



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

EXCELENCIA
SEVERO
OCHOA



<https://tinyurl.com/UNBigData2024>



doi:10.5281/zenodo.11566326



POLITÉCNICA

INIA
Instituto Nacional de Investigación
y Tecnología Agraria y Alimentaria

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

www.cbgp.upm.es

FAIR

Strategies and technologies for early wins

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

mark.wilkinson@upm.es

*Keynote to #UNBigData2024
June 10, 2024*

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 exploration of global knowledge

IT WORKS! We can now prove it!

We have also experienced some notable failures...





Now with >13,500 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 by data providers

These behaviors can be concretely described

Software is then written to leverage these behaviors to find, access, and correctly reuse the data

“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

This means that FAIRness is, by definition,

measurable

by automated agents

The FAIR Evaluator

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

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



<http://w3id.org/AmIFAIR>

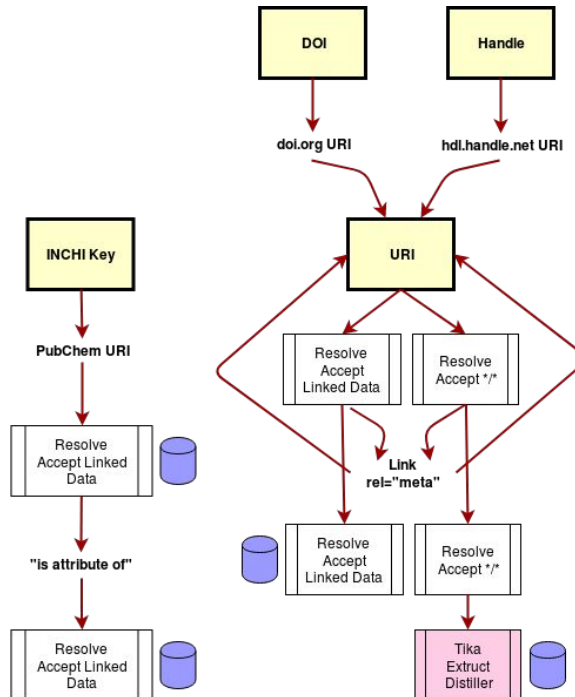


The screenshot shows the homepage of the FAIR Evaluation Services website. The header includes the site name 'FAIR Evaluation Services' and navigation links for 'HOME', 'EVALUATIONS', 'METRICS', and 'COLLECTIONS'. A search bar is located in the top right corner. The main content area features a large blue gradient banner with the title 'FAIR Evaluation Services' and the subtitle 'Resources and guidelines to assess the FAIRness of digital resources.' Below the banner, there are social media statistics: 'Fork', 'Star 46', and 'Watch 22'. The page is divided into three main sections, each with an icon and a 'Get started' button: 1. 'Import Maturity Indicators' (gear icon) with the description 'Import Maturity Indicators as YAML smartAPI interface annotation'. 2. 'Create collections' (list icon) with the description 'Assemble Maturity Indicators into community centered collections'. 3. 'Evaluate resources' (scales icon) with the description 'Evaluate resources FAIRness against Maturity Indicator Collections'.

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

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

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

(there are at least 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





FAIR Assessment a cottage industry!

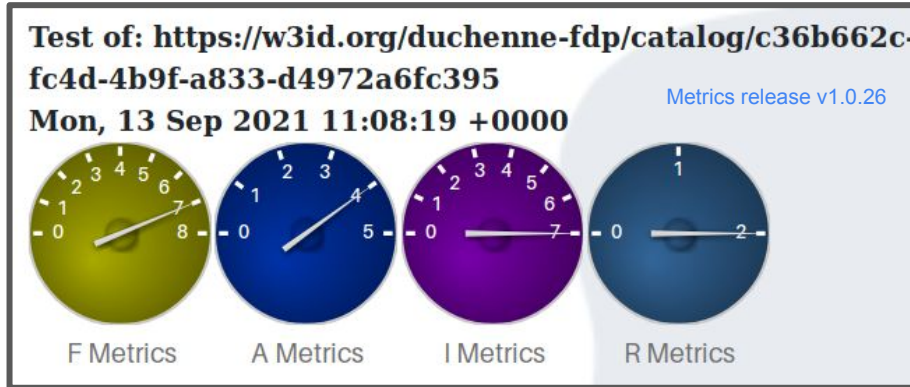
- 22 independent FAIR assessment platforms
- Most are questionnaire-based
- All of them focus on (primarily) metadata
- **But, there is a problem! The scores from each platform are different...**

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
FAIR Model	Manual - checklist

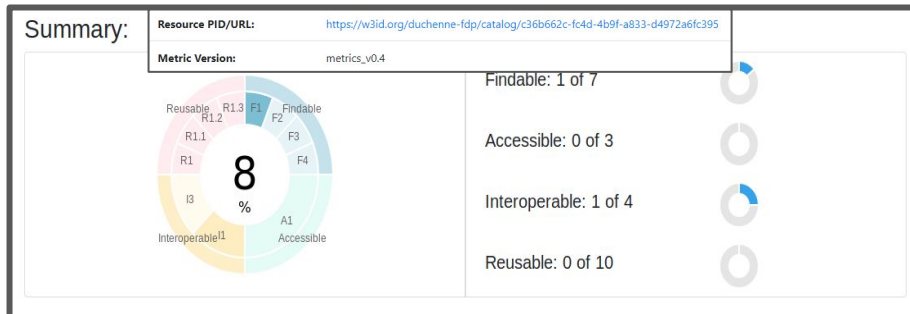
<https://fairassist.org>

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

The problem of metadata discovery and interpretation

EOSC calls for an investigation of the FAIR assessment discrepancies





EOSC Task Force on FAIR Metrics and Data Quality

Co-Chairs:

