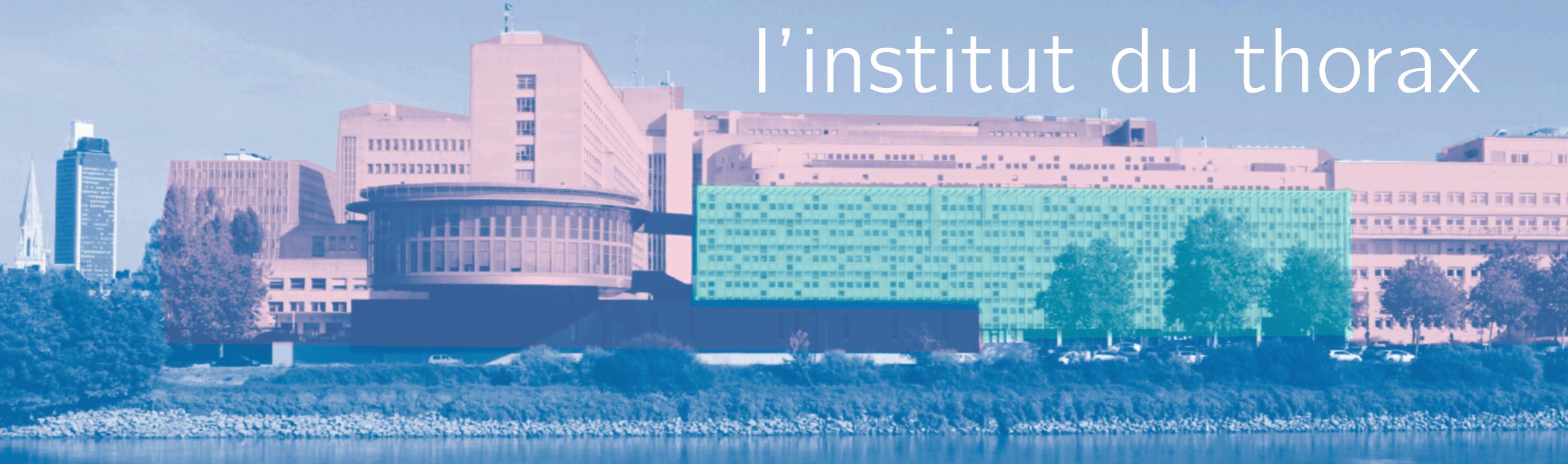


Beyond FAIR assessment black boxes: ontology-based evaluations and the challenge of comparing FAIR scores

Alban Gaignard
CNRS, institut du thorax, Nantes, France

October 20, 2024
CMLS 2025 - Poitiers, France

Context



Better understanding of cardio-vascular and metabolic diseases

Gene-function associations

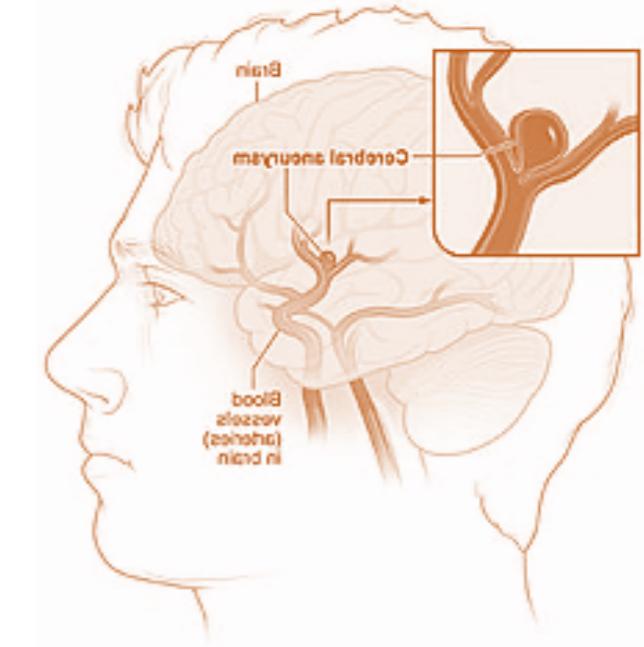
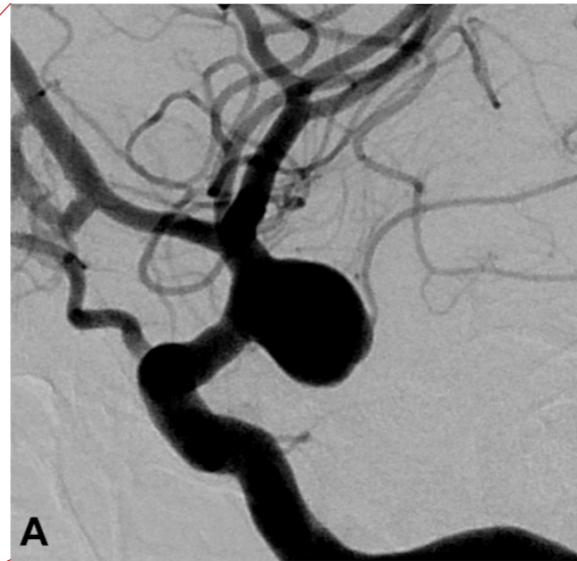
Translational medicine

Real proximity between the university hospital and the research lab

Bioinformatics

- ▶ Massive production of genomic sequence & health data
→ high performance computing
- ▶ Integration of multi-modal and multi-scale data
- ▶ Predictive models

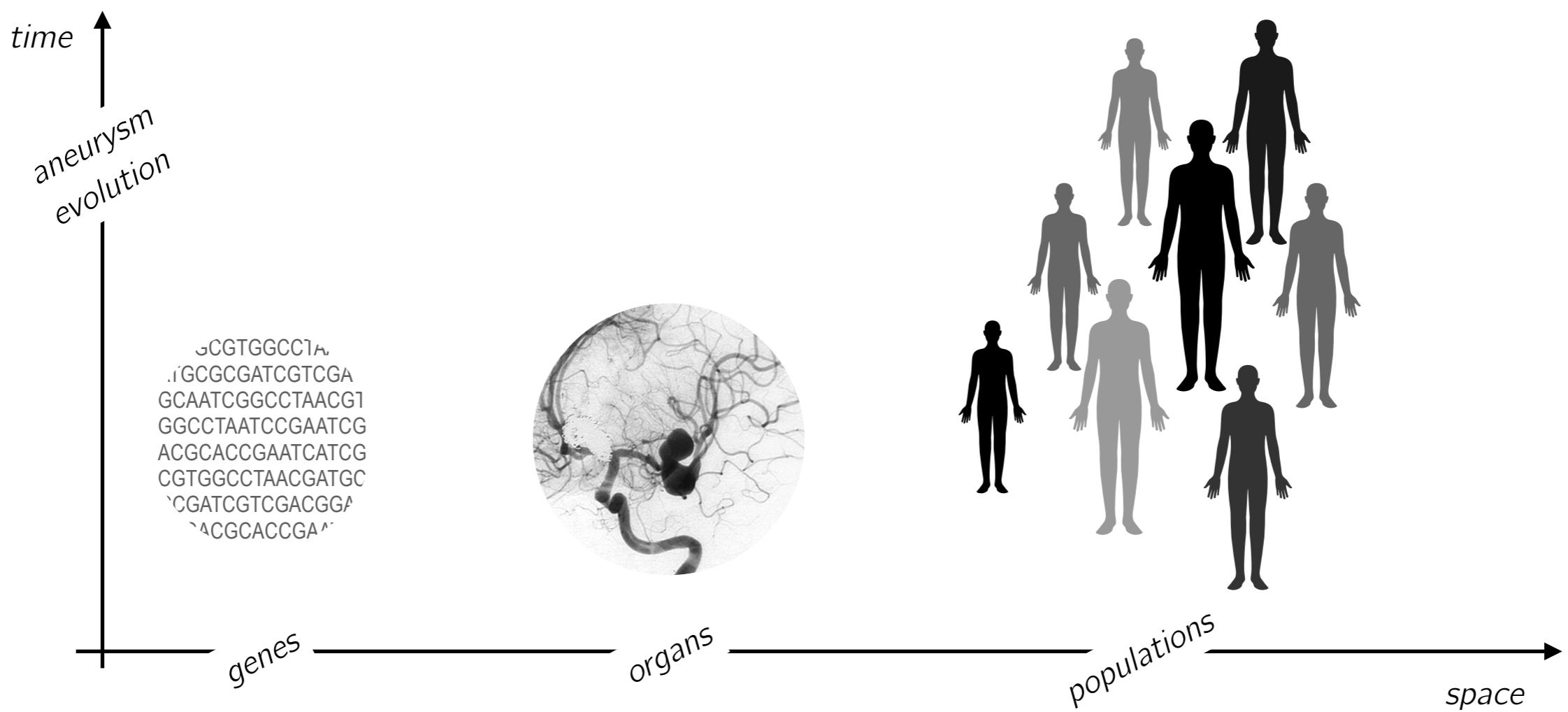
Intracranial aneurysm (ICA)



- ▶ 3% of the general population
- ▶ unpredictable rupture
- ▶ 50% of death in case of rupture

Multi-factorial disease → multi-scale data

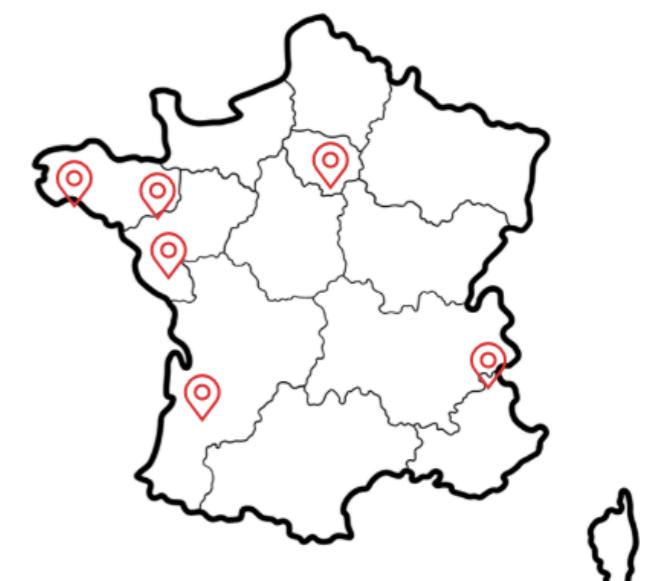
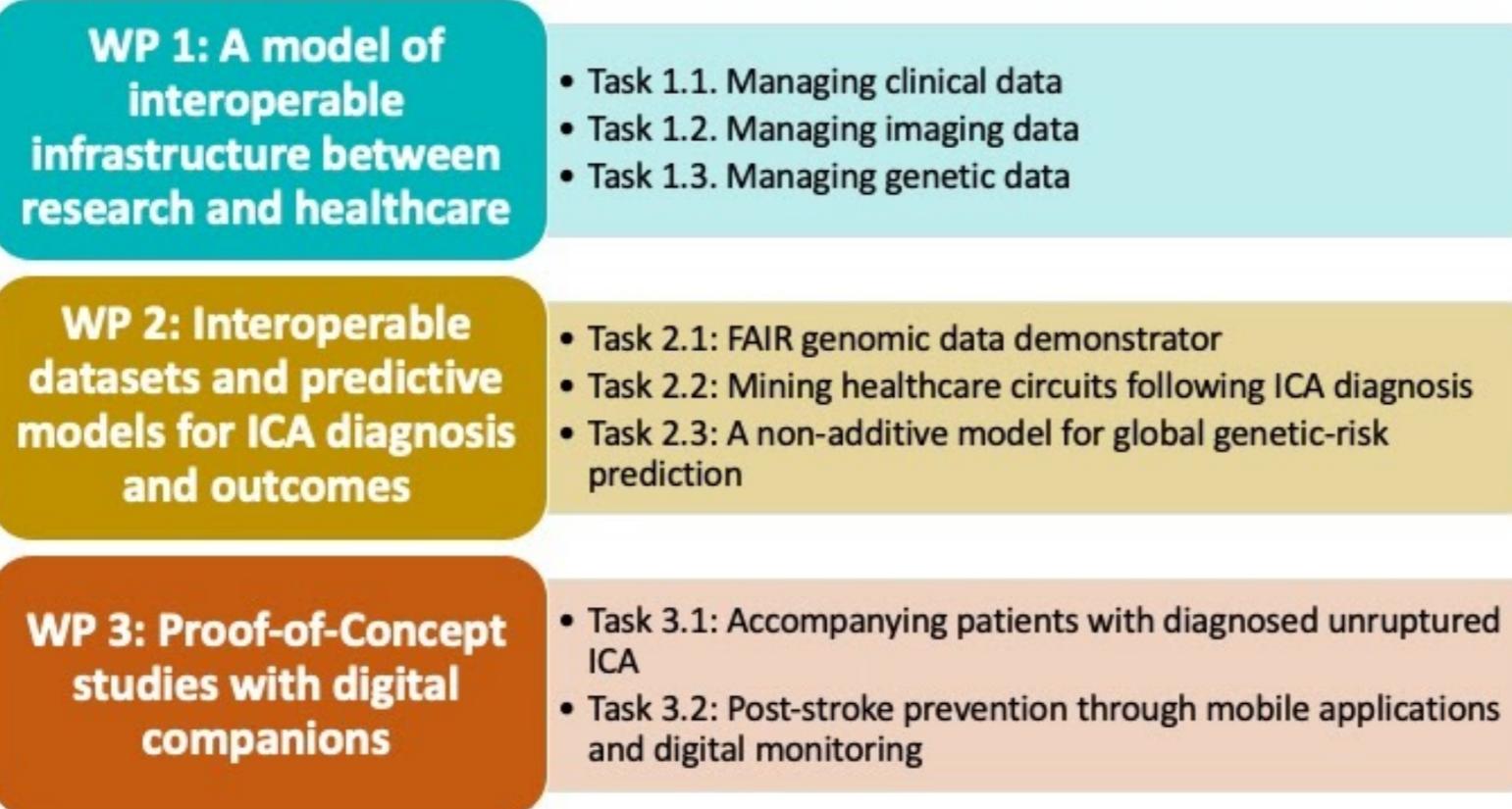
- ▶ Inter-disciplinary efforts needed for a better understanding of the pathology
- ▶ Specific data produced at very specific scales



Neurovasc project

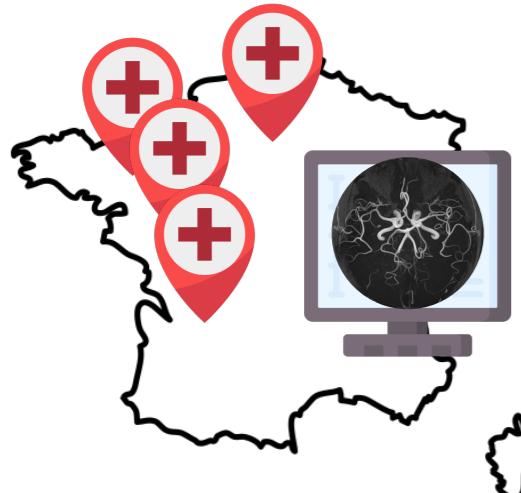


- ▶ Neurovasc: a national programme funded for 4 years by the french research agency to build a **digital infrastructure** to manage and exploit **intracranial aneurysm data**
 - 3 Research Institutes (Inria, Inserm, IMT Atlantique)
 - 2 Clinical Research Teams (Brest & Nantes academic hospitals)
 - 3 Universities (Bordeaux, Paris-Saclay, Nantes)



Data integration & sharing challenges

① How to **collect high-quality medical images** from multiple hospitals/ MRIs ?



