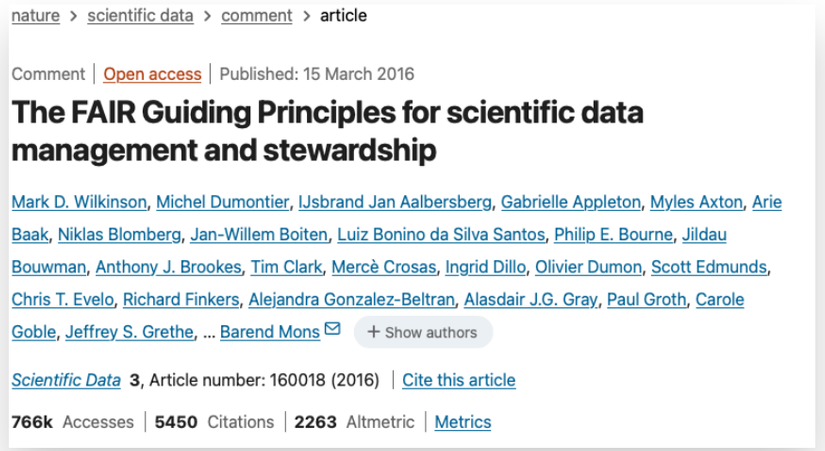
A generic framework to better understand and compare FAIRness measures

Philippe Lamarre, Jennie Andersen, Alban Gaignard, Sylvie Cazalens

EKAW 2024 - Amsterdam

FAIR principles ≠ technical specifications

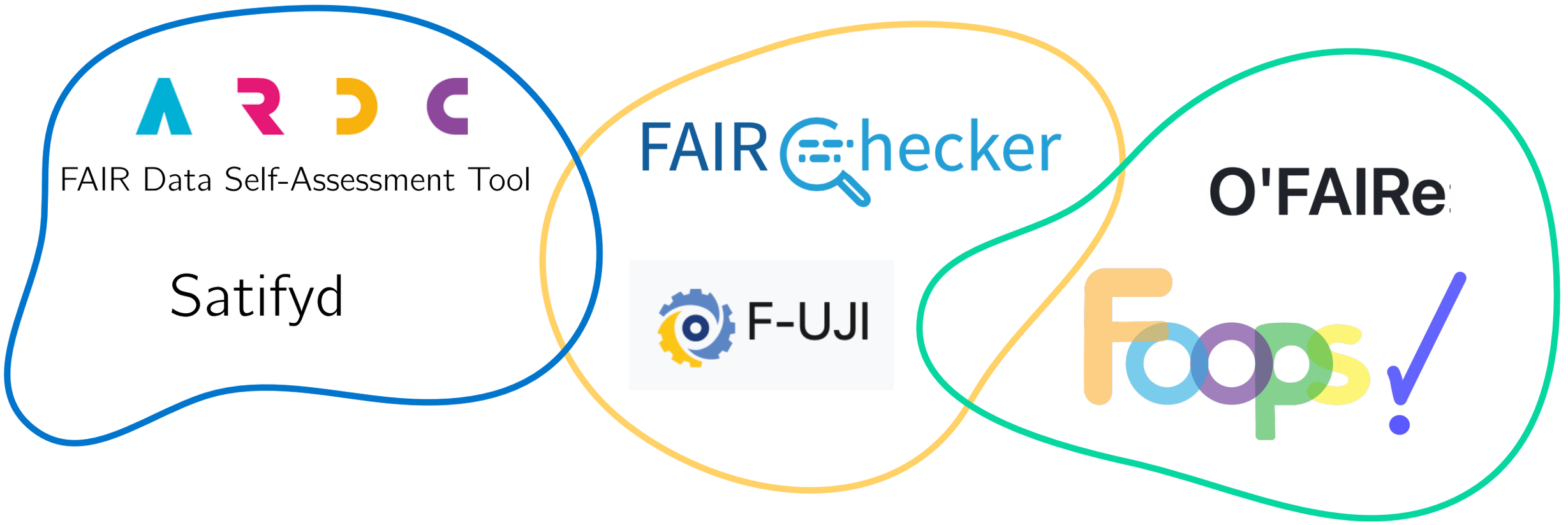
- Largely **adopted** by research funders and organizations
- Key for more open & reproducible sciences



However

- Principles, non-technical guidelines
- can be **interpreted** differently by specific communities
- **many** ways of **implementing** them