- Mark D Wilkinson
- Chris Schubert
- Carlo Lacagnina (retired)

Established November 2021

Chairs



Mark Wilkinson
UPM

Carlo Lacagnina
BSC

Members

Aguilar-Gómez, Fernando
CSIC

Al-Zoubi, Raed
ASREN

Bertino, Andrea
SWITCH

Biehlmaier, Oliver
Biozentrum, University of Basel

Cappiello, Cinzia
Politecnico di Milano

David, Romain
ERINHA AISBL

Dennis, Richard
Copenhagen University Library

Gingold, Arnaud
CNRS

Hajič, Jan
Charles University

Hecker, David
German Aerospace Center

Kleemola, Mari
CESSDA ERIC

Kuusniemi, Mari Elisa
OpenAIRE

Nikiforova, Anastasija
BBMRI

Nordling, Josefine
CSC

Papadopoulou, Eli
ATHENA RC

Sansone, Susanna-Assunta
University of Oxford

Schubert, Chris
TU Wien

Smit, Eefke
STM Association

Stryeck, Sarah
Graz University of Technology

Thiemann, Hannes
DKRZ

Velupillai, Sumithra
Swedish Research Council

von Stein, Ilona
DANS

Wright, Louise
EURAMET

Board Liaison



Sarah Jones
GÉANT

Outputs

Task Force charter



Three key outputs v.v. FAIR Testing



FAIR Metrics and Data Quality
Task Force

<https://doi.org/10.5281/zenodo.7463421>

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

Authorship Community:

Mark D Wilkinson^{1,3,*}, Susanna-Assunta Sansone^{2,4,*}, Marjan Grootveld^{2,5}, Josefine Nordling^{2,6},
Richard Dennis^{2,7}, David Hecker^{2,8} on behalf of the EOSC FAIR Metrics subgroup



FAIR Metrics and Data Quality
Task Force

<https://doi.org/10.5281/zenodo.10490289>

Report on "FAIR Signposting" and its uptake by the community

Mark D Wilkinson^{1,3}, Susanna-Assunta Sansone^{2,4}, Marjan Grootveld^{2,5}, Richard
Dennis^{2,6}, David Hecker^{2,7}, Robert Huber⁸, Stian Soiland-Reyes⁹, Herbert Van de
Sompel⁵, Andreas Czerniak¹⁰, Milo Thurston⁴, Allyson L. Lister⁴, Alban Gaignard¹¹



FAIR Metrics and Data Quality
Task Force

<https://doi.org/10.5281/zenodo.7390482>

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 Méndez⁵
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 Castro¹²

ostrails

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

Among its deliverables:

- Harmonization of FAIR Assessment tools, workflows, and APIs*
- Bootstrapping a global governance process for FAIR testing*

* These activities are open to non-project participants, so please contact me if you wish to become involved!

An example of a successful large-scale FAIRification initiative

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





+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)

EJP RD in numbers



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





+1800
people

EJP RD in numbers

91 beneficiaries
 10 hospitals
 12 research institutes

35 participating countries

26 EU MS, 7 associated (Algeria, Argentina, Brazil, Canada, Chile, Israel, Japan, Korea, Mexico, New Zealand, Norway, Oman, Portugal, Saudi Arabia, South Korea, Switzerland, Taiwan, Thailand, Turkey, United States)