② How to **interlink and query multi-modal and multi-scale data** while preserving privacy constraints ?



③ How to mine and **model patient trajectories** from EHR data ? can we predict clinical outcomes ?

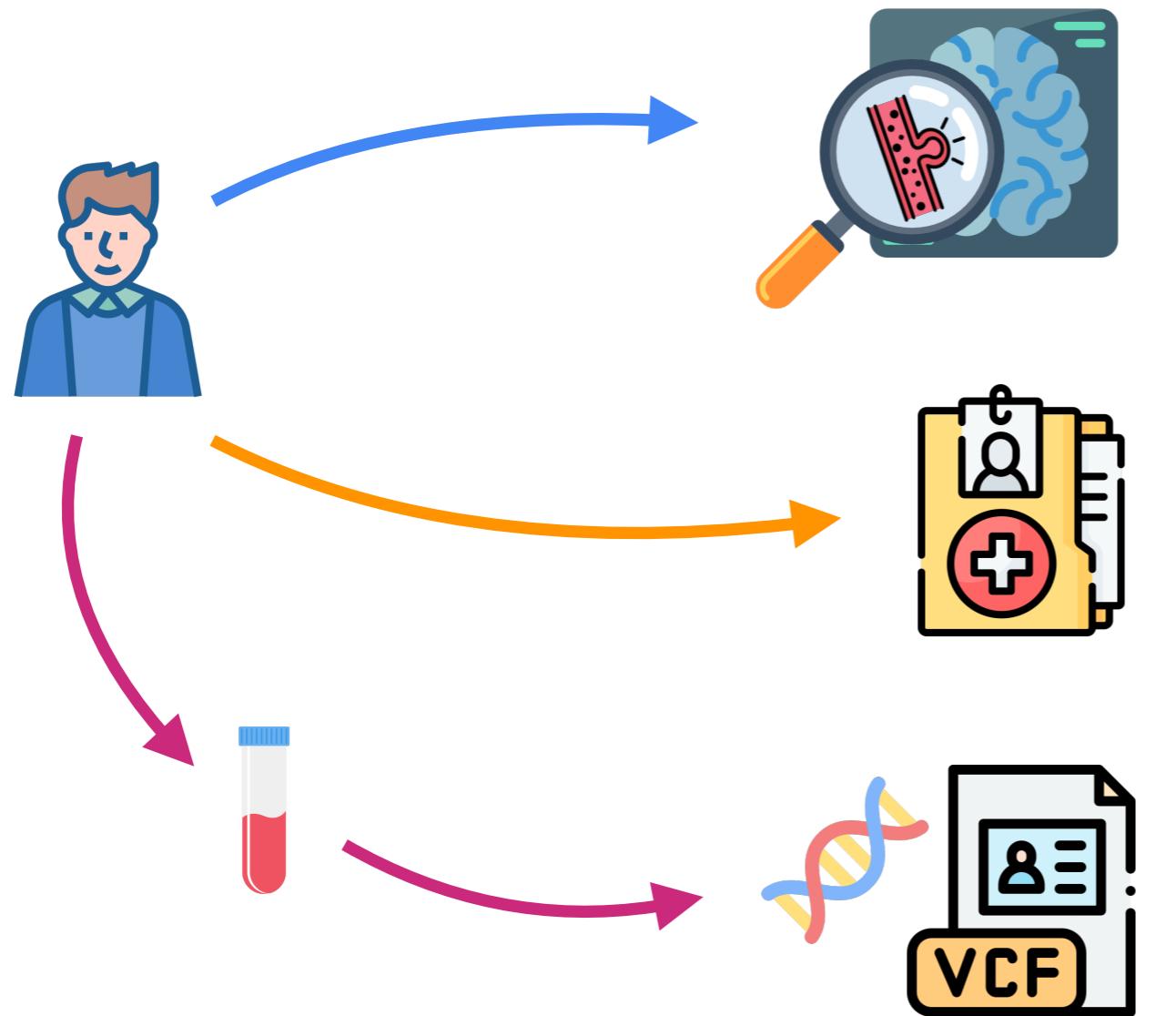


We need FAIR data !

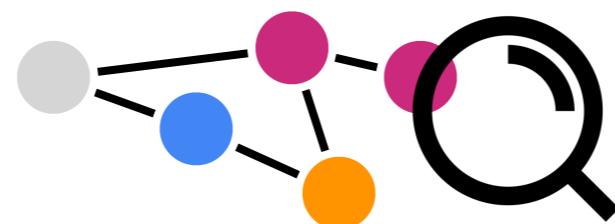


FAIRifying clinical & genomic data

Alexandrina
Bodrug



A clinical and genomic intracranial aneurysm knowledge graph



to find & exchange phenotypes/variants with reference terminologies !



Marie-Dominique Devignes



Thomas Rosnet



Frédéric de Lamotte

The screenshot shows the journal's header with 'Journal of Biomedical Semantics' and navigation links for Home, About, Articles, Collections, and Submission Guidelines. A blue button says 'Submit manuscript'. Below the header, it says 'Software | Open access | Published: 01 July 2023'. The main title of the article is 'FAIR-Checker: supporting digital resource findability and reuse with Knowledge Graphs and Semantic Web standards'. It lists authors: Alban Gaignard, Thomas Rosnet, Frédéric De Lamotte, Vincent Lefort, & Marie-Dominique Devignes. It also shows the journal name 'Journal of Biomedical Semantics', volume '14', article number '7 (2023)', and citation information: '5024 Accesses | 31 Citations | 3 Altmetric | Metrics'. There is a small logo in the top left corner of the screenshot.

How can we evaluate that research data are FAIR enough ?

FAIR principles require tooling



Australian Research Data Commons

FAIR principles

- ▶ critical for open & reproducible sciences
- ▶ result in many guidelines
- ▶ technology agnostic guidelines

How to implement the principles ...
... and go beyond checklists ?

Web URIs HTTP, RDF/SPARQL

Domain ontologies

**Resource provider
and developers
need help and tooling.**

Usage scenarios



Data producers



- Dataset
- Training
- Software tools
- etc ...

discoverable !

Where to publish ?

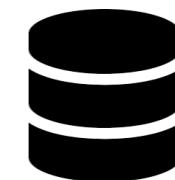
Which registry ?

Does it provide metadata ?

Is it enough to be FAIR ?



Registry developers



Make

more FAIR !

- Dataverse
- Bio.tools
- Zenodo,
- Pubmed ...

Improve metadata quality ?

Community specific
standards ?

Which technology ?

Why a (nother) tool ?

Assumptions

- ▶ "Linked Data" and Semantic Web technologies are key in most of the FAIR principles (especially F, I, and R)
... but technical skills are needed.

Objectives

- ▶ Provide a web interface for resource providers to evaluate FAIR metrics and make progress on FAIRification (iterative testing)
- ▶ Provide additional tools for developers Leverage semantic web technologies (RDF, SPARQL, SHACL) to enhance the quality of metadata
- ▶ Share implementations, and contribute to larger FAIR evaluation initiatives



Approach

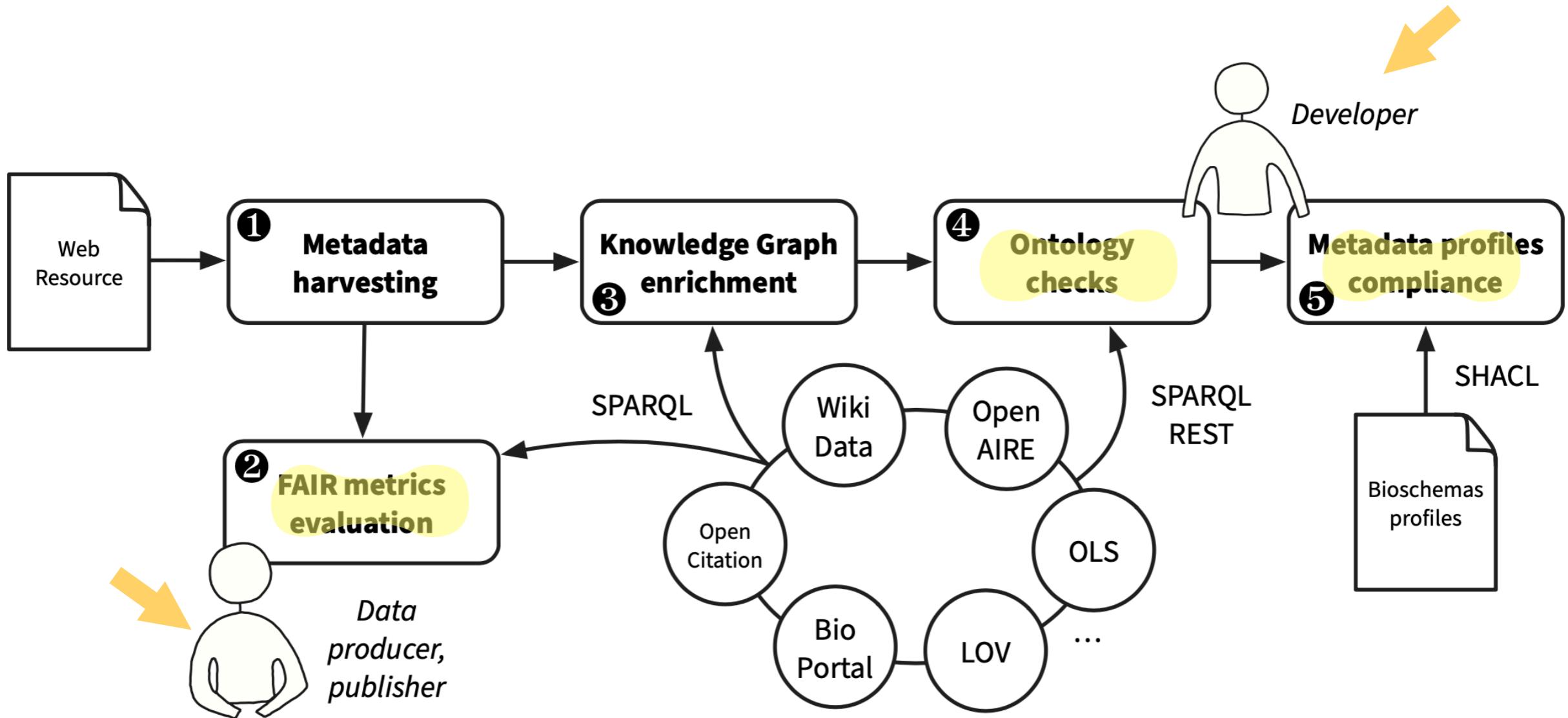
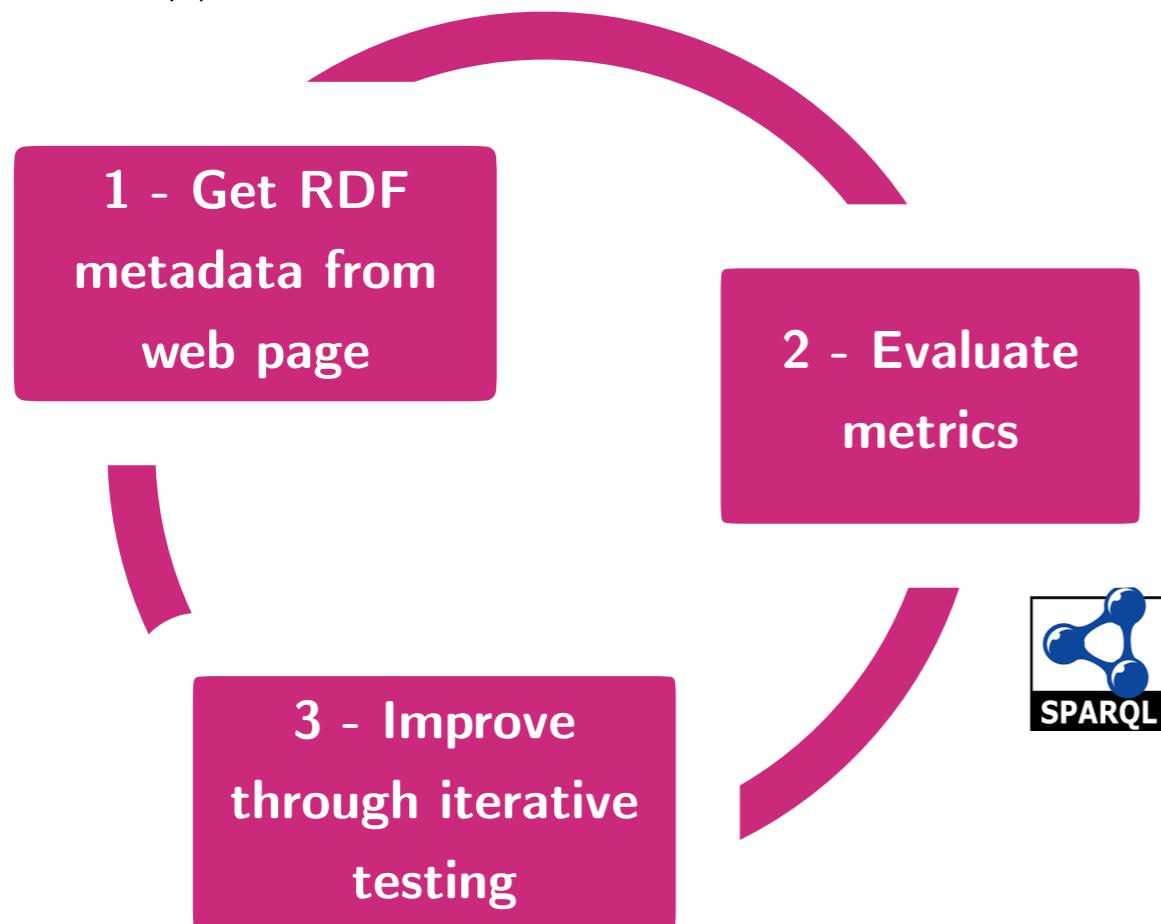


Figure 1 Gathering, enriching and analyzing semantic web annotations in line with FAIR principles.