Multiple FAIR assesment tools

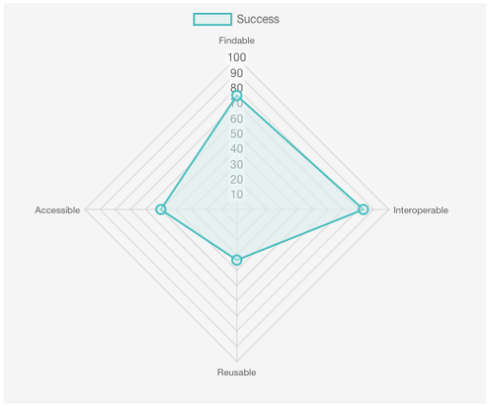


- Check lists / questionnaires
- Automated tools

- Some tools are "community oriented"
- Some tools are "technologically opinionated"

How FAIR is a bioinformatics software ?

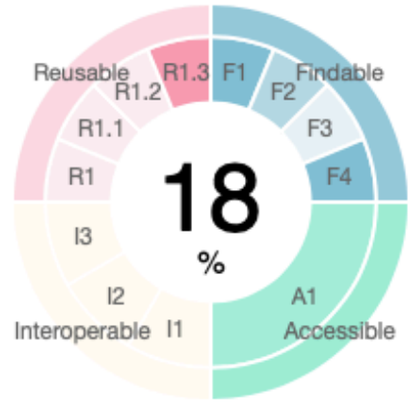
FAIR Checker



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?