ALL 24 EU countries

101 Budget

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

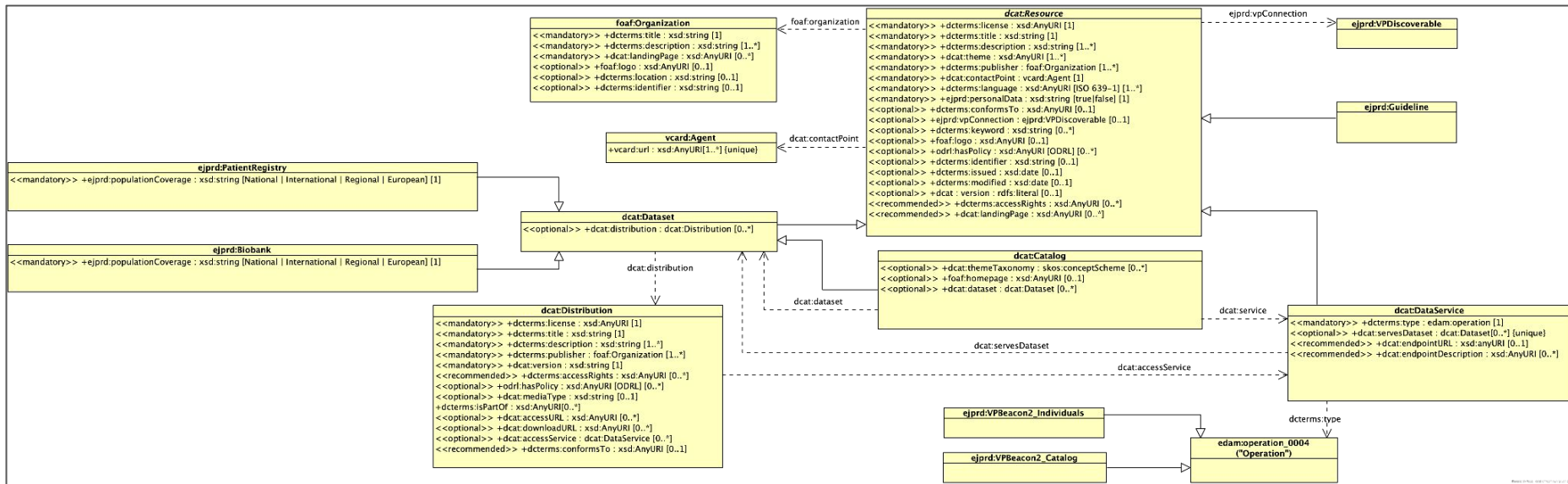
Challenge: Make it all FAIR

...but how?

funding bodies/ ministries
 hospital universities
 research centres
 foundations
 third parties
 associated networks



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 metadata for human exploration
- Publishes 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

Metadata for machines

Log in

FAIR Data Points

Filter:

All 26
Active 13
Inactive 5
Unreachable 1
Invalid 7
Unknown 0

Endpoint ▲ ▼	Registration ▲ ▼	Modification ▲ ▼	Status
https://w3id.org/ctsr-fdp/	15-12-2023, 06:00:00	08-06-2024, 06:00:00	ACTIVE
https://directory.bbMRI-eric.eu/api/fdp	04-10-2023, 16:12:15	07-06-2024, 14:00:00	ACTIVE
http://fairdatapointorphanet.info:7070	08-04-2024, 13:32:46	07-06-2024, 12:03:29	ACTIVE
http://45.88.81.224:7070/	19-03-2024, 10:46:34	07-06-2024, 11:46:36	ACTIVE
https://w3id.org/simpathic/fdp	11-07-2023, 13:23:57	05-06-2024, 13:23:57	ACTIVE
https://ejp-rd-fdp.ega-archive.org	02-03-2024, 11:53:09	05-06-2024, 12:53:06	ACTIVE
https://fair.ciroco.org	25-07-2023, 13:03:00	04-06-2024, 13:20:52	ACTIVE
https://w3id.org/fairvasc-fdp/	15-03-2023, 16:14:48	04-06-2024, 12:49:36	ACTIVE
https://w3id.org/duchenne-fdp	25-02-2023, 15:41:17	04-06-2024, 08:43:16	ACTIVE
https://fdp.wikipathways.org/index.ttl	27-02-2024, 22:34:01	04-06-2024, 03:02:14	ACTIVE

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

EUROPEAN JOINT PROGRAMME RARE DISEASES FAIR Data Point Metadata for machines

Search FAIR Data Point... Log in Advanced

FAIR Data Points

Filter: All 26 Active 13 Inactive 5 Unreachable 1 Invalid 7 Unknown 0

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

URL	Registration	Modification	Status
https://ejp-rd-fdp.ega-archive.org	02-03-2024, 11:53:09	05-06-2024, 12:53:06	ACTIVE
https://fair.ciroco.org	25-07-2023, 13:03:00	04-06-2024, 13:20:52	ACTIVE
https://w3id.org/fairvasc-fdp/	15-03-2023, 16:14:48	04-06-2024, 12:49:36	ACTIVE
https://w3id.org/duchenne-fdp	25-02-2023, 15:41:17	04-06-2024, 08:43:16	ACTIVE
https://fdp.wikipathways.org/index.ttl	27-02-2024, 22:34:01	04-06-2024, 03:02:14	ACTIVE