① Checking metadata

Checking metadata

http://...



Semantic web technologies

- ▶ RDF
- ▶ SPARQL



 FAIR cookbook

 RDMkit

<https://fairplus.github.io/the-fair-cookbook/content/home.html>

<https://rdmkit.elixir-europe.org>

What is evaluated ?

Findability F1B, F2	Accessibility A1.2	Reuse (licenses) R1.1	Reuse (provenance) R1.2
dct:identifier	odrl:hasPolicy	schema:license	prov:wasGeneratedBy
schema:identifier	dct:rights	dct:license	prov:wasDerivedFrom
dct:title	dct:accessRights	doap:license	prov:wasAttributedTo
dct:description		dbo:license	prov:used
dcat:accessURL		cc:license	prov:wasInformedBy
dcat:downloadURL		xhv:license	prov:wasAssociatedWith
dcat:endpointDescription		sto:license	prov:startedAtTime
dcat:endpointURL		nie:license	prov:endedAtTime
			dct:hasVersion
			dct:isVersionOf
			dct:creator
			dct:contributor
			dct:publisher
			pav:hasVersion
			pav:version
			pav:hasCurrentVersion
			pav:createdBy
			pav:authoredBy
			pav:retrievedFrom
			pav:importedFrom
			pav:createdWith
			pav:retrievedBy
			pav:importedBy
			pav:curatedBy
			pav:createdAt
			pav:previousVersion
			schema:creator
			schema:author
			schema:publisher
			schema:provider
			schema:funder

DC-Terms
DCAT
Schema.org

+ ODRL

+ DOAP,
DBO, CC ...

+ PROV-O, PAV

Table 2 Summary of the selected ontology properties relevant to assess three specific FAIR principles in *FAIR-Checker*

Resource identifier (URL/DOI)

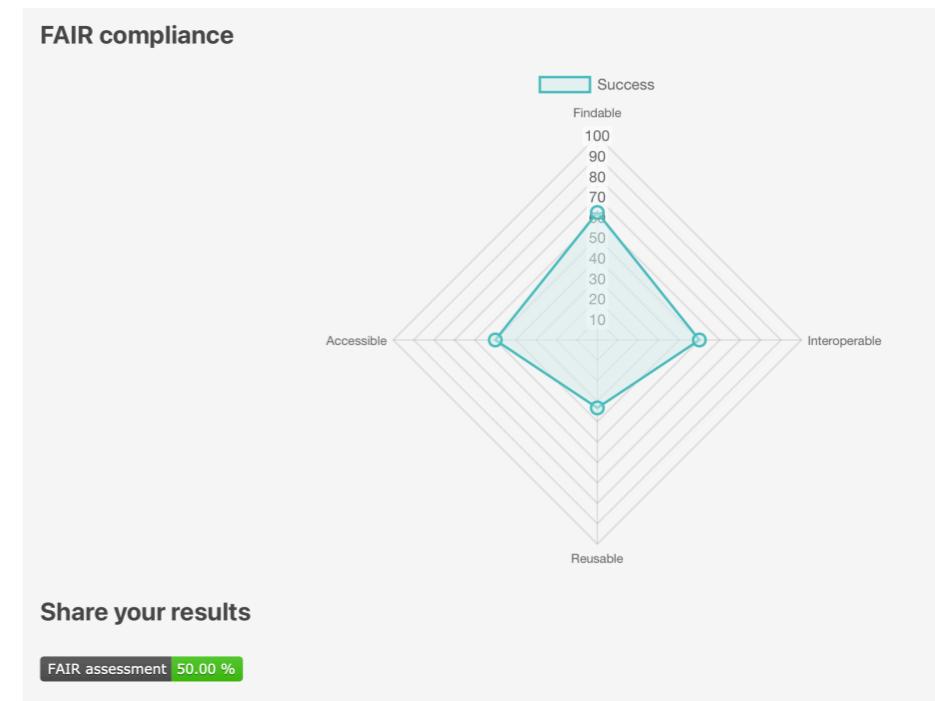
https://commons.datacite.org/doi.org/10.3205/09dgnc137 ✓

All metrics

Valid URL/DOI – The input contains the following DOIs that you can also test: 10.3205/09dgnc137

Clean results

Dataset Dataverse Workflow Publication Datacite Dataset Tool



F2B: Shared vocabularies for metadata	<button>Check</button>	FAIR principle F2B 2/2		<button>i</button>
A1.1: Open resolution protocol	<button>Check</button>	FAIR principle A1.1 2/2		<button>i</button>
A1.2: Authorisation procedure or access rights	<button>Check</button>	FAIR principle A1.2 0/2	You should describe the access policy in metadata by using at least one of the Read more ▾	<button>i</button>
I1: Machine readable format	<button>Check</button>	FAIR principle I1 1/2	You should provide discoverability oriented metadata with one of the following properties: dct:title dct:description dcat:accessURL dcat:downloadURL dcat:endpointDescription dcat:endpointURL Read less ▲	<button>i</button>
I2: Use shared ontologies	<button>Check</button>	FAIR principle I2 2/2		<button>i</button>
I3: External links	<button>Check</button>	FAIR principle I3 0/2	You should enrich your metadata with more diversified external links. Here we did not Read more ▾	<button>i</button>
R1.1: Metadata includes license	<button>Check</button>	FAIR principle R1.1 0/2	You should include information about licence in your metadata using one of the Read more ▾	<button>i</button>

1. Provide resource URL
2. Retrieve metadata
3. Evaluate relevant metadata fields (ontology classes and properties)
4. Propose **relevant recommendations** from FAIR Cookbook and RDMkit
5. Improve metadata and **try again ...**

② Inspecting metadata

Approach

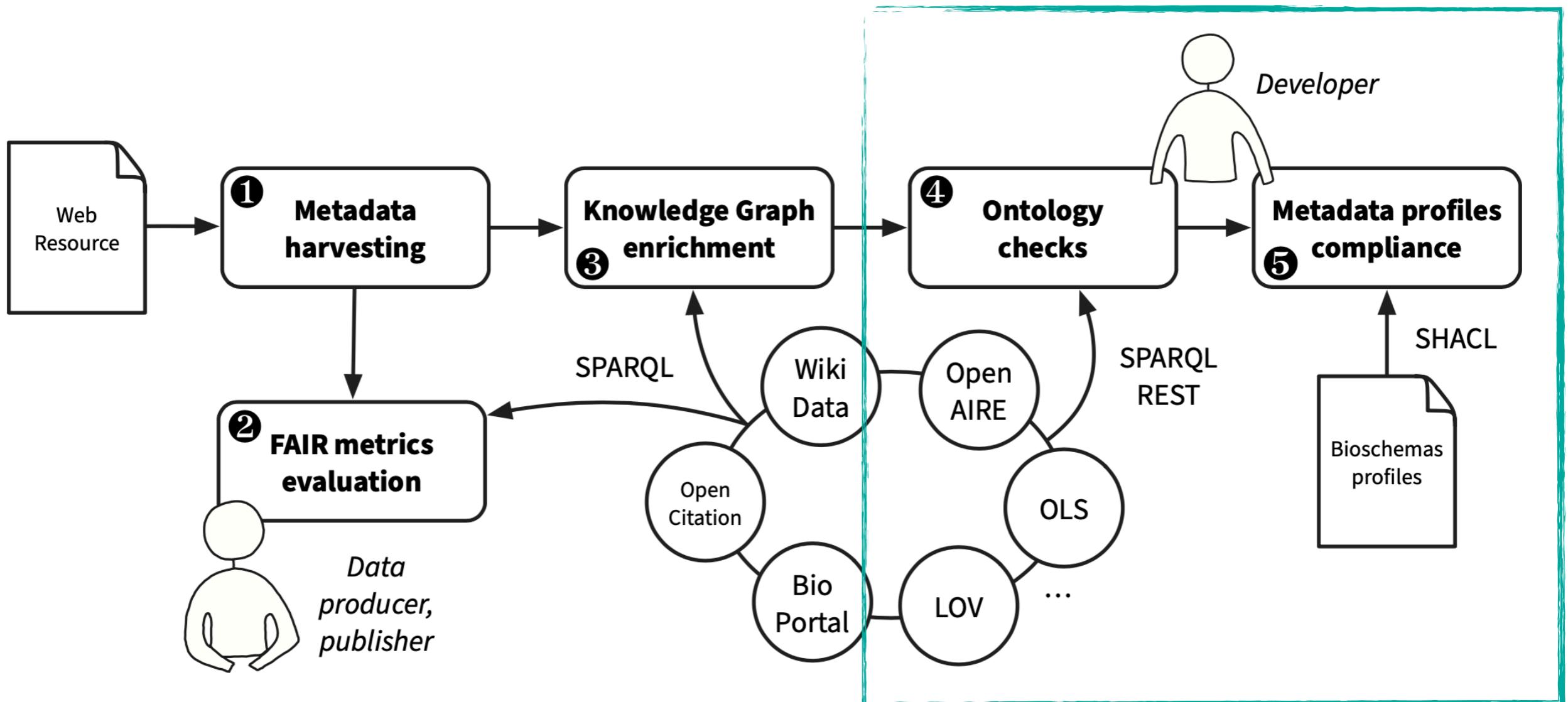


Figure 1 Gathering, enriching and analyzing semantic web annotations in line with FAIR principles.

Full Hierarchy

Schema.org is defined as two hierarchies: one for textual property values, and one for the things that they describe.

This is the main schema.org hierarchy: a collection of types (or "classes"), each of which has one or more parent types. Although a type may have more than one super-type, here we show each type in one branch of the tree only. There is also a parallel hierarchy for **data types**.

Types:[Close hierarchy](#) / [Open hierarchy](#)**Thing**

- ▶ Action +
- ▶ BioChemEntity +
- ▶ CreativeWork +
- ▶ Event +
- ▶ Intangible +
- ▶ MedicalEntity +
- ▶ Organization +
- ▶ Person +
- ▶ Place +

Product

- DietarySupplement
- Drug
- IndividualProduct
- ProductCollection
- ProductGroup

- ▶ General purpose **lightweight** ontology
- ▶ Aimed at **annotating web pages**
- ▶ Targetting **FINDABILITY**
- ▶ Originating from major search engines



Schema.org is massively adopted

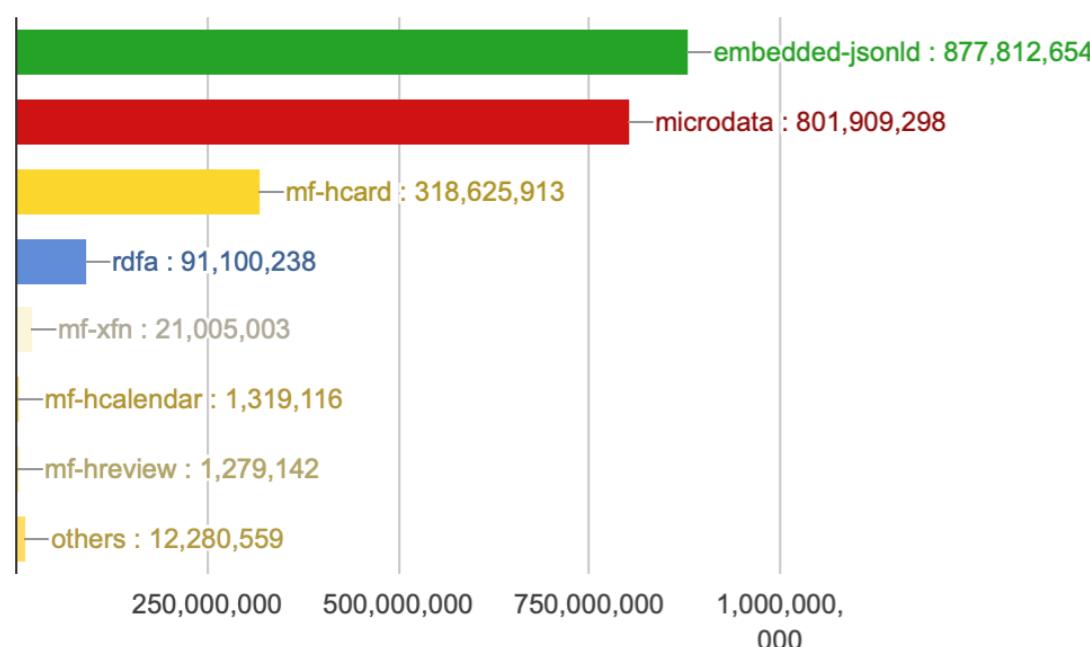
Web Data Commons

Extracting Structured Data from the Common Crawl



Crawl Date	October 2022	
Total Data	82.71 Terabyte	(compressed)
Parsed HTML URLs	3,048,746,652	
URLs with Triples	1,518,609,988	
Domains in Crawl	33,820,102	
Domains with Triples	14,235,035	
Typed Entities	19,072,628,514	
Triples	86,462,816,435	
Size of Extracted Data	1.6 Terabyte	(compressed)

URLs with Triples



Top Domains by Extracted Triples

1. [blogspot.com](#) (879,564,145 triples)
2. [wordpress.com](#) (458,770,038 triples)
3. [wikipedia.org](#) (190,087,065 triples)
4. [yummly.com](#) (87,112,540 triples)
5. [hotels.com](#) (81,991,039 triples)
6. [boohoo.com](#) (79,884,394 triples)
7. [kayak.com](#) (77,623,248 triples)
8. [google.com](#) (73,729,078 triples)
9. [yahoo.com](#) (65,317,838 triples)
10. [southleedslife.com](#) (63,758,451 triples)
11. [indiatimes.com](#) (58,899,559 triples)
12. [freepik.com](#) (56,124,447 triples)
13. [airbnb.com](#) (51,964,983 triples)
14. [pinterest.com](#) (47,251,484 triples)
15. [soundcloud.com](#) (45,745,317 triples)
16. [apple.com](#) (42,410,414 triples)
17. [hostadvice.com](#) (42,309,867 triples)
18. [elpais.com](#) (42,136,136 triples)
19. [vsemayki.ru](#) (38,167,517 triples)
20. [smugmug.com](#) (38,031,434 triples)
21. [More](#)

Schema.org for
Life-Science resources ?

