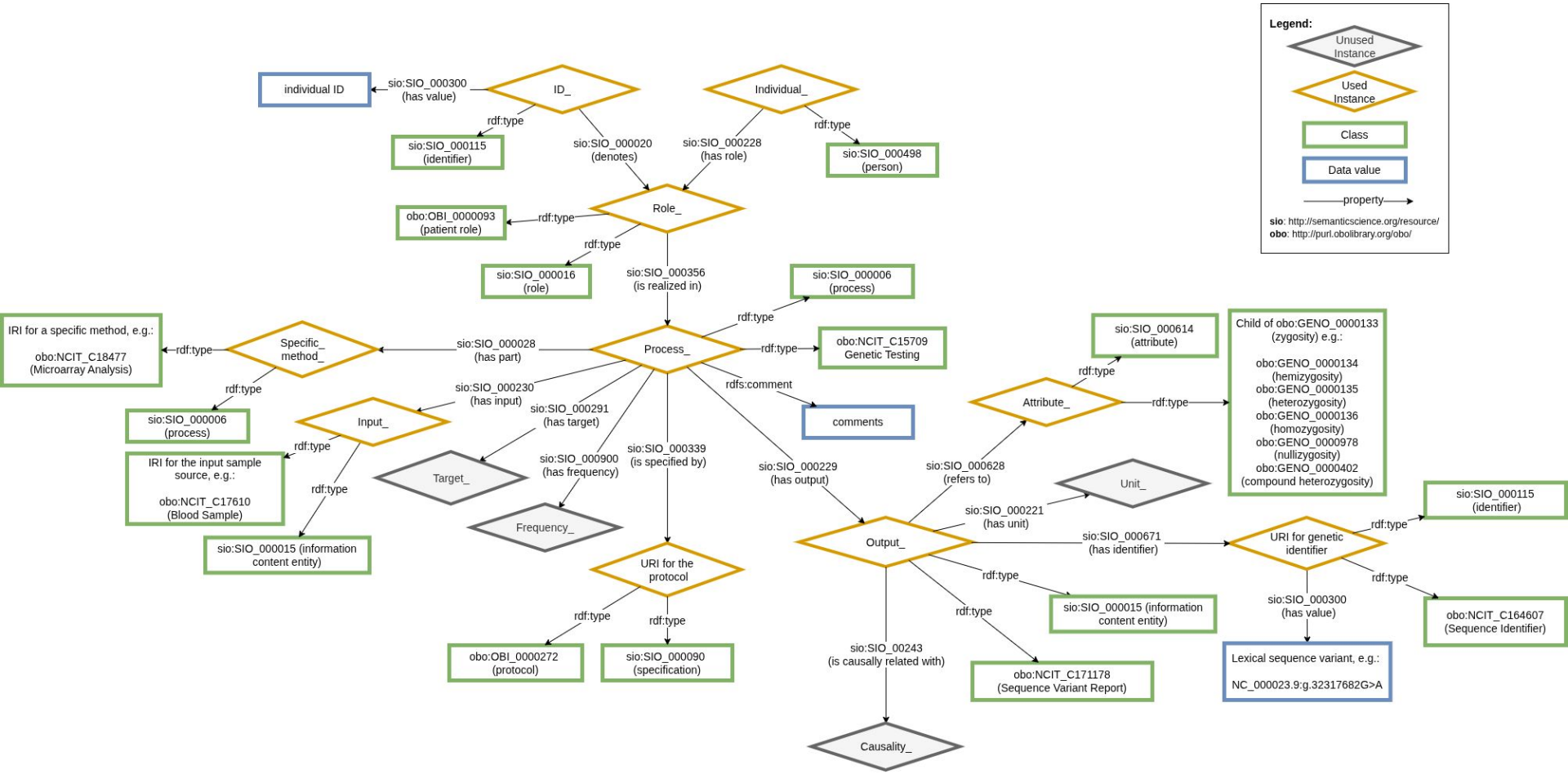
CARE-SM Core model



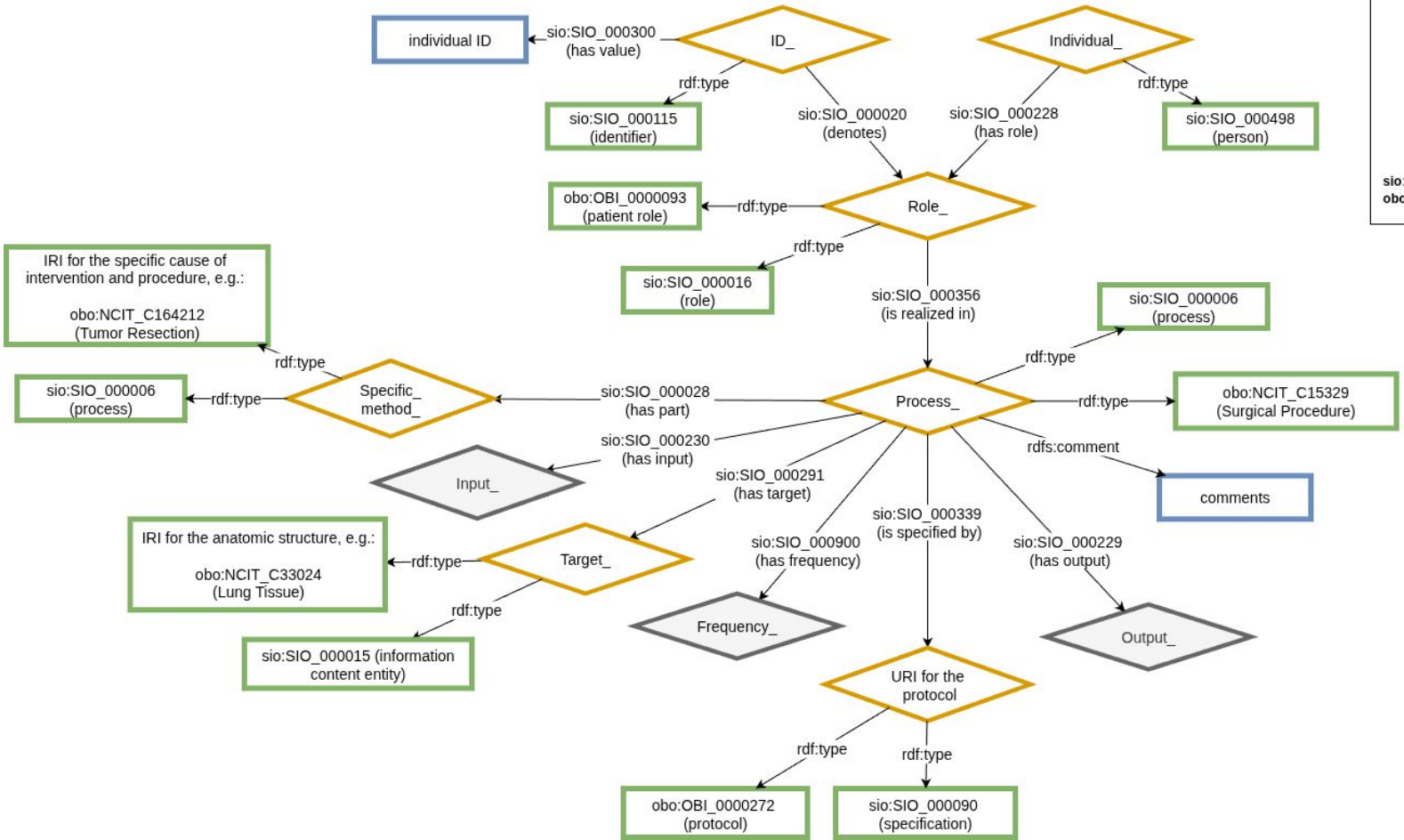
Birthdate


Legend:
Unused Instance
Used Instance
Class
Data value
property
sio: http://semanticscience.org/resource/
obo: http://purl.obolibrary.org/obo/
individual ID
sio:SIO_000300 (has value)
ID_
rdf:type
sio:SIO_000115 (identifier)
sio:SIO_000020 (denotes)
Individual_
rdf:type
sio:SIO_000498 (person)
sio:SIO_000228 (has role)
Role_
obo:OBI_0000093 (patient role)
rdf:type
rdf:type
sio:SIO_000016 (role)
sio:SIO_000356 (is realized in)
sio:SIO_000006 (process)
rdf:type
Specific_ process_
sio:SIO_000028 (has part)
Process_
rdf:type
obo:NCIT_C142470 (Data Capture)
obo:NCIT_C68615 (Birth Date)
rdf:type
Attribute_
rdf:type