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



Challenge

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

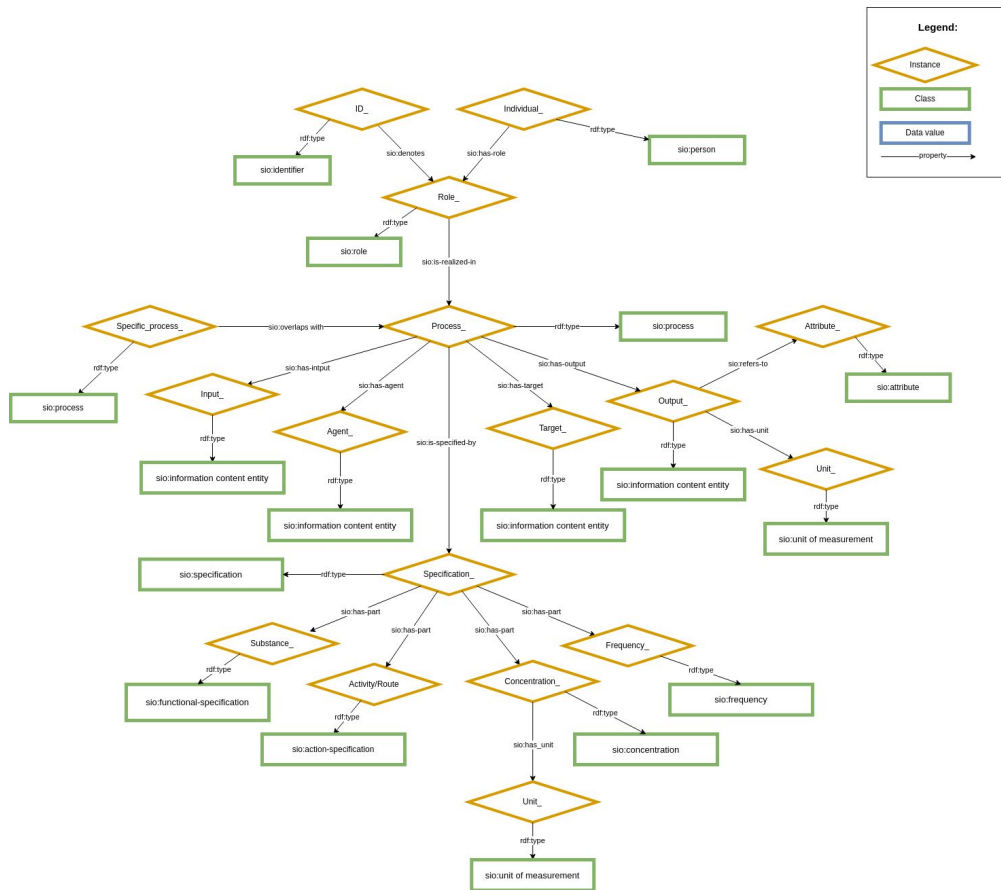
Partners have similar data (patient registry or biobank samples)

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 FAIR data model



Dr. Michel Dumontier,
UMaastricht,
SemanticScience
Integrated Ontology



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

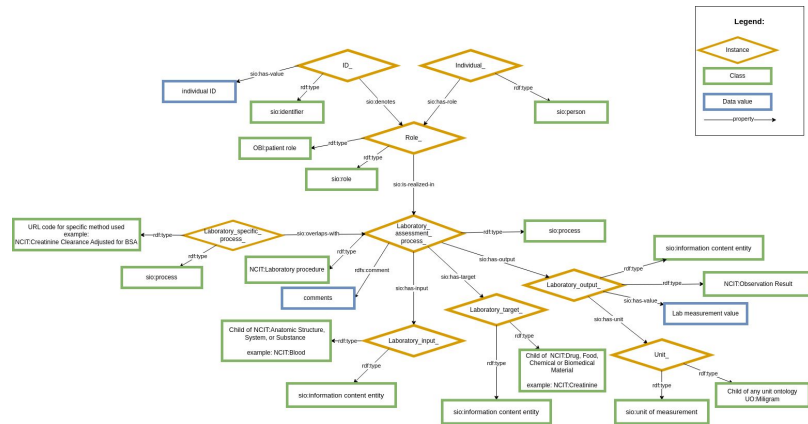
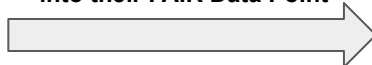


Step 3: Use CSV as a “lingua franca” for all partners

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?

The image shows five overlapping screenshots of the FAIR Data Point interface, each representing a different NMD registry:

- ER N EURO-NMD FAIR Data Point:** Shows the registry's description and a list of catalogs including EURO-NMD Registry Catalog, Top-level Catalog, and Top level metadata catalog.
- Duchenne Data Platform:** Describes the platform as a patient registry and lists distributions.
- CRAMP FAIR Data Point:** Describes it as an exemplar FAIR Data Point for the CRAMP Registry.
- DMScope FDP:** Describes it as an exemplar FAIR Data Point.
- SMARTCARE FAIR Data Point:** Describes it as a joint initiative of neurologists, pediatricians, and patient organizations, and includes a table of metadata.

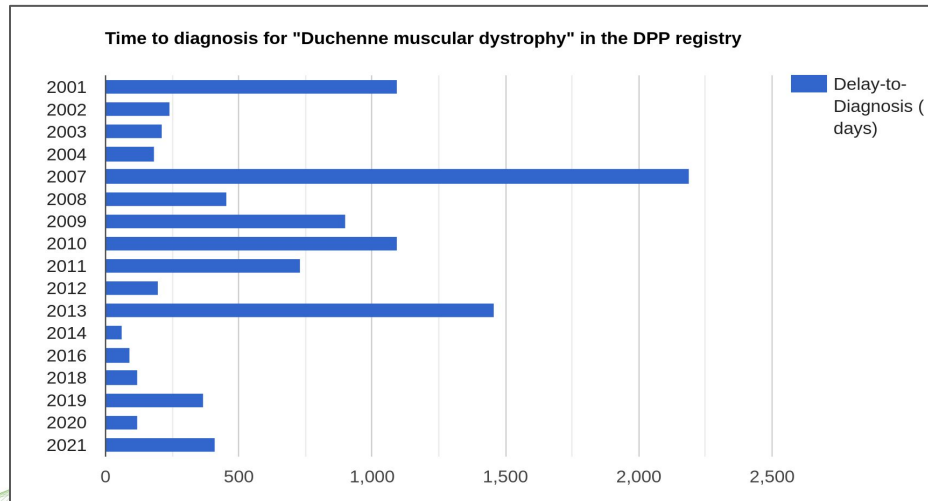
Metadata	Issued	Modified
Version	1.0	
Language	en	
Version	4.0	
Metadata Identifier	Identifier	
Metadata issued	27-02-2023	
Metadata modified		27-02-2023