37 ± Life Science profiles



Name	Group	Use Cases	Cross Walk	Task & Issues	Examples	Live Deploy
ChemicalSubstance (v0.4-RELEASE) 07 April 2020	Chemicals					
ComputationalTool (v1.0-RELEASE) 11 October 2021	Tools					
ComputationalWorkflow (v1.0-RELEASE) 09 March 2021	Workflow					
DataCatalog (v0.3-RELEASE-2019_07_01) 01 July 2019	Data Repositories					
Dataset (v0.3-RELEASE-2019_06_14) 14 June 2019	Datasets					
FormalParameter (v1.0-RELEASE) 09 March 2021	Workflow					
Gene (v1.0-RELEASE) 07 April 2021	Genes					
MolecularEntity (v0.5-RELEASE) 07 April 2020	Chemicals					
Protein (v0.11-RELEASE) 07 April 2020	Proteins					
Sample (v0.2-RELEASE-2018_11_10) 10 November 2018	Samples					
Taxon (v0.6-RELEASE) 07 April 2020	Biodiversity					

- ▶ different usage of schema.org for life sciences
- ▶ Communities agree on **minimal / recommended / optional** annotation

Metadata completeness

R1.3: (Meta)data meet domain-relevant community standards

Marginality: Recommended.				
applicationCategory	Text URL	Schema: Type of software application, e.g. 'Game, Multimedia'. Bioschemas: Type of tool e.g. Command-line tool, Web application etc. Note: Bioschemas have changed URL to Text in the Expected Types. This will be reverted once Bio.Tools provides stable URIs for tool types.	MANY	Please use terms from the 'Tool type' table in the biotools documentation 
applicationSubCategory	Text URL	Schema: Subcategory of the application, e.g. 'Arcade Game'. Bioschemas: Use an EDAM:Topic to describe the category of application	MANY	EDAM:Topic 
author	Organization Person	Schema: The author of this content or rating. Please note that author is special in that HTML 5 provides a special mechanism for indicating authorship via the rel tag. That is equivalent to this and may be used interchangeably.	MANY	
citation	CreativeWork IIRI	Schema: A citation or reference to another creative work, such as another publication	MANY	

```
ex:myTool    rdf:type    schema:SoftwareApplication, prov:SoftwareAgent ;  
            schema:description "This tool does ... " ;  
            schema:license <https://spdx.org/licenses/MIT.html> ;  
            schema:codeRepository <http://github.com/...> .
```

SHACL constraint template

```
1      ns:{{shape_name}} a sh:NodeShape ;
2      {% for c in target_classes %}
3          sh:targetClass {{c}} ;
4      {% endfor %}

5
6      {% for min_prop in min_props %}
7          sh:property [
8              sh:path {{min_prop}} ;
9              sh:minCount 1 ;
10             sh:severity sh:Violation
11         ] ;
12     {% endfor %}

13
14     {% for rec_prop in rec_props %}
15         sh:property [
16             sh:path {{rec_prop}} ;
17             sh:minCount 1 ;
18             sh:severity sh:Warning
19         ] ;
20     {% endfor %}
21 .
```

Metadata completeness

R1.3: (Meta)data meet domain-relevant community standards

Validation of Bioschemas profiles:

- rank missing metadata
- developer focus
on minimal metadata first

Check BioSchemas

<https://workflowhub.eu/workflows/18?version=1> has type <http://schema.org/ComputationalWorkflow>

Using <https://bioschemas.org/profiles/ComputationalWorkflow/1.0-RELEASE> for validation, specified from the **dct:conformsTo** property.

Required missing properties	Improvements
https://schema.org/input must be provided	https://schema.org/citation should be provided
https://schema.org/output must be provided	https://schema.org/contributor should be provided
	https://schema.org/creativeWorkStatus should be provided
	https://schema.org/documentation should be provided
	https://schema.org/funding should be provided
	https://schema.org/hasPart should be provided
	https://schema.org/isBasedOn should be provided
	https://schema.org/maintainer should be provided
	https://schema.org/publisher should be provided
	https://schema.org/runtimePlatform should be provided
	https://schema.org/softwareRequirements should be provided
	https://schema.org/targetProduct should be provided

Reuse of ontologies



Linked Open
Vocabularies
(LOV)

F4: (Meta)data are registered or indexed in a searchable resource

I2: (Meta)data use vocabularies that follow the FAIR principles

R1.3: (Meta)data meet domain-relevant community standards

Step 3: Metadata quality checks

Controlled vocabularies

Bioschemas

We now have a Knowledge Graph grounded to ontology concepts (classes) and relations (properties). Are these classes and properties already known in reference ontology registries such as LOV, OLS or BioPortal ?

Check Vocabularies

Congratulations ! All Classes and Properties are referenced in one or more of the registries checked !

Classes

http://schema.org/DataDownload	OLS	LOV	BioPortal
http://schema.org/Organization	OLS	LOV	BioPortal
http://schema.org/Person	OLS	LOV	BioPortal
https://schema.org/Dataset	OLS	LOV	BioPortal

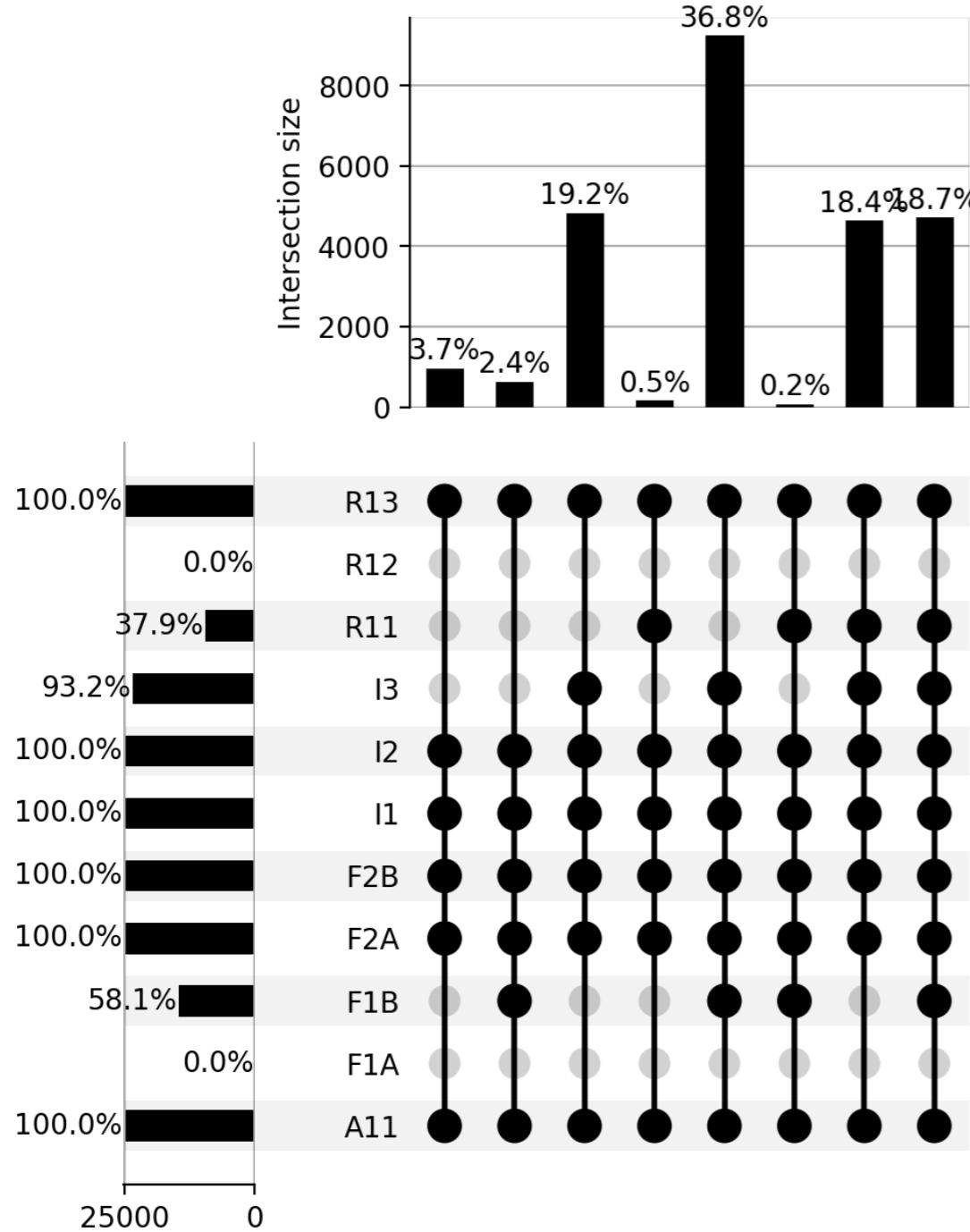
Properties

http://ogp.me/ns#description	OLS	LOV	BioPortal
http://ogp.me/ns#site_name	OLS	LOV	BioPortal
http://ogp.me/ns#title	OLS	LOV	BioPortal
http://ogp.me/ns#url	OLS	LOV	BioPortal
http://schema.org/affiliation	OLS	LOV	BioPortal
http://schema.org/author	OLS	LOV	BioPortal
http://schema.org/contentSize	OLS	LOV	BioPortal
http://schema.org/contentUrl	OLS	LOV	BioPortal
http://schema.org/creator	OLS	LOV	BioPortal
http://schema.org/dateCreated	OLS	LOV	BioPortal
http://schema.org/dateModified	OLS	LOV	BioPortal
http://schema.org/datePublished	OLS	LOV	BioPortal

Using FAIR-Checker at
large scale

Large-scale FAIR metrics evaluations

How FAIR are Bio.Tools registered softwares ?

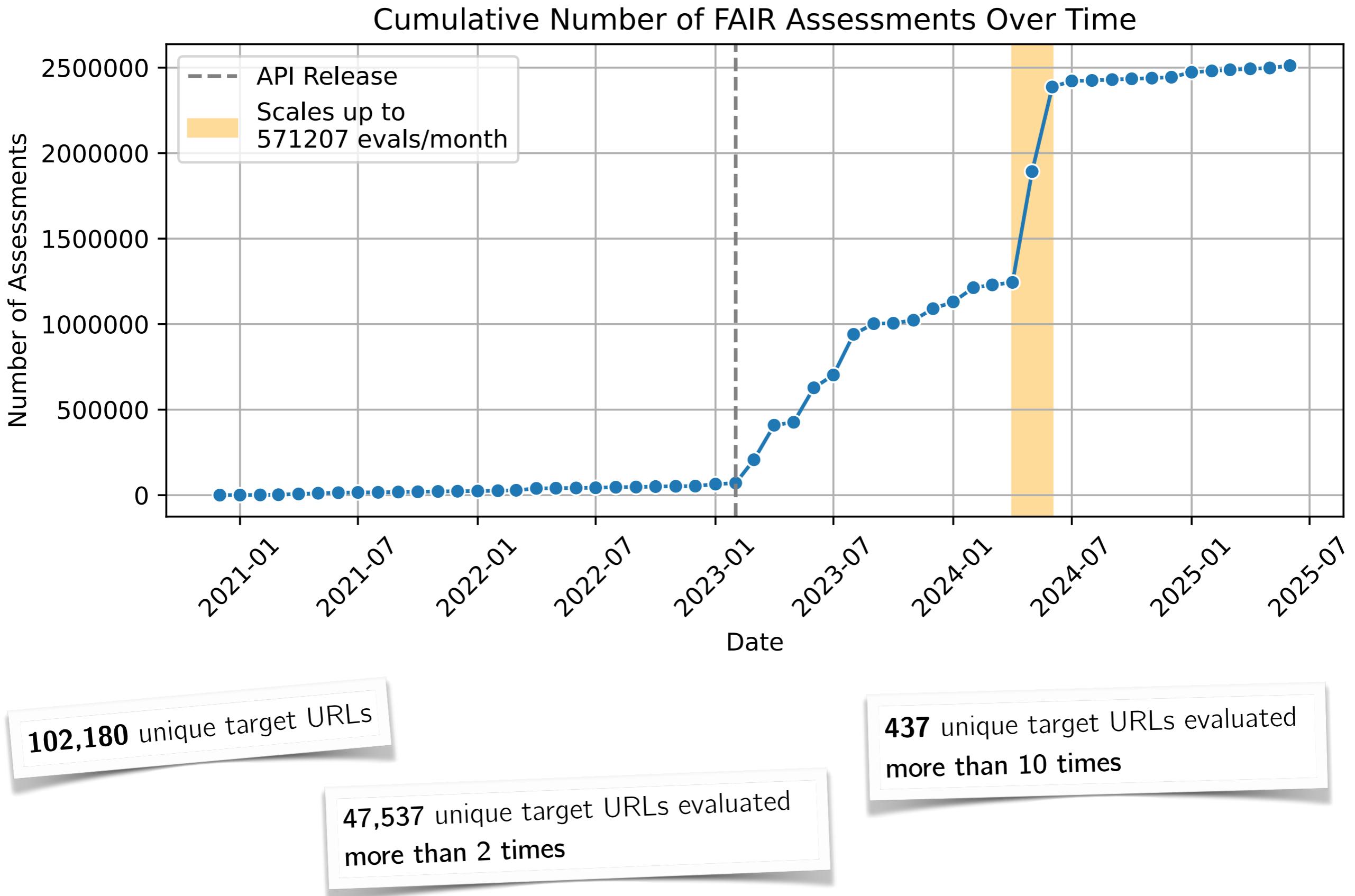


Running FAIR-Checker over
25.000+ bioinformatics softwares
from Bio.tools.