sio:SIO_000614 (attribute)
sio:SIO_000230 (has input)
sio:SIO_000291 (has target)
rdfs:comment
Input_
comments
Target_
sio:SIO_000900 (has frequency)
sio:SIO_000339 (is specified by)
sio:SIO_000229 (has output)
sio:SIO_000628 (refers to)
Unit_
Frequency_
sio:SIO_000221 (has unit)
URI for the protocol
Output_
sio:SIO_000671 (has identifier)
Output_ identifier_
rdf:type
sio:SIO_000015 (information content entity)
rdf:type
rdf:type
rdf:type
obo:OBI_0000272 (protocol)
sio:SIO_000090 (specification)
sio:SIO_000300 (has value)
obo:NCIT_C70856 (Observation Result)
sio:SIO_00243 (is causally related with)
Causality_
ISO 8601 formatted date

Genetic diagnosis



Surgical intervention



FAIR Oversight and Testing

Who gets to decide what is and is not FAIR?

EOSC FAIR Metrics and Data Quality Task Force

Co-chair

eosc

FAIR-related
Task Forces

Co-chair

Chris Schubert
Head of Data Centre
Climate Change Centre Austria

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

Key Output from that TF

The choice of Signposting for achieving FAIR objectives

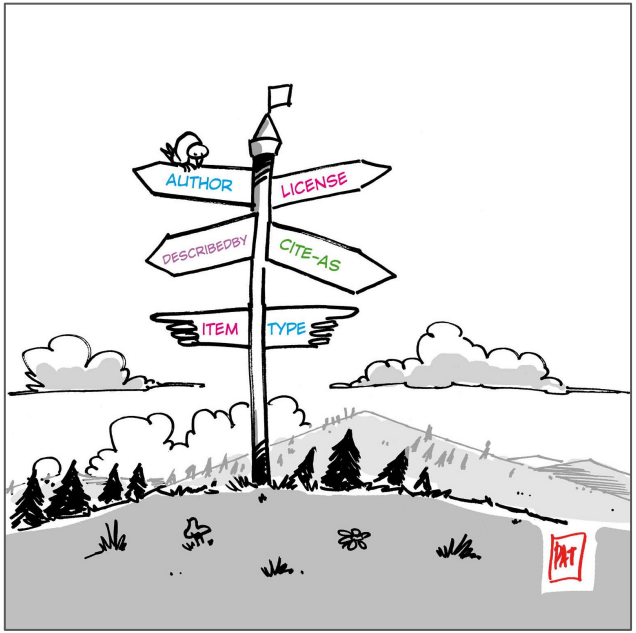


Image courtesy of Patrick Hochstenbach.

A way to publish research outputs on the Web that has
explicit pointers between identifiers, metadata, and data

Key Output from that TF

The choice of Signposting for achieving FAIR objectives

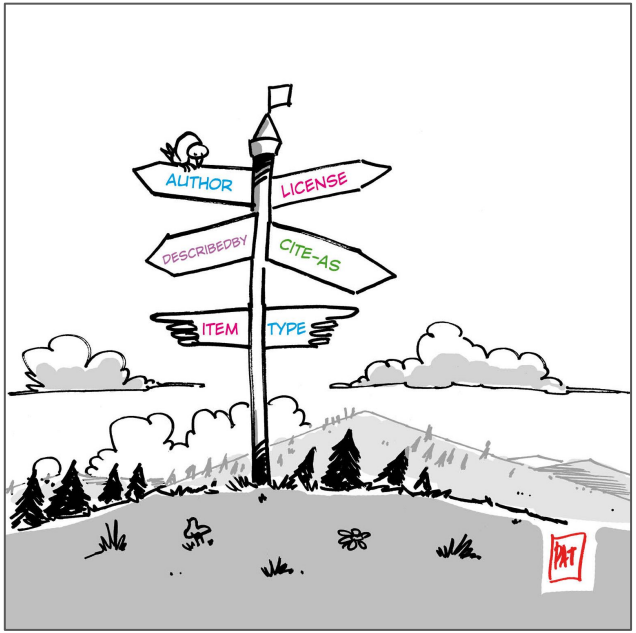


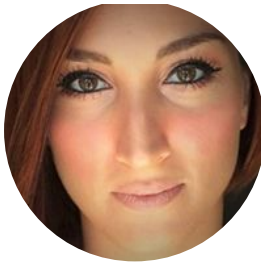
Image courtesy of Patrick Hochstenbach.