One digital object …

… but different FAIR results 🤔

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

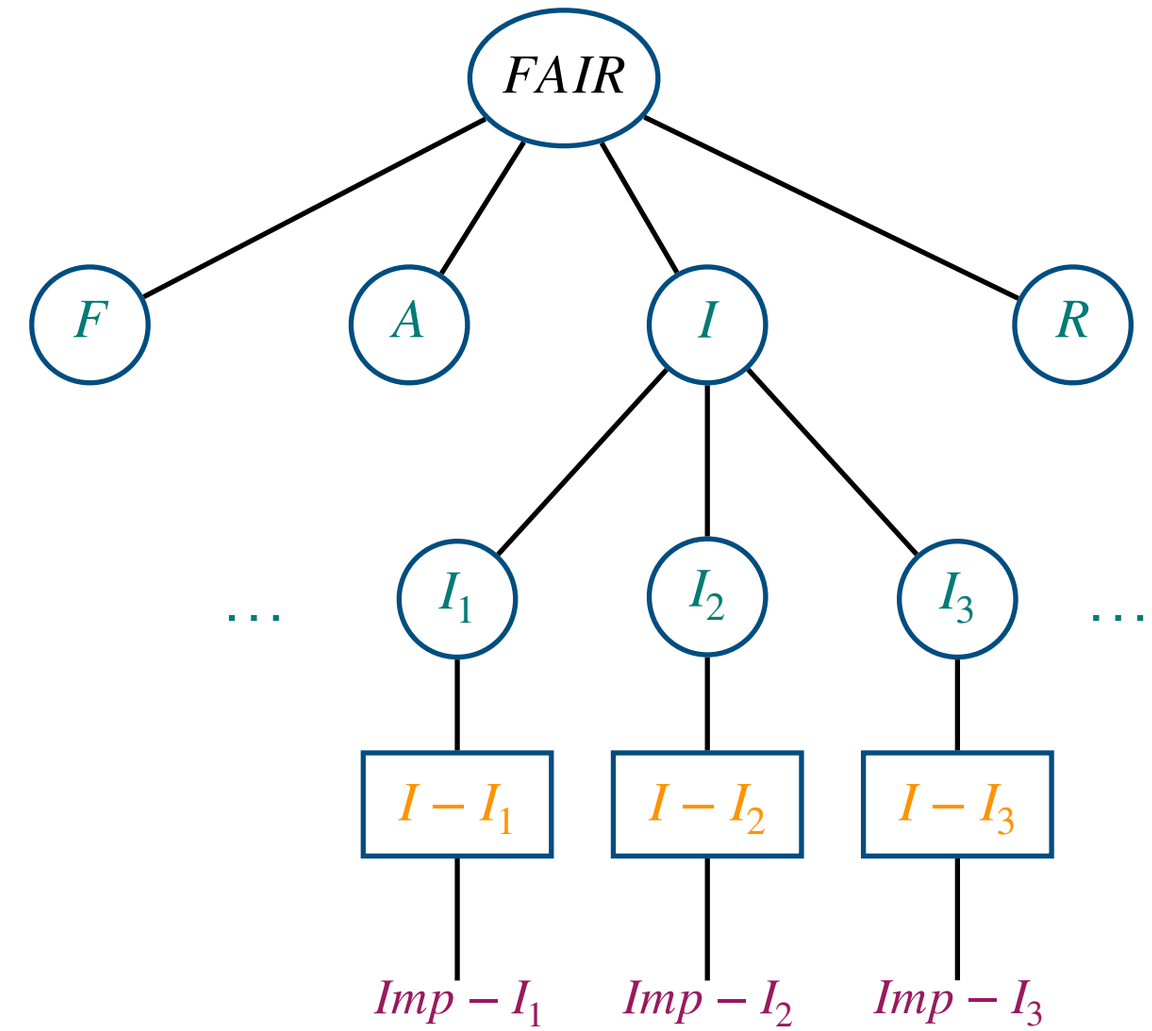
Objectives

(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 $I(\mathscr{M})$: the specification of principles, (i.e. what has to be verified)
- Implementations $Imp(\mathscr{M})$: the implementations of the principles



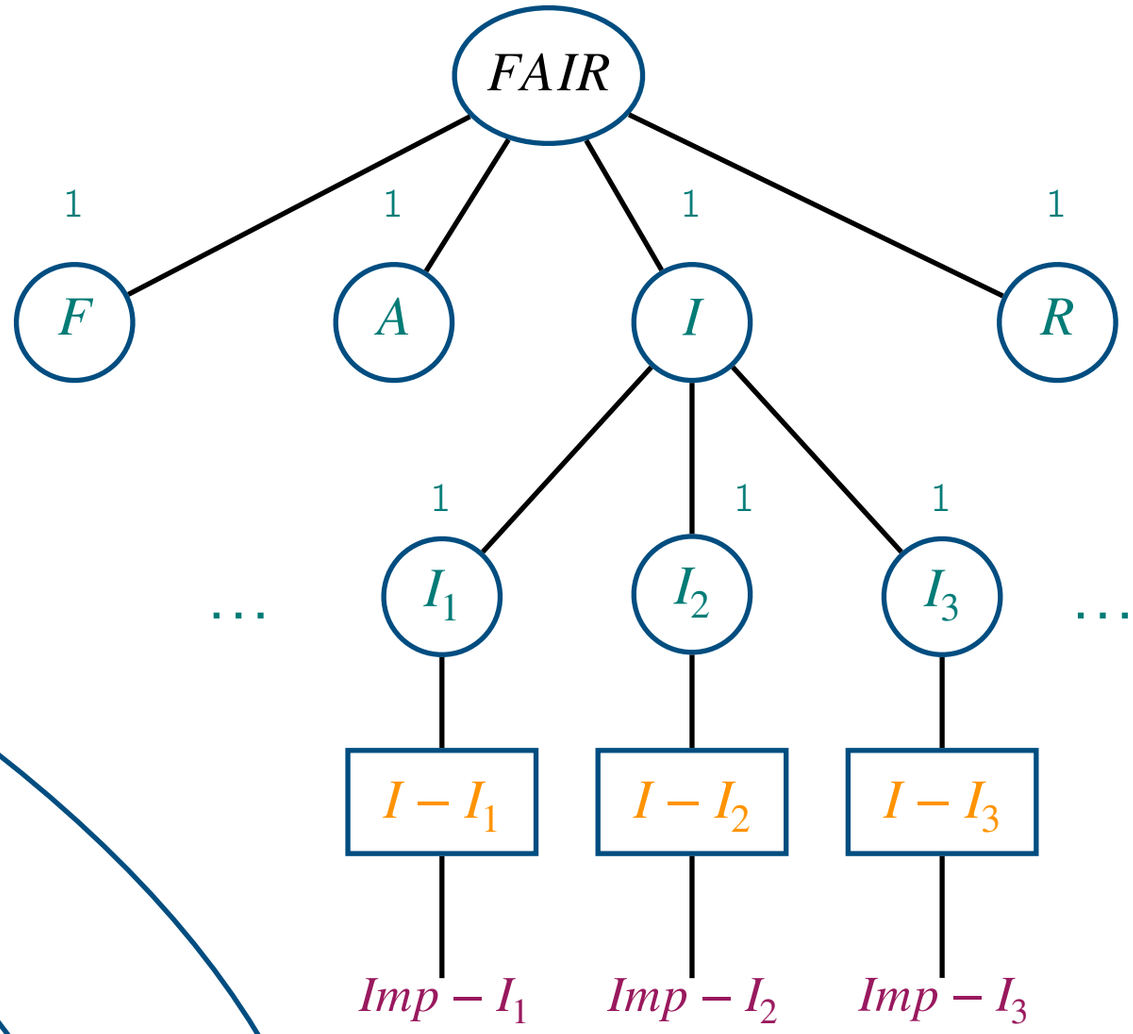
$$\mathscr{M} = (\overbrace{V, E, FAIR}^{structure}, \underbrace{\Diamond, w, v_{max}, D}_{score})$$