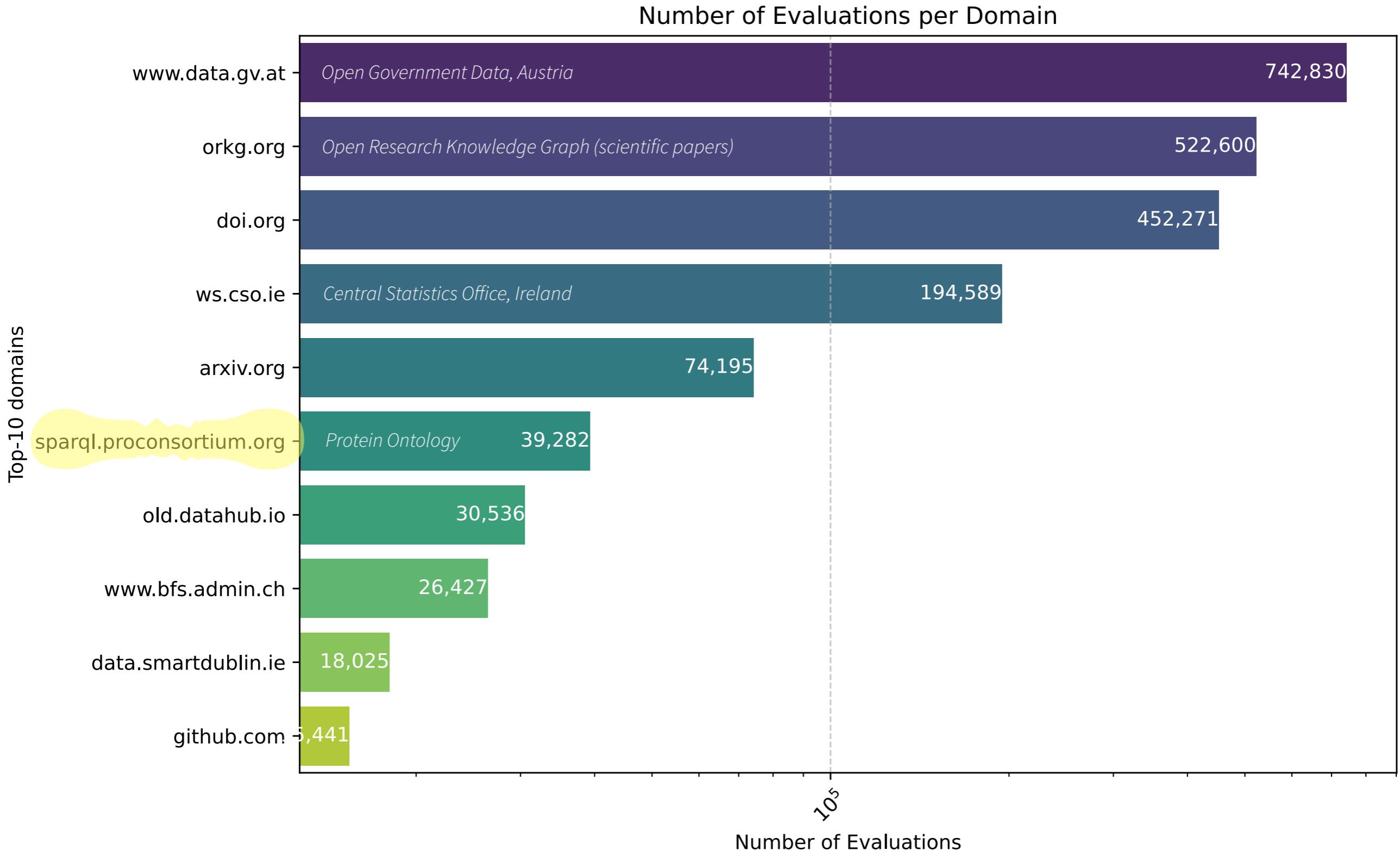
R1.1: Only 37,9% of the tools
expose a **licence**

R1.2: **No provenance** metadata
→ massive impact if bio.tools
developers provide PROV / PAV
ontology terms

Usage statistics

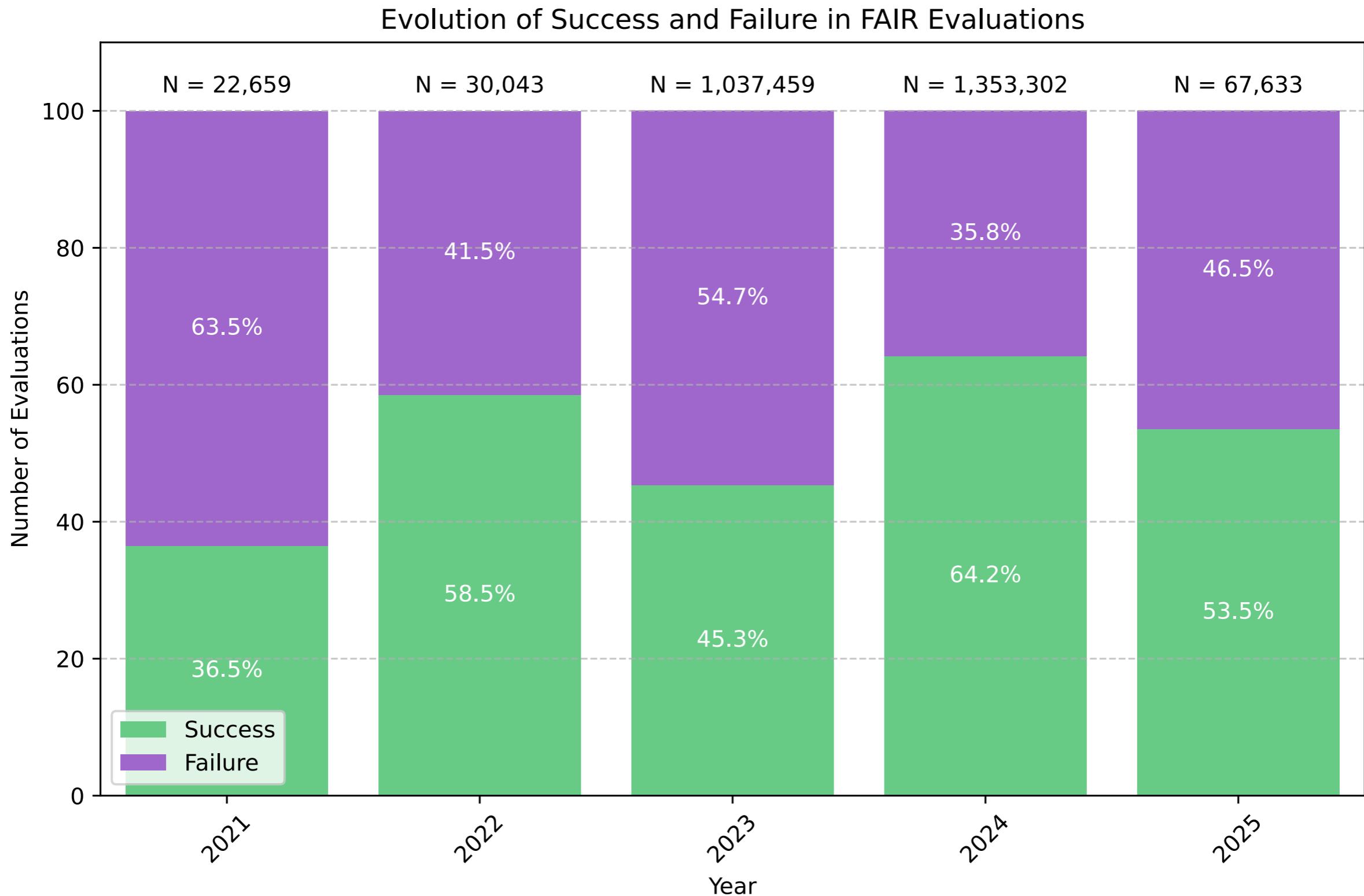


Most evaluated domains



Evolution of FAIR scores

Work in progress



A community service, with possible improvements

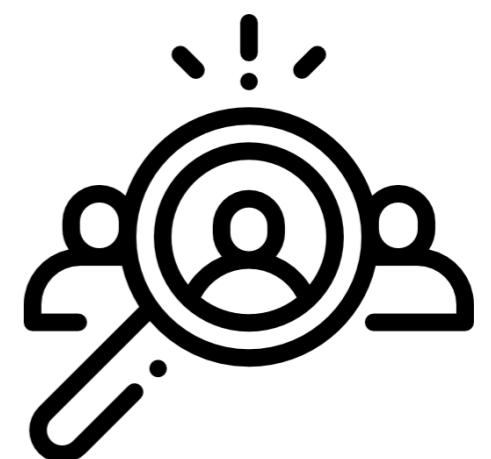


<https://fair-checker.france-bioinformatique.fr>

A screenshot of a project page from the EOSC Macro-Roadmap. The header includes the EOSC logo and the text "Macro-Roadmap". The main title is "FAIR-Checker" with a "FAIR Checker" logo next to it. Below the title, there are several data points: "Funding source: National", "In-kind value: < €100k", "Timeframe: 2021–ongoing", "Target group: Researchers", "Scale: International", and "Year of reporting: 2024". A section titled "Good practice" describes FAIR-Checker as an online tool for automating FAIR assessments. On the right side, there are logos for "cnrs", "Inserm", "INRAE", and "cea". A "Collaborators:" section lists "National Institute for Health and Medical Research (Inserm)", "National Research Institute for Agriculture, Food and Environment (INRAE)", and "French Alternative Energies and Atomic Energy Commission (CEA)".

Future works

- ▶ Support "FAIR-Signposting" for **better metadata consumption**
- ▶ Bioschemas **profile recommender**
- ▶ Allow users to provide **missing metadata**
- ▶ Suggest semantic metadata based on **AI generation** pipeline
- ▶ Retrospective **usage study**





Jennien Andersen



Sylvie Cazalens



Philippe Lamare

Many FAIR assessment tools ...
→ do they provide similar results ?

A Generic Framework to Better Understand and Compare FAIRness Measures

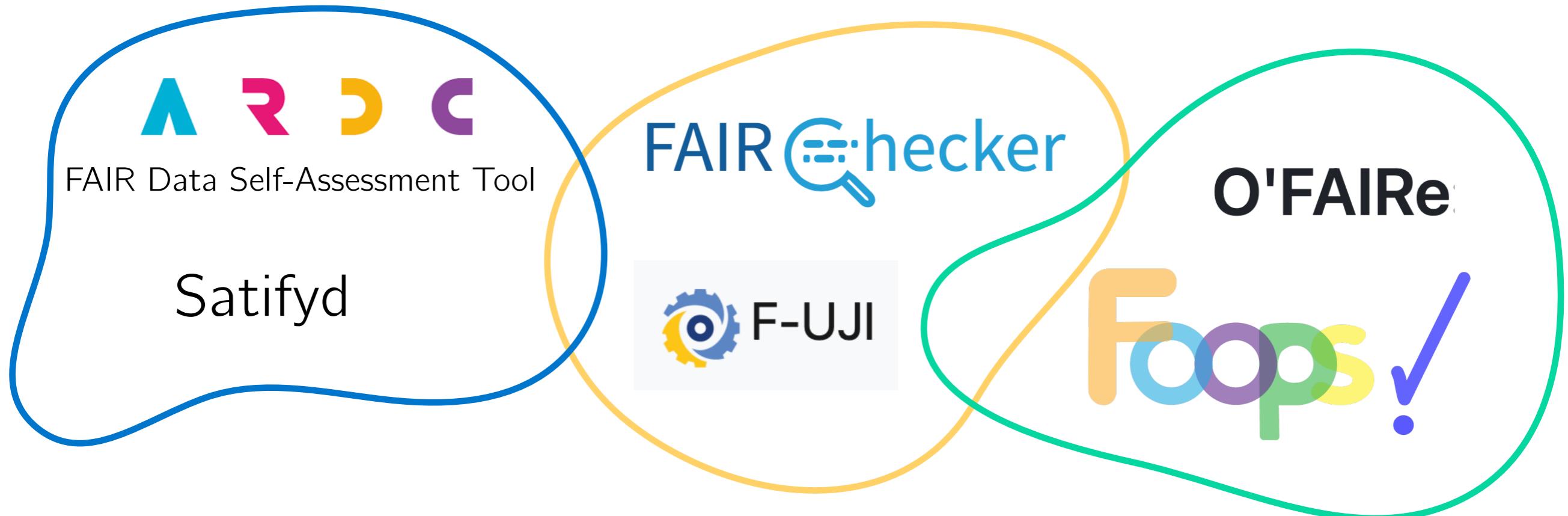
Philippe Lamarre¹, Jennie Andersen¹, Alban Gaignard², and Sylvie Cazalens¹

¹ INSA Lyon, CNRS, Ecole Centrale de Lyon, Université Claude Bernard Lyon 1,
Université Lumière Lyon 2, LIRIS, UMR5205 69621 Villeurbanne, France
firstname.lastname@insa-lyon.fr

² Nantes Université, CNRS, INSERM, l'institut du thorax, F-44000 Nantes, France
alban.gaignard@univ-nantes.fr

Abstract. In recent years, the adoption of the FAIR principles has achieved notable success. This progress has led to the development of numerous assessment tools originating from diverse fields of application, thus addressing diverse object types, interpretations and implementations. Given the plethora of proposals available, it is crucial for users to precisely understand these measures, compare them effectively, make

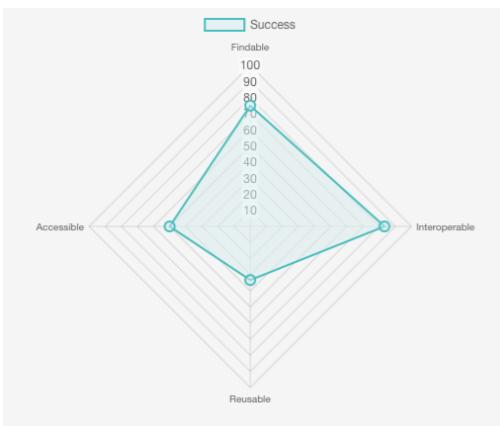
Multiple FAIR assessment tools



- ▶ Check lists / questionnaires
- ▶ Automated tools
- ▶ Some tools are "community oriented"
- ▶ Some tools are "technologically opinionated"

How FAIR is a bioinformatics software ?