Five FAIR Data Points for partners representing NMDs

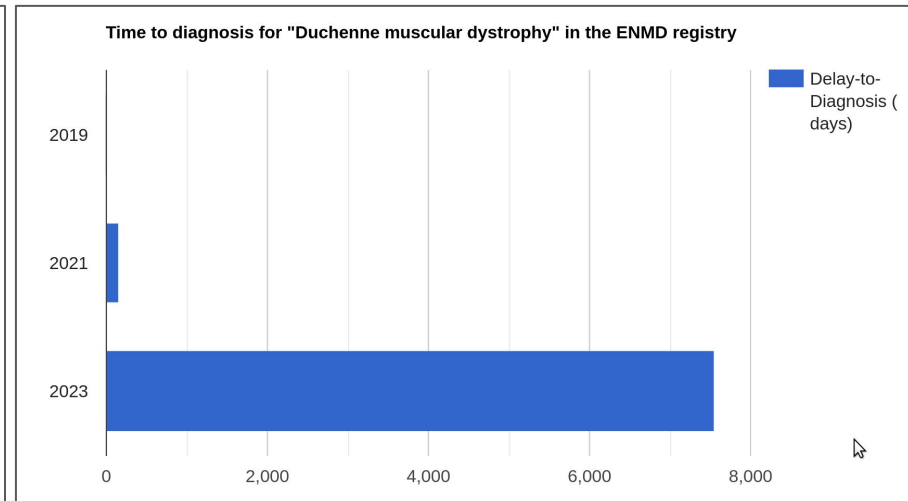
“What is the delay between symptom onset and diagnosis?”

- 1) Metadata allowed automated agent to discover and interact with participants capable of providing question-relevant data
- 2) The same query sent to all participants (shared model)
- 3) Integrate the output

Duchenne Parent Project (DPP)



EURO-NMD (mock data)



Caveat emptor!!

FAIR Data alone is NOT sufficient to
achieve interoperability!



This paper compares two independent FAIRification efforts over identical data items

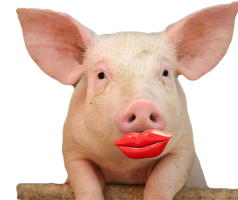


Journal of the Society for
Clinical Data Management

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 Braunt†, Emily Hartley†, Daniel Olson†, Nirupama Benis‡, Ronald Cornet‡, Mark D. Wilkinson* and Ramona L. Walls†

Interoperability was almost zero!
Shared models are necessary

Segue into the panel session
coming up next

TED2021-130788B-I00



FLAIR-GG

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

Dr. Santiago Moreno Vasquez

Dr. Mark D. Wilkinson

Oussama Mohammed Benhamed, PhD Candidate

Alberto Camara Ballesteros, PhD Candidate





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

The EURISCO Web catalogue automatically receives data from the European National Inventories (NI). It provides information at the accession level of PGR conserved in European genebanks or other collections. EURISCO is hosted at and maintained by [IPK Gatersleben](#) on behalf of the Secretariat. [Click here for further information and access.](#)

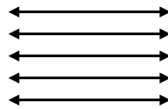


- **Genesys** (global portal to information about Plant Genetic Resources for FA)
- **GLIS** (Global Information System for PGRFA)



USERS

Breeders
 Scientists
 Producers
 Taxonomists
 Archaeologists
 Etc.



National focal points
(European National Inventories)



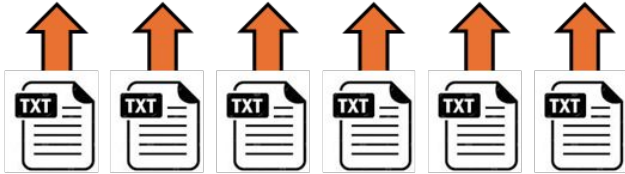
National Genebanks Networks
(Crops conserved *ex-situ*)

EURISCO
(European Search Catalogue for Plant Genetic Resources)



National focal points

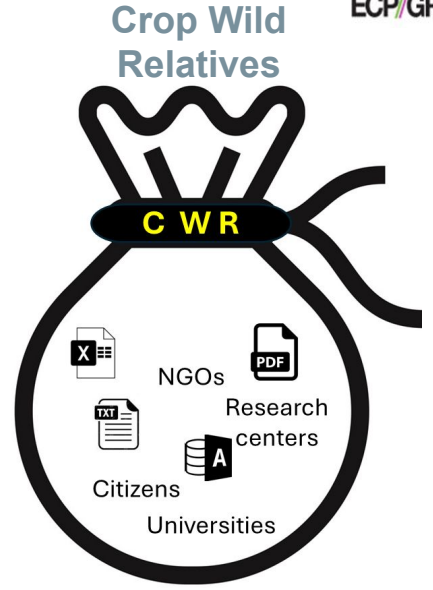
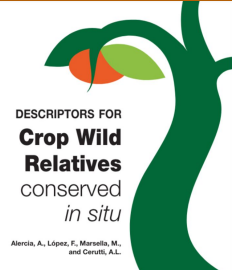
(European National Inventories on PGR)



National Genebanks Networks
 (Crops conserved *ex-situ*)



- *environmental data
- *legal regulations on territories
- *legal regulations on species
- *conservation status
- *research data on species



A plethora of new stakeholders with highly varying technical experience & varying levels of compliance with emergent FAO descriptors

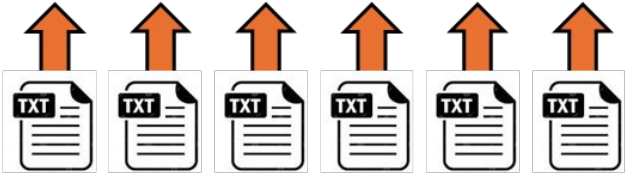
This is identical to the situation faced by the registries and biobanks in EJP-RD!