A way to publish research outputs on the Web
SO THAT A MACHINE CAN TRULY REUSE IT!

EOSC Task Force Phase 2

Co-Chair

FAIR-related
Task Forces



Co-Chair

Elli Papadopoulou,
Athena Research Center

PHASE 2 (2024-2026)

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?

This work is ongoing... stay tuned!

Making FAIRness part of every project, at birth!

EOSC Infrastructure, led from OpenAIRE, Greece

OBJECTIVE: Plan, Track and Assess all EU Research Activity

PLAN: Machine-actionable Data Management Plans (maDMPs) can be “investigated” to test completeness and plausibility

TRACK: maDMPs can be “executed” to test compliance (did you do what you said you would do?)

ASSESS: All data can be automatically assessed for its completeness, standards-compliance, and FAIRness

FAIR Assessment

Checking on progress…

FAIR Champion

What do you want to do today?

- List all Tests in OSTrails Registry
- Run individual Tests on a Digital Object (this sends you to the "Tests" server; Champion tests only!)
- Execute a Benchmark Quality Assessment on a Digital Object
- List Champion Benchmark Quality Assessment Algorithms
- Register a new Benchmark Quality Assessment Algorithm
- Register a new Metric Test

FAIR Champion

What do you want to do today?

- List all Tests in OSTrails Registry
- Run individual Tests on a Digital Object (this sends you to the "Tests" server; Champion tests only!)
- Execute a Benchmark Quality Assessment on a Digital Object
- List Champion Benchmark Quality Assessment Algorithms
- Register a new Benchmark Quality Assessment Algorithm
- Register a new Metric Test

125 Tests Available in the Registry

!!! And growing!!!

Generic CURL command:

```
curl -H "Accept: application/json" https://tools.ostrails.eu/champion/tests
```

Keyword Filter (one word, case-insensitive):

filter by keyword

FAIR Champion: Data Authorization

Test a discovered data GUID for the ability to implement authentication and authorization in its resolution protocol. Currently passes InChI Keys, DOIs, Handles, and URLs. It also searches the metadata for the Dublin Core 'accessRights' property, which may point to a document describing the data access process. Recognition of other identifiers will be added upon request by the community.

API Definition

▶ Additional Details

▶ Execute Test

FAIR Champion

What do you want to do today?

- List all Tests in OSTrails Registry
- Run individual Tests on a Digital Object (this sends you to the "Tests" server; Champion tests only!)
- Execute a Benchmark Quality Assessment on a Digital Object
- List Champion Benchmark Quality Assessment Algorithms
- Register a new Benchmark Quality Assessment Algorithm
- Register a new Metric Test

Benchmarks: The new approach to FAIR Testing

Consortium of European Social Science Data Archives (CESSDA) Benchmark

GENERAL INFORMATION

cessda

FAIR Benchmark - CESSDA Data Catalogue (FB-CDC)

10.25504/FAIRsharing.e4d13f

Type	Benchmark
Registry	FAIRassist
Description	The CESSDA Data Catalogue Benchmark is intended to ensure that metadata records it contains are FAIR. It is of primary utility for data curators, re assessors, and data infrastructure maintainers working with social science research data. This benchmark coordinates the FAIR assessment of CDC generic and specialised metrics.
Homepage	https://github.com/cessda/cessda.fair-tests/blob/main/README.md
Year of Creation	2026
Maintainers	pabloalarcon, john.shepherdson
Countries developing this resource	United Kingdom
Object types in scope for this resource	Dataset
Subjects	Humanities And Social Science
Domains	Benchmarking, FAIR
Taxonomic Range	Not Applicable
User Defined Tags	N/A

FAIRsharing.org
standards, databases, policies

Consortium of European Social Science Data Archives (CESSDA) Benchmark

FAIR Champion

What do you want to do today?