FAIR  checker

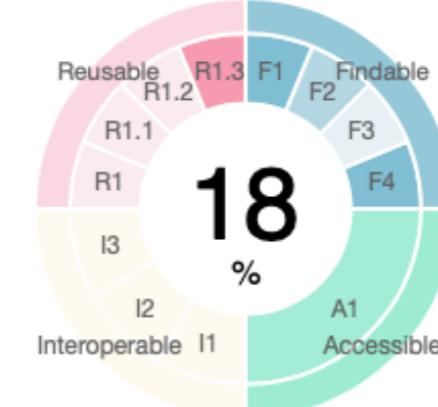


FAIR assessment 62.50 %



<https://bio.tools/seurat>

 F-UJI



	Score earned:	Fair level:
Findable:	2.5 of 7	 initial
Accessible:	1 of 3	 initial
Interoperable:	0 of 4	 incomplete
Reusable:	1 of 10	 initial

- ▶ Why are results so different?
- ▶ Which tool should I use?
- ▶ What's behind the scene?

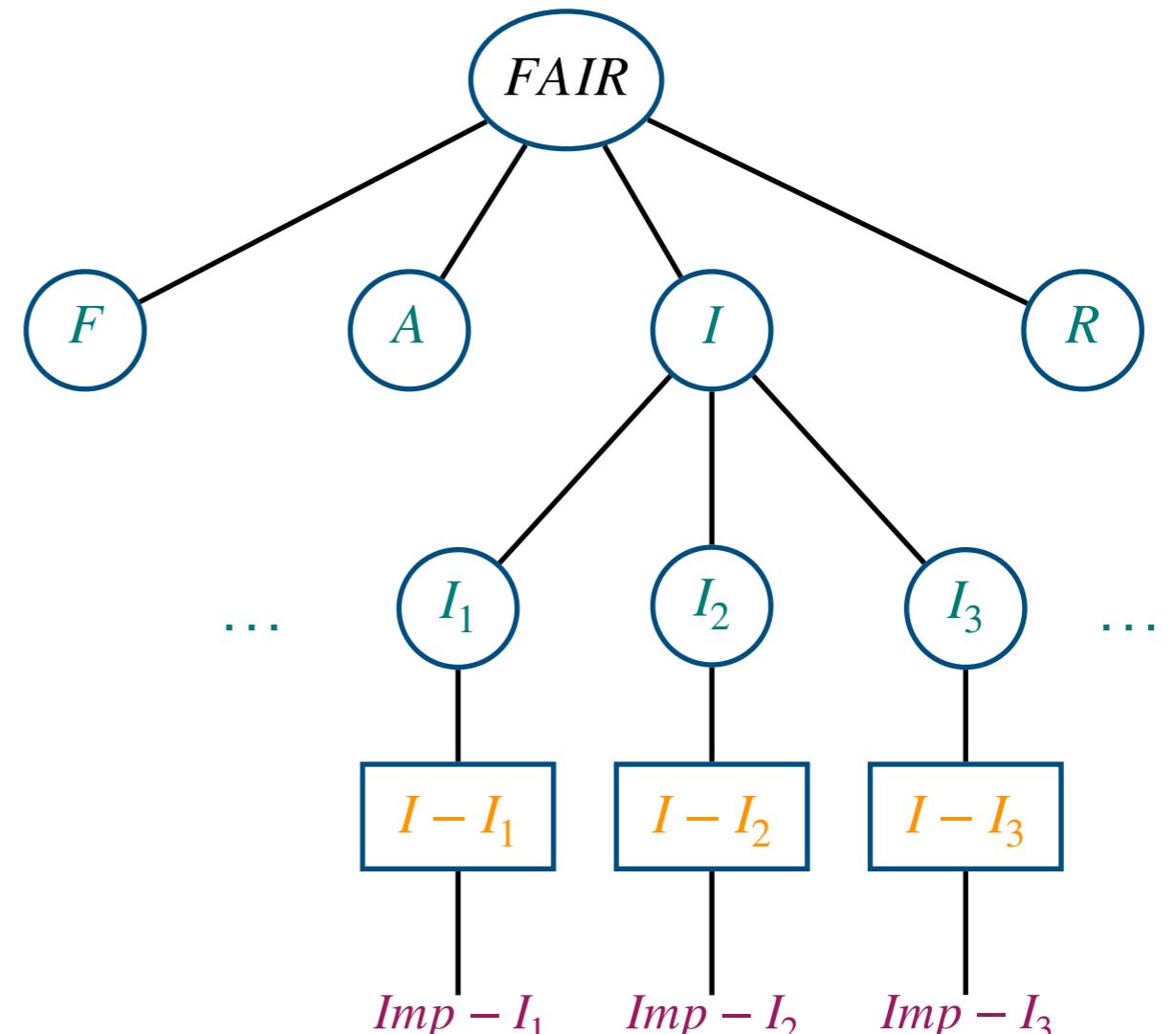
① We need a computable
framework for *modeling*
FAIR measures

Objectives

- (i) **specify measures** in a uniform and computable **model**
- (ii) propose specific **quantities** to **analyse and compare**
FAIR assessment approaches

Typical structure for FAIR assessment methods

- ▶ Principles P : the FAIR principles and their sub-principles
- ▶ Indicators $I(\mathcal{M})$: the specification of principles, (i.e. what has to be verified)
- ▶ Implementations $Imp(\mathcal{M})$: the implementations of the principles



$$\mathcal{M} = \underbrace{(V, E, FAIR)}_{\text{structure}}, \underbrace{\diamondsuit, w, v_{max}, D}_{\text{score}}$$

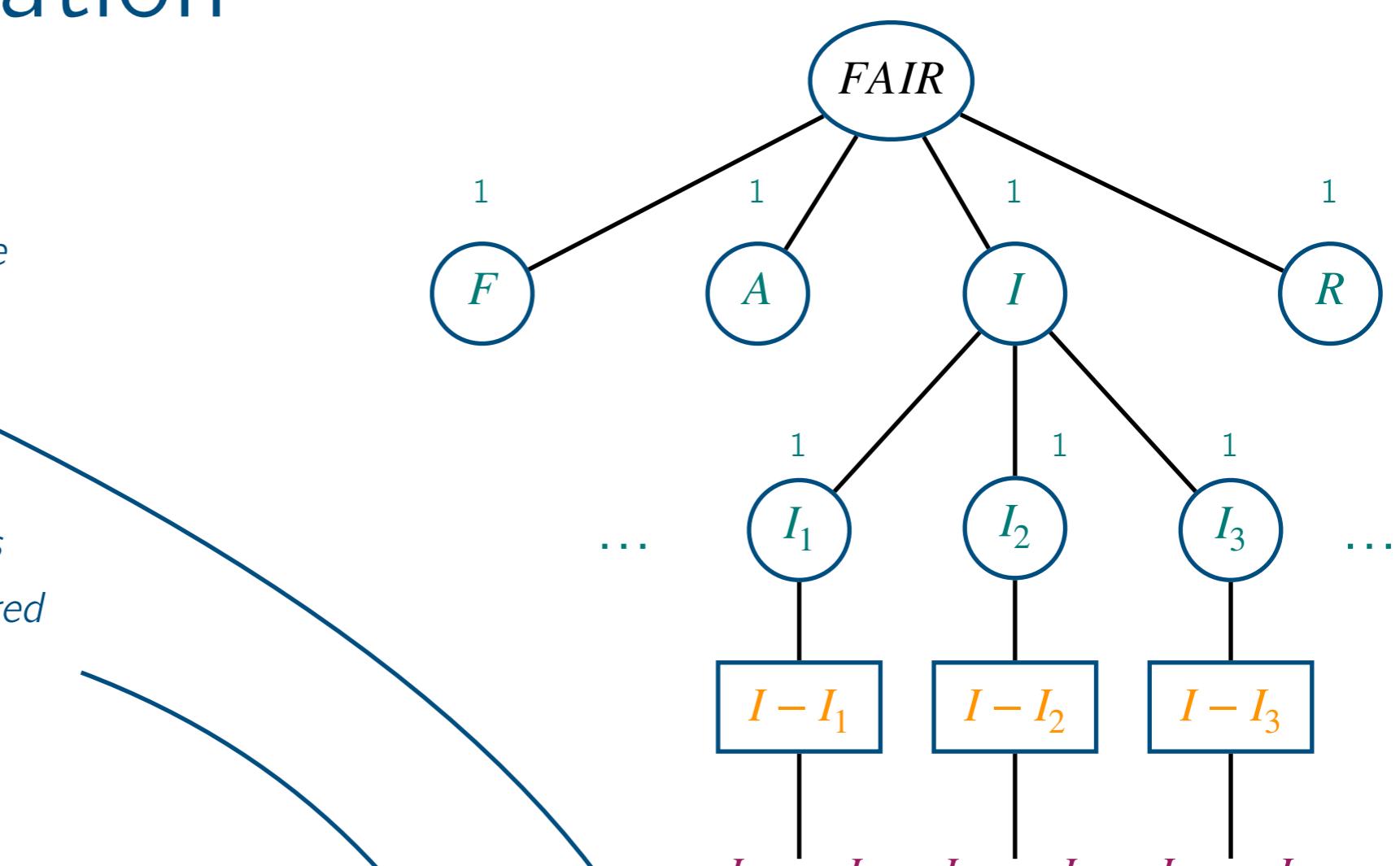
Score computation

Max function: $v_{max}(i)$ returns the maximum reachable score for an indicator or an implementation

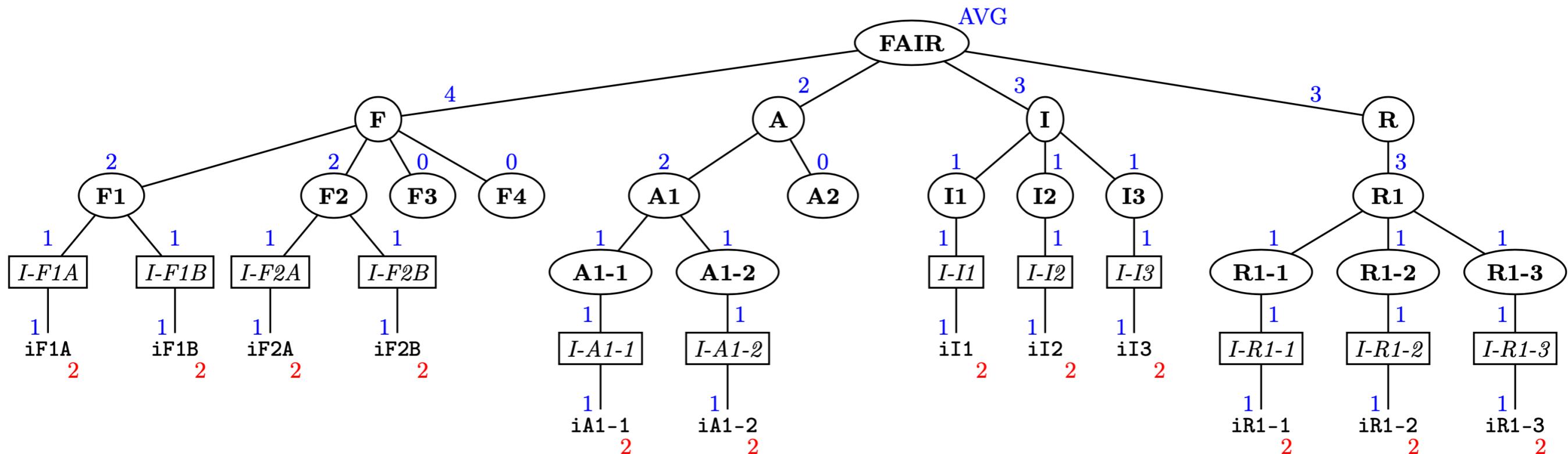
Weighting function: $w(n)$ returns the weight of the node n compared to its siblings