Score computation

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

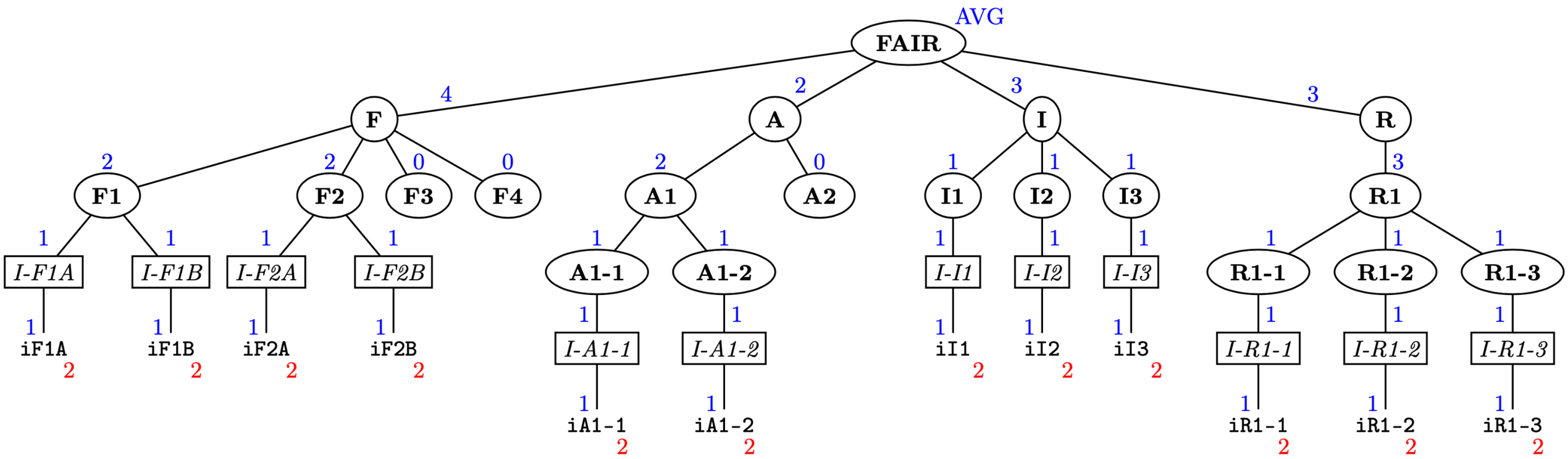
Weighting function: $w(n)$ *returns the weight of the node* n *compared to its siblings*

Aggregation function : weighted sum or weighted average to collect scores



$$\mathscr{M} = (\overbrace{V, E, FAIR}^{structure}, \underbrace{\Diamond, w, v_{max}}_{score}, D)$$