EURISCO
(European Search Catalogue for Plant Genetic Resources)



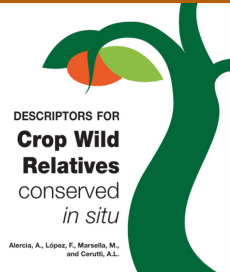
National focal points

(European National Inventories on PGR)



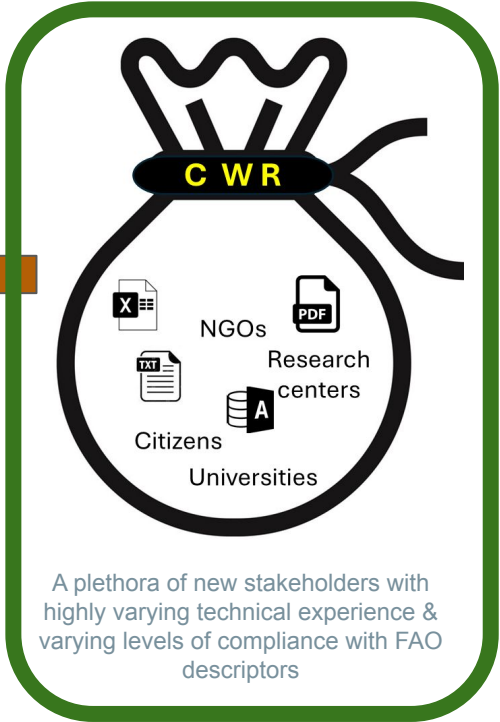
National Genebanks Networks
(Crops conserved *ex-situ*)

*environmental data
 *legal regulations on territories
 *legal regulations on species
 *conservation status
 *research data on species



Alencia, A., López, E., Marotta, M. and Cerutti, A.L.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
 on behalf of
 THE INTERNATIONAL TREATY ON PLANT GENETIC RESOURCES
 FOR FOOD AND AGRICULTURE
 Rome, 2022 Revised version



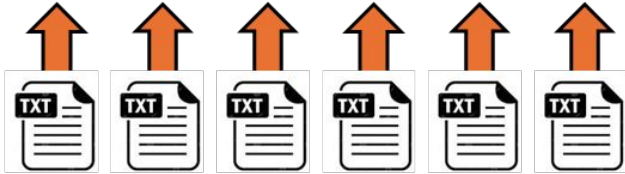
FAIR Data facilitates communication between partners **within** the CWR community, and **should also improve communication with other indexes** such as EURISCO

EURISCO
(European Search Catalogue for Plant Genetic Resources)



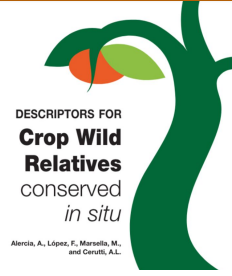
National focal points

(European National Inventories on PGR)

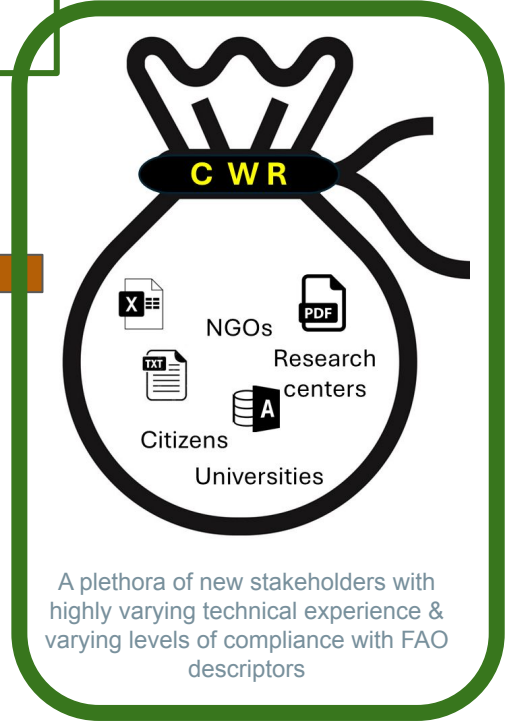


National Genebanks Networks
(Crops conserved *ex-situ*)

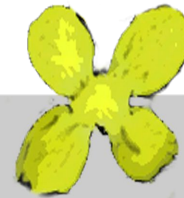
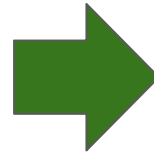
- *environmental data
- *legal regulations on territories
- *legal regulations on species
- *conservation status
- *research data on species



FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
 on behalf of
 THE INTERNATIONAL TREATY ON PLANT GENETIC RESOURCES
 FOR FOOD AND AGRICULTURE
 Rome, 2022 Revised version



FLAIR-GG Objectives



BGV-UPM

"César Gómez Campo"

Replicate EJP-RDs success with
BGV as our first target

...then expand!