- List all Tests in OSTrails Registry
- Run individual Tests on a Digital Object (this sends you to the "Tests" server; Champion tests only!)
- Execute a Benchmark Quality Assessment on a Digital Object
- List Champion Benchmark Quality Assessment Algorithms
- Register a new Benchmark Quality Assessment Algorithm
- Register a new Metric Test

Benchmarks: Can be automatically Applied

Execute Benchmark Assessment Algorithm

Algorithm Configuration Google Spreadsheet URL: (Example)

https://docs.google.com/spreadsheets/d/1Nk0vM4yBpVQTo_UbB62NY_fz93aRZRHBZGh5fG-khOw

You may now apply the selected algorithm to a new Digital Object, or an existing Test Result Set

GUID: https://datacatalogue.cessda.eu/detail/a1667722691cdc8f1bfe7224

OR

Apply Algorith to Existing Test Result Set (JSON-LD):

Choose File No file chosen

Run Benchmark Quality Assessment

Benchmark Assessment Outputs

▶ Details

Test Results

Test ID	Result	Weight
F1_PID	indeterminate	0.0
F1_GUID	pass	1.0
F2A	pass	3.0
F2B	pass	1.0
F4	indeterminate	0.0
A1_1	indeterminate	0.0
A1_2	indeterminate	0.0
I1_A	indeterminate	0.0
I2_A	fail	-1.0
R1_2_CPI	indeterminate	0.0
R1_3_CEK	indeterminate	0.0
R1_3_CTV	indeterminate	0.0
R1_3_DMOCV	indeterminate	0.0
R1_3_DAUV	indeterminate	0.0
R1_3_DTMV	indeterminate	0.0
R1_3_DSPV	indeterminate	0.0

Individual Test Results

All tests were defined by ***your Community Members*** as being relevant and appropriate

You are being “judged by your peers”!

This is a completely novel approach to FAIR Assessment

MOST IMPORTANT

Conclusions:

Condition	Discussion	Guidance
1	Unacceptable: metadata are not assigned persistent identifiers	Identifiers use ARK, DOI, HANDLE or URN PID schemas. More Information DANS FAIR Guidance. More Information
2	Acceptable: metadata are assigned globally unique identifiers	
3	Acceptable: data are described with structured metadata	
4	Acceptable: data are described with grounded metadata	
5	Unacceptable: metadata are not searchable	FAIR Principles F4. More Information DANS FAIR Guidance. More Information