Aggregation function : weighted sum or weighted average to collect scores

$$\mathcal{M} = \underbrace{(V, E, FAIR, \diamond, w, v_{max}, D)}_{score}$$



Model instantiation with FAIR-Checker

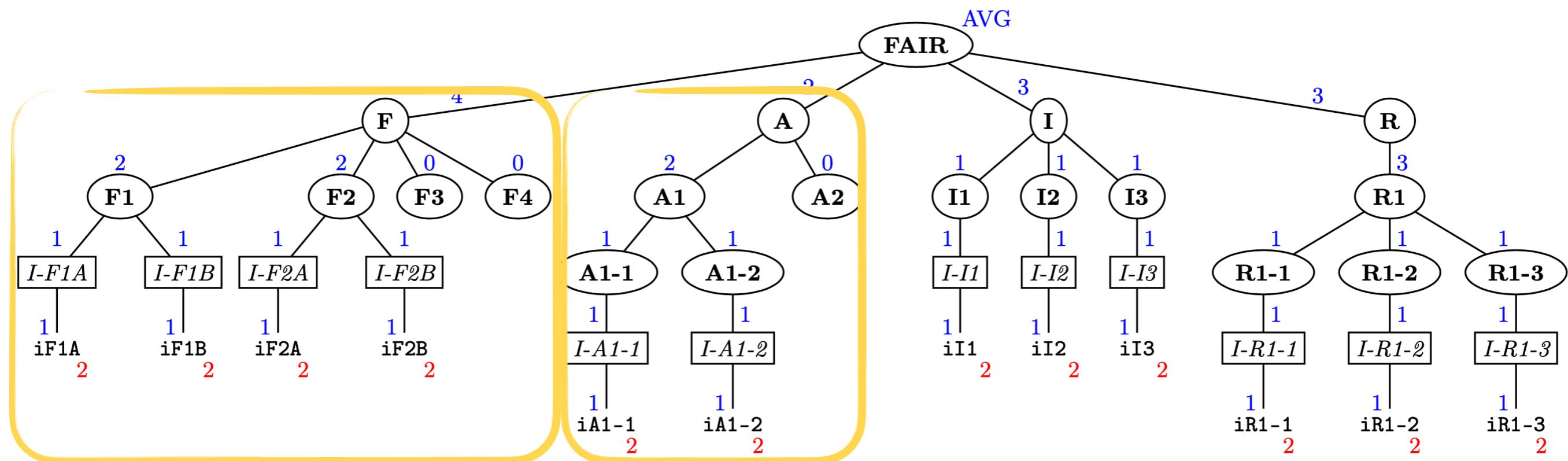


- ▶ The tool fits the tree-based structure $\{P, I, Imp\}$
- ▶ needs the maximum reachable score per implementation $\rightarrow 2$
- ▶ needs an aggregation function $\rightarrow AVG$ (*normalized scores*)
- ▶ needs cumulative weights for each principle $\rightarrow w \in [0,4]$

② We need quantities to characterize and compare FAIR measures

Quantifying the *granularity* of a measure

Idea: a global metric computed as the mean number of indicators per principles with at least one indicator. **If we have one principle with many implementations → important granularity.**

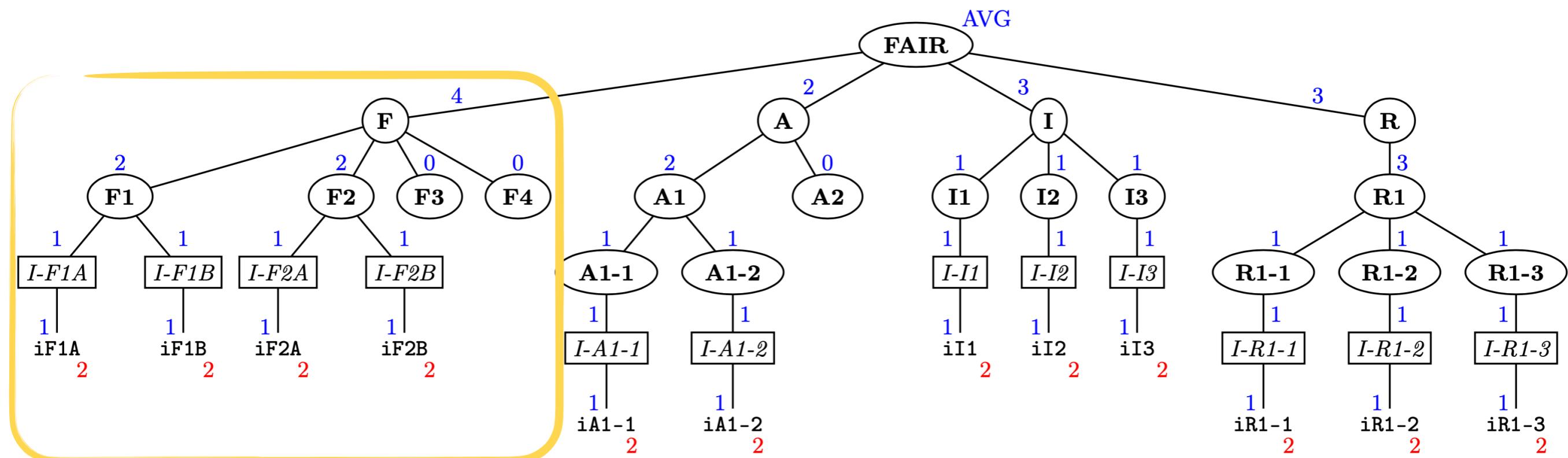


F has 4 indicators, A has only 2 → principles have not the same granularity

$$\text{granularity}(F) = 4/2 = 2 ; \text{granularity}(FAIR) = 12/10 = 1.2$$

Quantifying the *coverage* of a measure

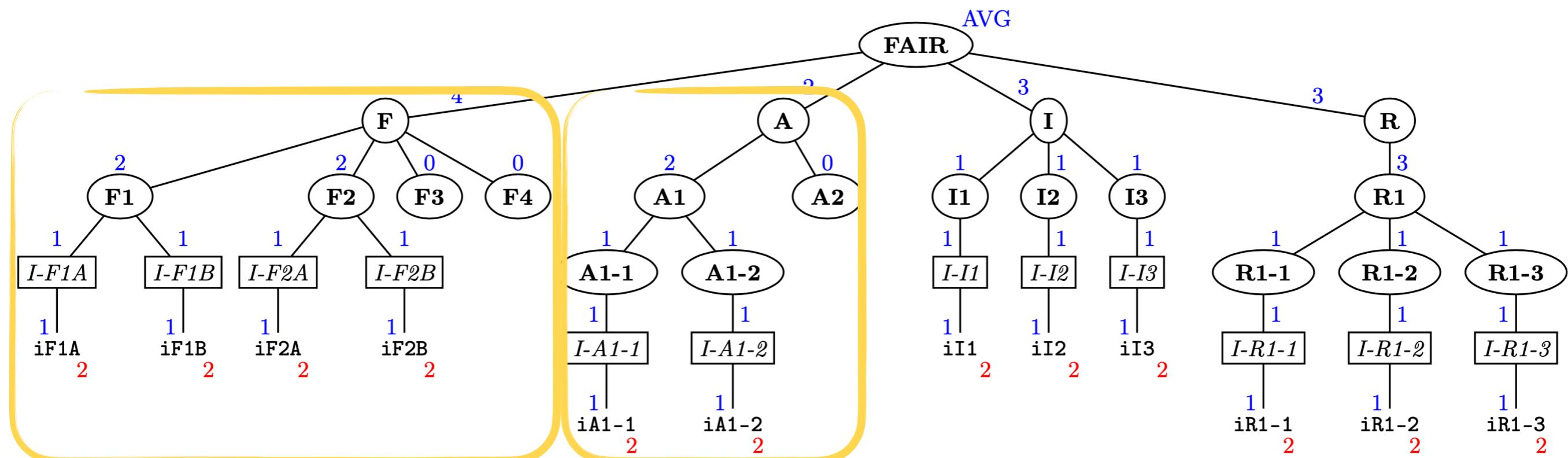
Idea: a principle is considered as covered if it exists at least one implementation for the principle or, its sub-principles.



F3 & F4 are never evaluated → F has only 50% of coverage

Quantifying the *impact* of principles

Idea: the impact of a principle corresponds to the **global score** obtained when **all** its underlying **implementations** are **successful** (without considering other principles).



$$\text{imp}(F) = 4*2 / 12*2 = 2/6$$

$$\text{imp}(A) = 2*2 / 12*2 = 1/6$$

→ F contributes 2 times more compared to A in the global FAIR score

Experimental results

Experimental setup

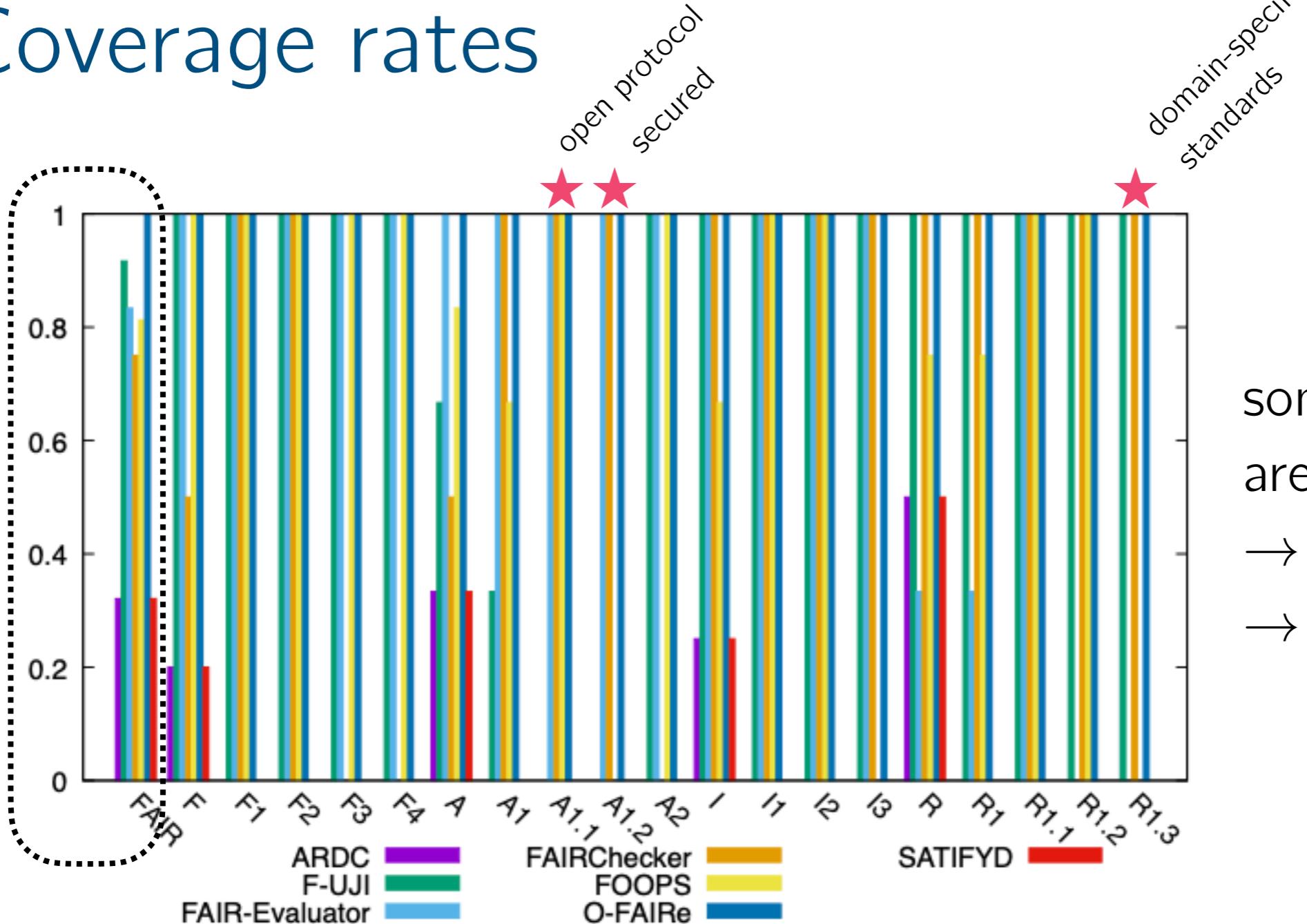
- ▶ Evaluated tools
 - automated: F-UJI, FAIR Evaluator, FOOPS!, O'FAIRe, FAIR-Checker,
questionnaires: ARDC, SATiFYD
- ▶ 10 selected web resources for FAIR assessments, covering
 - datasets descriptions
 - ontologies
 - online courses
 - bioinformatics software
 - RDF files

Do engines reach consensus on FAIR assessment ?

Resource	F-UJI (%)	FAIR-Checker (%)	Std dev
Dataset (PANGAEA) [31]	91	91.70	0.49
Gene Ontology (OLS) [21]	18	16.70	0.92
Dataset (Harvard Dataverse) [23]	75	79.20	2.97
Dataset (Kaggle) [26]	60	70.80	7.64
Online course (Moodle) [28]	4	16.70	8.98
Dataset (Governmental platform) [22]	52	70.80	13.29
Dataset (WHO) [39]	27	50.00	16.26
Training material (TeSS) [36]	39	70.80	22.49
Bioinformatics tool (bio.tools) [6]	18	54.20	25.60
Dataset (RDF metadata) [33]	43	87.50	31.47

- ▶ Higher scores for FAIR-Checker
- ▶ Last two entries: std. dev. > 25 % ?