BGV FAIR Data Point

Metadata of the Banco de Germoplasma Vegetal de la UPM

Datasets Services

Datasets

Administrative data from the BGV

Information about the institute and/or collection team responsible for the germplasm deposit

Administrative Contact Institution

Issued 03-11-2023 Modified 01-06-2024 Keywords Administrative

BGV June 2023

Metadata snapshot of BGV taken in June 2023

Issued 28-06-2023 Modified 01-06-2024 Keywords

Location Information

Geolocation information for the germplasm deposit. This will include features such as country name/abbreviations, latitude and longitude, and soil conditions at that position.

Collection site Environmental Geolocation Soil

Issued 03-11-2023 Modified 01-06-2024 Keywords Collection site

Services

BGV FAIR Data Point Metadata SPARQL server

Metadata Service serving the DCAT for the César Gómez Campo Banco de Germoplasma Vegetal de la UPM

format_3790


Issued 28-12-2023 Modified 28-12-2023 Keywords

FLAIR-GG Status

FAIR Data Point:

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

FLAIR-GG Status



BGV FAIR Data Point

Metadata of the Banco de Germoplasma Vegetal de la UPM

Datasets
Services

Datasets

Administrative data from the BGV

Information about the institute and/or collection team responsible for the deposit

Administrative
Contact
Institution

Issued 03-11-2023 **Modified** 01-06-2024 **Keywords** Administrative

BGV June 2023

Metadata snapshot of BGV taken in June 2023

Issued 28-06-2023 **Modified** 01-06-2024 **Keywords**

Location Information

Geolocation information for the germplasm deposit. This will include country name/abbreviations, latitude and longitude, and soil conditions

Collection site
Environmental
Geolocation
Soil

Issued 03-11-2023 **Modified** 01-06-2024 **Keywords** Collection site

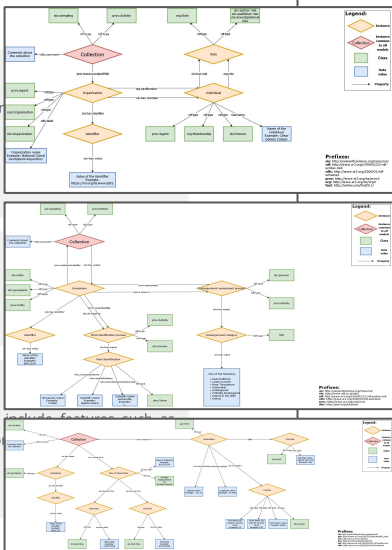
Services

BGV FAIR Data Point Metadata SPARQL server

Metadata Service serving the DCAT for the César Gómez Campo Banco de Germoplasma Vegetal de la UPM

format_3790

Issued 28-12-2023 **Modified** 28-12-2023 **Keywords**



The diagrams illustrate the data models and templates for BGV FAIR Data Point. Each diagram shows a hierarchical structure of data elements, with a legend and a list of prefixes.

Legend:

- Red diamond: Class
- Yellow diamond: Property
- Green rectangle: Instance
- Blue rectangle: Property

Prefixes:

- dc: http://purl.org/dc/terms/
- dcat: http://www.w3.org/ns/dcat#
- prov: http://www.w3.org/ns/prov#
- rdfs: http://www.w3.org/2000/01/rdf-schema#
- skos: http://www.w3.org/2004/02/skos/core#
- time: http://www.w3.org/2006/time#
- vann: http://www.w3.org/ns/vann/
- vocab: http://www.w3.org/ns/vocab/
- xsd: http://www.w3.org/2001/XMLSchema#

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.

www.cbgp.upm.es

FLAIR-GG Status

FAIR Data Point
Metadata for machines

Search FAIR Data Point... Log in

Advanced

FAIR Data Points


Filter: All 1 Active 1 Inactive 0 Unreachable 0 Invalid 0 Unknown 0

Endpoint ▲ ▼	Registration ▲ ▼	Modification ▲ ▼	Status
https://w3id.org/bgv-fdp	24-07-2023, 14:04:50	06-06-2024, 14:27:21	ACTIVE

FAIR Data Point Federated Partner Index:

- So far... we're quite lonely!
- Several seedbank* partners identified who should be onboard within a few weeks
- Everyone is welcome to participate!!
Contact me!