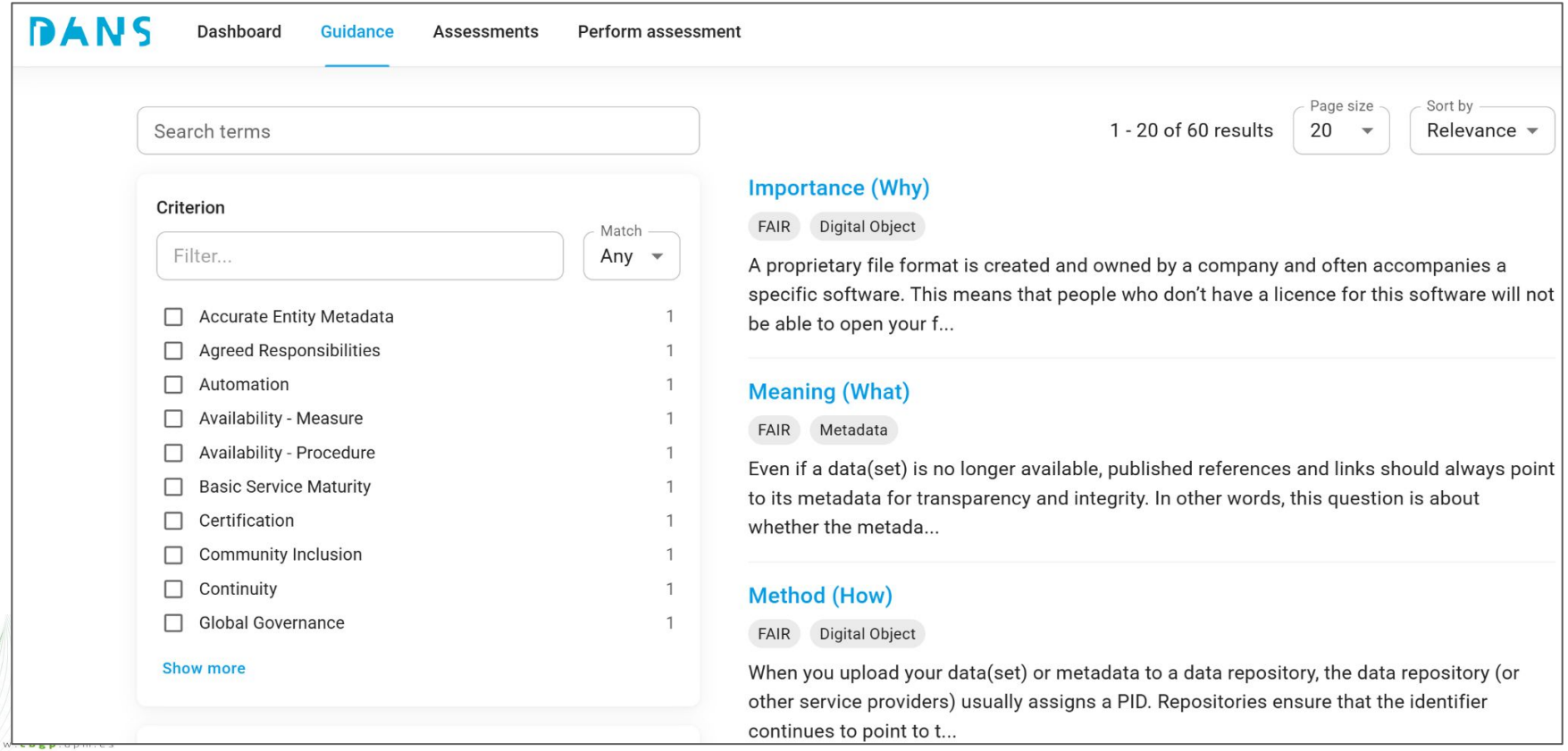
MOST IMPORTANT

Conclusions:

Condition	Discussion	Guidance
1	Unacceptable: metadata are not assigned persistent identifiers	Identifiers use ARK, DOI, HANDLE or URN PID schemas. More Information DANS FAIR Guidance. More Information
2	Acceptable: metadata are assigned globally unique identifiers	
3	Acceptable: data are described with structured metadata	
4	Acceptable: data are described with grounded metadata	
5	Unacceptable: metadata are not searchable	FAIR Principles F4. More Information DANS FAIR Guidance. More Information

Links to GUIDANCE
You aren't being judged, you're being helped!

“Guidance Elements”



Projects that put FAIR at the center

Programme:	HORIZON.2.1.5 - Tools, Technologies and Digital Solutions for Health and Care, including personalised medicine
Coordination:	INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, France
Budget:	146,000,000€
Consortium:	152 member organizations

FAIR technologies ~identical to those described for FLAIR GG

ERDERA as a service provider

- Onboarding
- Data FAIRification
- Data Management Plan creation
- Ontology recommendation and implementation

FIN


Sometimes, all a researcher needs is that one missing peace...
FAIR Data

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

www.cbgp.upm.es

Class: Thank You

Term IRI: http://purl.obolibrary.org/obo/NCIT_C94783

Definition: An expression of gratitude.

[Definition Source: NCI]

Dr. Mark D. Wilkinson
Dr. Pablo Alarcón
Dr. Bruno Cuevas
Sara Curiel

Contact me:
alberto.camara.ballesteros@alumnos.upm.es

ERDERA: Horizon Europe Proyecto HZ240050425A

Proyecto TED2021-130788B-I00 financiado por MCIN/AEI /10.13039/501100011033 y por la Unión Europea NextGenerationEU/ PRTR

POLITÉCNICA

y Tecnología Agraria y Alimentaria
CSIC
Consejo Superior de Investigaciones Científicas

MINISTERIO DE CIENCIA E INNOVACIÓN
Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
AGENCIA ESTATAL DE INVESTIGACIÓN