Coverage rates

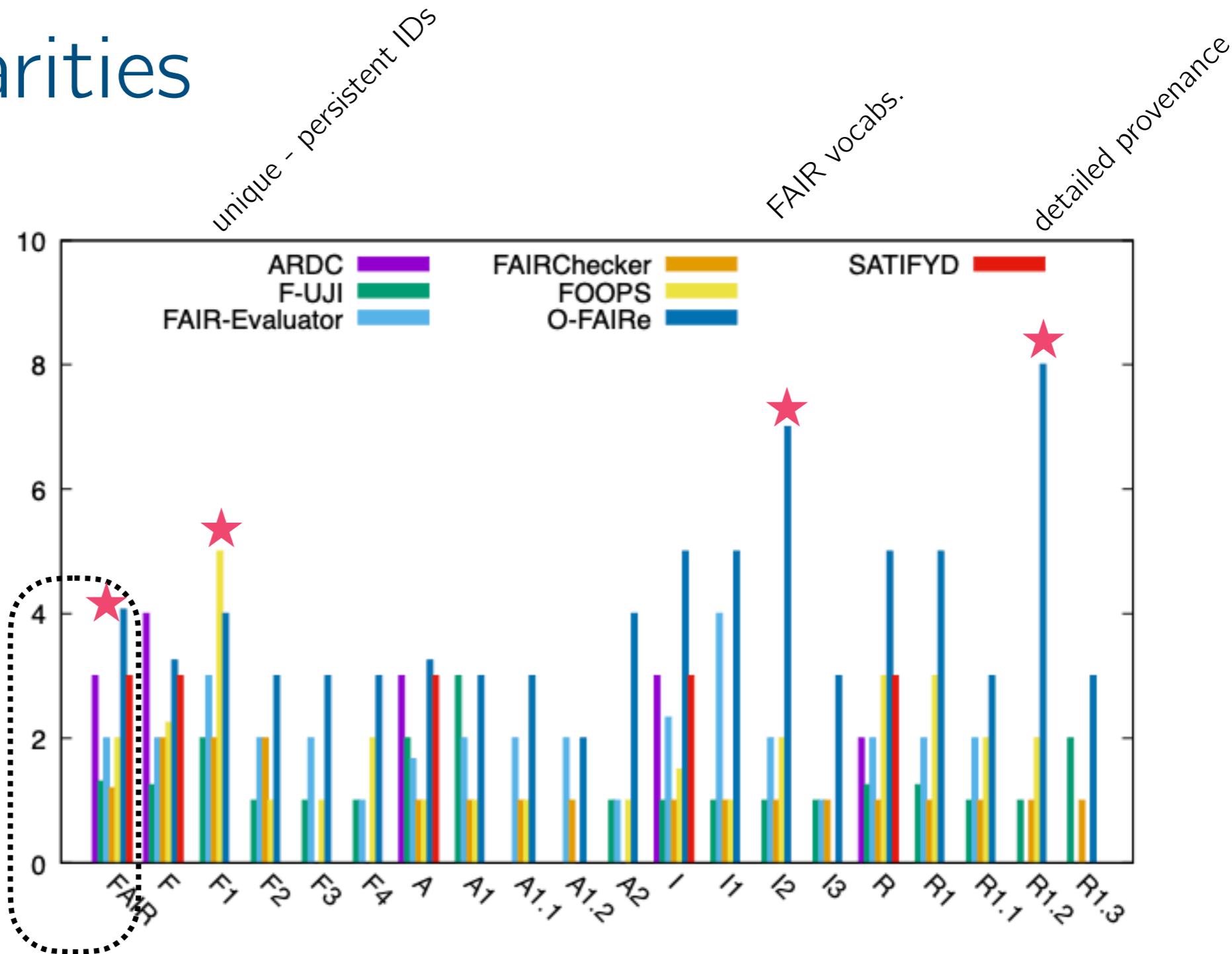


some principles (★)
are not implemented
→ technical issues ?
→ interpretation ?

{O-FAIRe} → the only one with a full coverage (1)

Low coverage for questionnaires (no sub-principles)

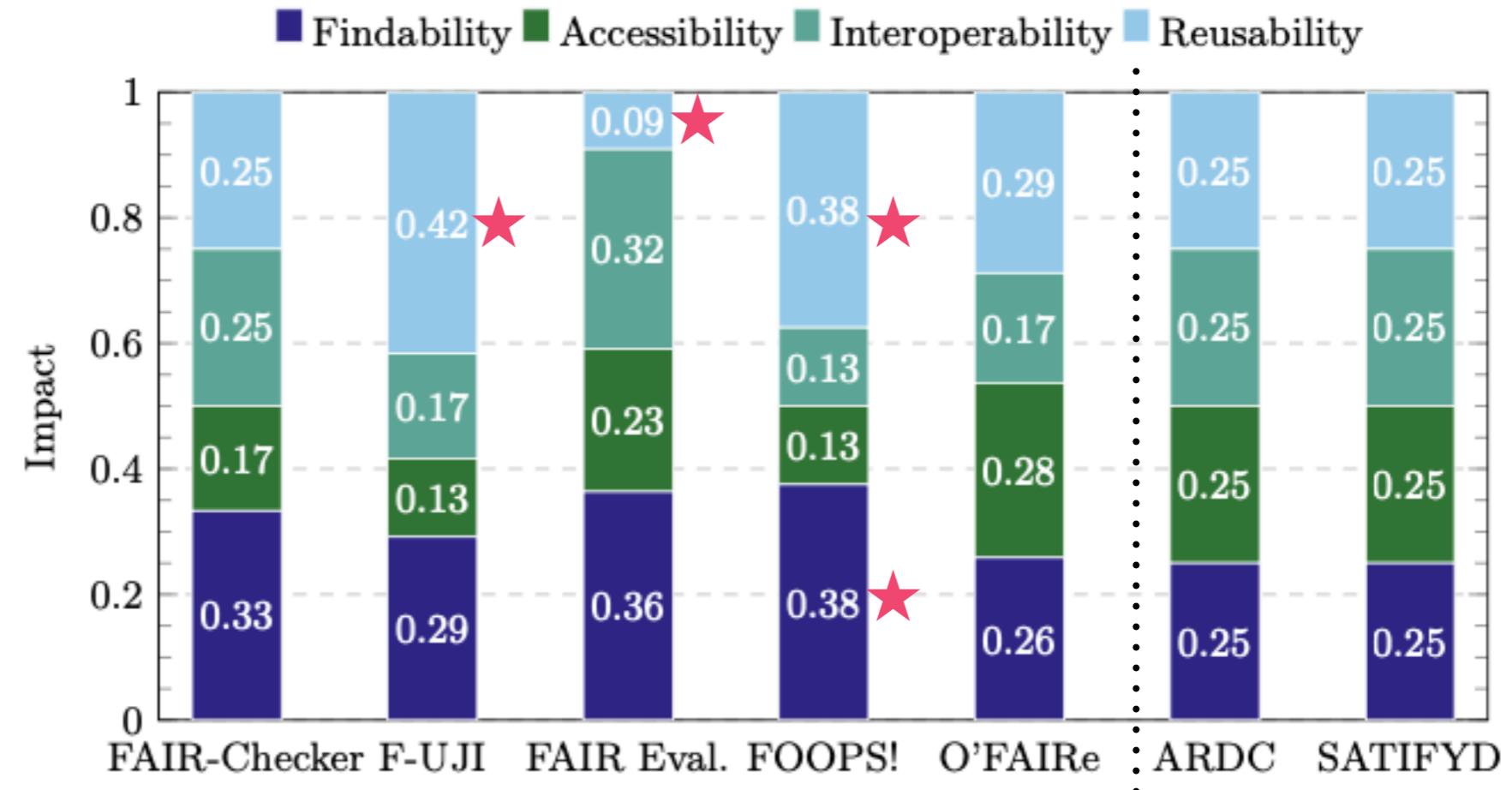
Granularities



- ▶ O-FAIRe (targeting ontologies) has the most fine-grained evaluation (e.g. I₂ and R_{1.2}) → great diversity in metadata for ontologies
- ▶ FOOPS deeply investigate F₁ (identifiers)

Impacts

Are all principle equally contributing to the global FAIR assessment score ?



→ *How to get a good FAIR score with a minimal effort ?*

- ▶ Pay attention to identifiers (F), license + provenance + domain-specific standards (R) if you use FOOPS!
- ▶ Not useful to spend energy on provenance or domain ontologies if you use FAIR-Evaluator ...
- ▶ ... but pay attention to it if you use F-UJI.

Conclusion and future works

Take-home message

- ▶ We need to understand the specificities of FAIR measurement tools
- ▶ We introduced a generic **model** for representing FAIR measures and computing their **granularity**, **coverage** and **impact**
- ▶ Some **tools are "biased"** (intentionally or not):
 - they explore more deeply some dimensions
 - which has an impact on the scores
- ▶ **Future works** include
 - better investigating **implementations** (e.g. dependencies), with tools developers
 - **share** these metadata on the web (e.g. DQV ontology)
 - contribute to **FAIR harmonization efforts** with other communities (softwares, workflows, machine learning resources ...)

Acknowledgements



Marie-Dominique Devignes



Thomas Rosnet



Frédéric de Lamotte



Philippe Lamare



Sylvie Cazalens



Jennien Andersen

Funding:

ANR AAPG2019 DeKaloG

French institute for Bioinformatics (Elixir-FR)

contact

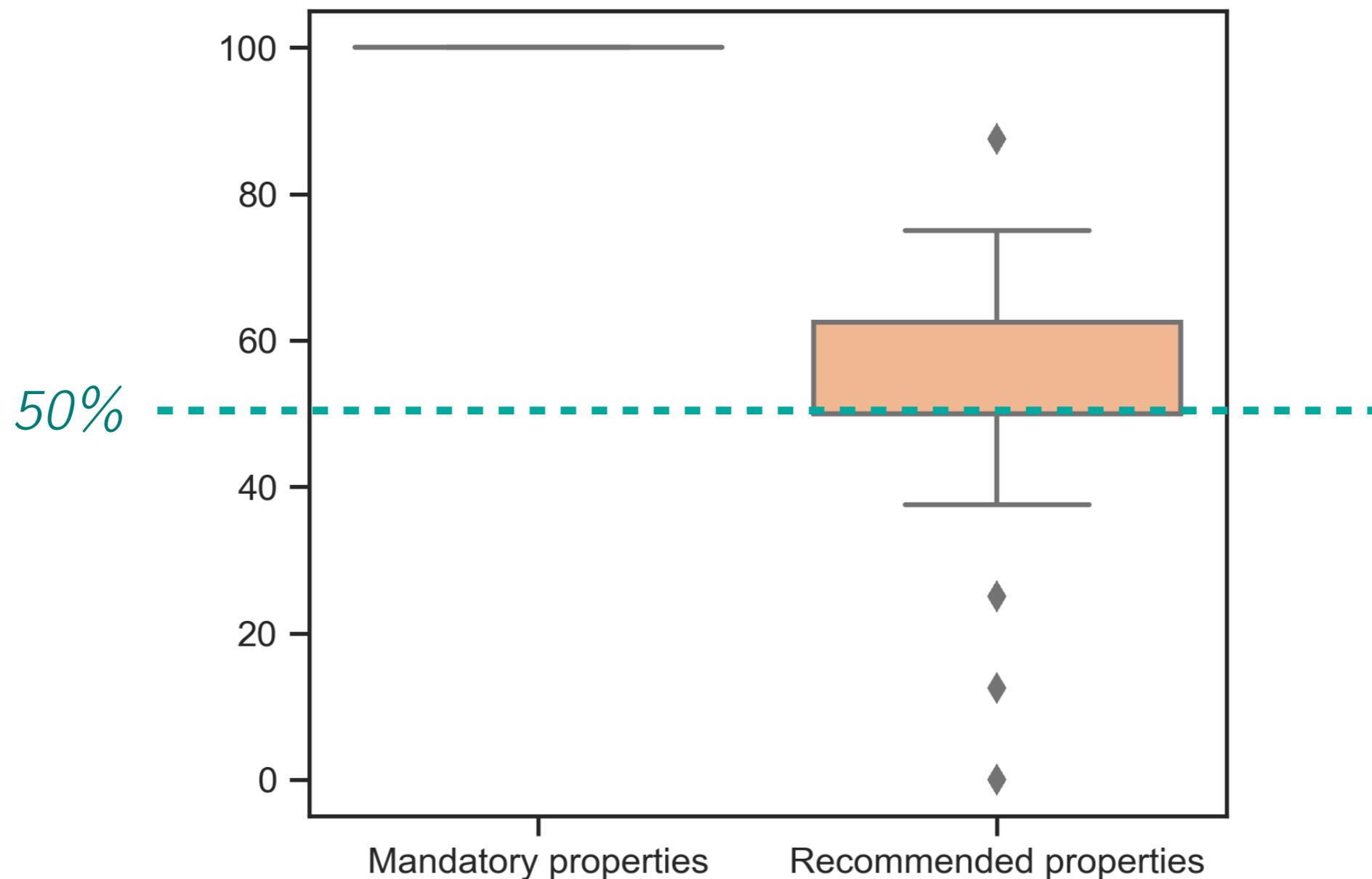
alban.gaignard@univ-nantes.fr

Back-up slides

SHACL shapes evaluations

→ machine-actionable bioschemas profile

Compliance of 25048 bioinformatics softwares
with the Computational Tool Bioschemas profiles



Impact

$$best_n(\textit{imp}) = \begin{cases} v_{max}(\textit{imp}) & \text{if } \textit{imp} \in \text{desc}(n) \\ 0 & \text{otherwise} \end{cases}$$

$$\text{impact}(\mathcal{M}, n) = \frac{\text{score}(\mathcal{M}, best_n, \text{FAIR})}{\text{score}(\mathcal{M}, best_{FAIR}, \text{FAIR})}$$

Coverage

$$\text{cover}(\mathcal{M}, n) = \begin{cases} 0 & \text{if } \text{children}_{\mathcal{M}}(n) = \emptyset \\ \frac{\left(\sum_{n' \in \text{children}_{\mathcal{M}}(n) \cap P} \text{cover}(\mathcal{M}, n') \right) + \text{local}(\mathcal{M}, n)}{|\text{children}_{\mathcal{M}}(n) \cap P| + \text{local}(\mathcal{M}, n)} & \text{else} \end{cases}$$

Granularity

$$\text{gran}(\mathcal{M}) = \text{gran}(\mathcal{M}, \text{FAIR}) \quad (5)$$

where, for $n \in P$:

$$\text{gran}(\mathcal{M}, n) = \frac{|\text{desc}_{\mathcal{M}}(n) \cap I(\mathcal{M})|}{|\{p \in (\{n\} \cup \text{desc}_{\mathcal{M}}(n)) \cap P, \text{children}_{\mathcal{M}}(p) \cap I(\mathcal{M}) \neq \emptyset\}|} \quad (6)$$