FLAIR-GG Status



FLAIR-GG

Connecting Germplasm Resources

Virtual Platform Resources: All Resources

Catalogs (1)
 Datasets (3)
 Distributions (0)

Services (4)
 Other (0)

SOURCE: <http://www.bancodegermoplasma.upm.es>

- Resource: [BGV Germplasm SPARQL Endpoint](#)
- Resource: [Administrative SPARQL Endpoint](#)
- Resource: [SPARQL Endpoint for Location data of BGV](#)
- Resource: [BGV FAIR Data Point Metadata SPARQL server](#)




Keyword Search:

Ontology URI:

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

© 2023/2024 Mark D Wilkinson

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

FLAIR-GG “Virtual Platform”

- 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)

FLAIR-GG Next Steps

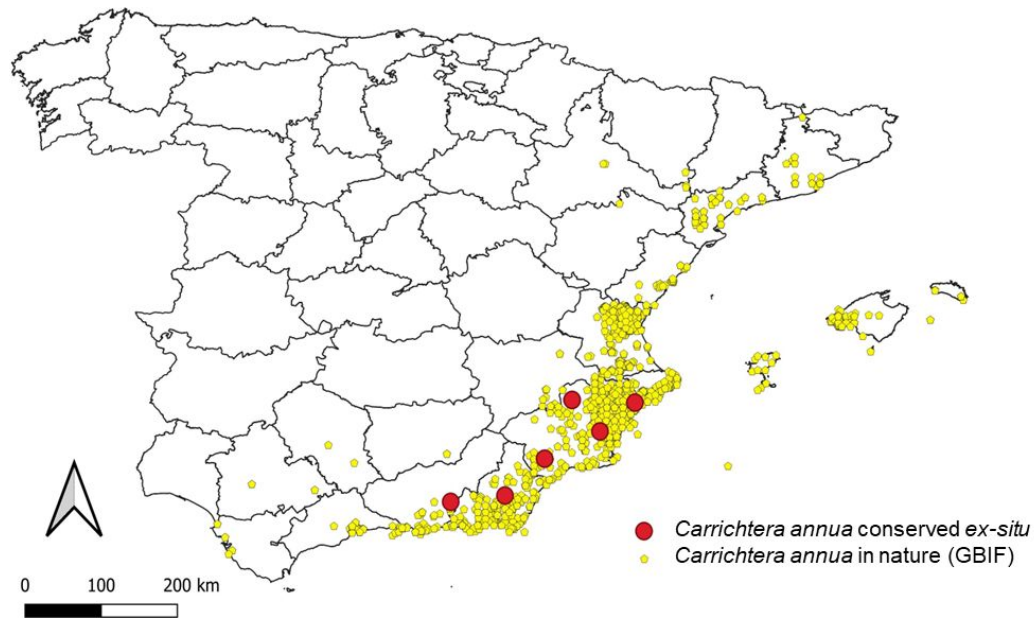
Begin constructing **shared queries** to help
build data-driven conservation strategies

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?”

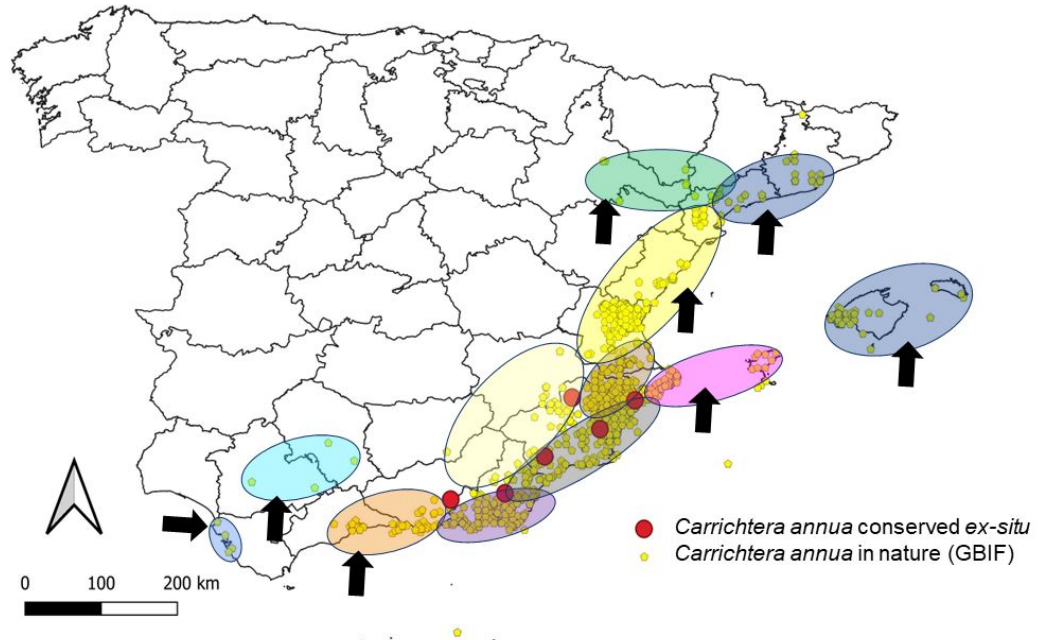


FLAIR-GG Next Steps

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?”



Take Home Messages

FAIR is metadata first!!!!

FAIR is measurable, but we need global governance of testing before agencies can trust FAIR quality assessments

FAIR Data does not, alone, lead to interoperability - shared models required!

Rich, high-quality metadata enhances the appropriate reuse of FAIR data

Technologies/strategies allow FAIR experts to assist data owners in creating FAIR data themselves → distribute the effort, rather than centralize/warehouse

Don't reinvent wheels - projects like EJP-RD and FLAIR-GG have generated a **mountain of reusable code and models** for FAIRification

Acknowledgements

<https://tinyurl.com/UNBigData2024>



My numerous and treasured collaborators and co-authors have been cited *in situ* throughout this slide deck



BGV-UPM

"César Gómez Campo"



TED2021-130788B-I00