Model instantiation with FAIR-Checker

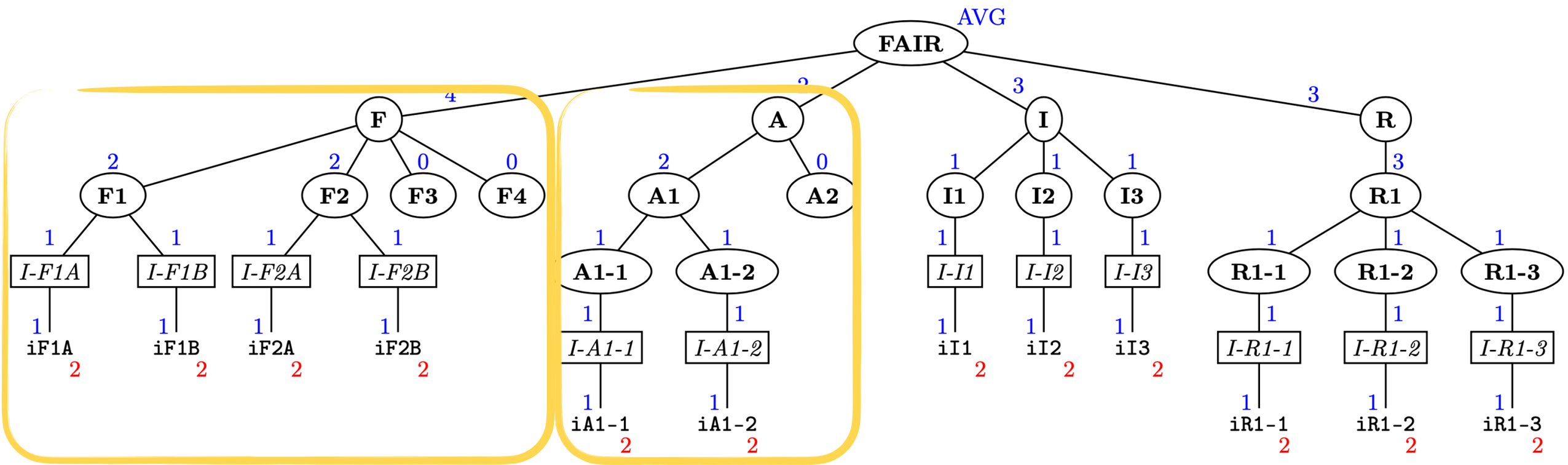


- The tool fits the tree-based structure {P,I,Imp}
- needs the maximum reachable score per implementation $\rightarrow 2$
- needs an aggregation function $\rightarrow$ *AVG (normalized scores)*
- needs cummulative weights for each principle $\rightarrow w \in [0,4]$

❷ 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* $\rightarrow$ *important granularity.*

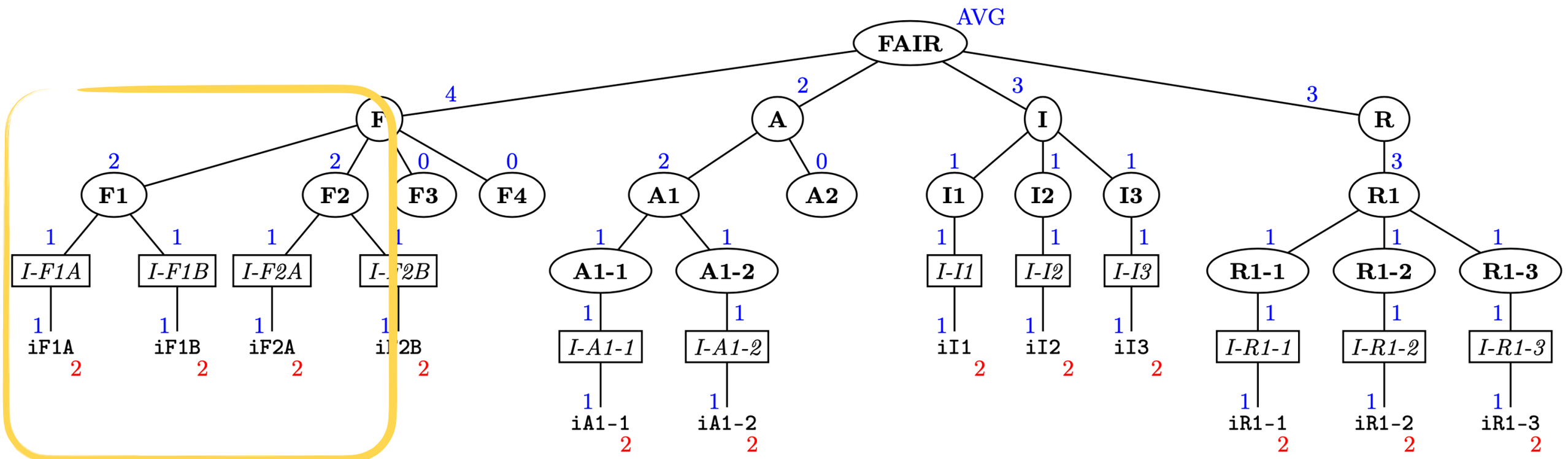


F has 4 indicators, A has only 2 $\rightarrow$ principles have not the same granularity

granularity(F) = 4/2 = 2 ; granularity(FAIR) = 12/10 = 1.2

Quantifying the ***coverage*** of a measure

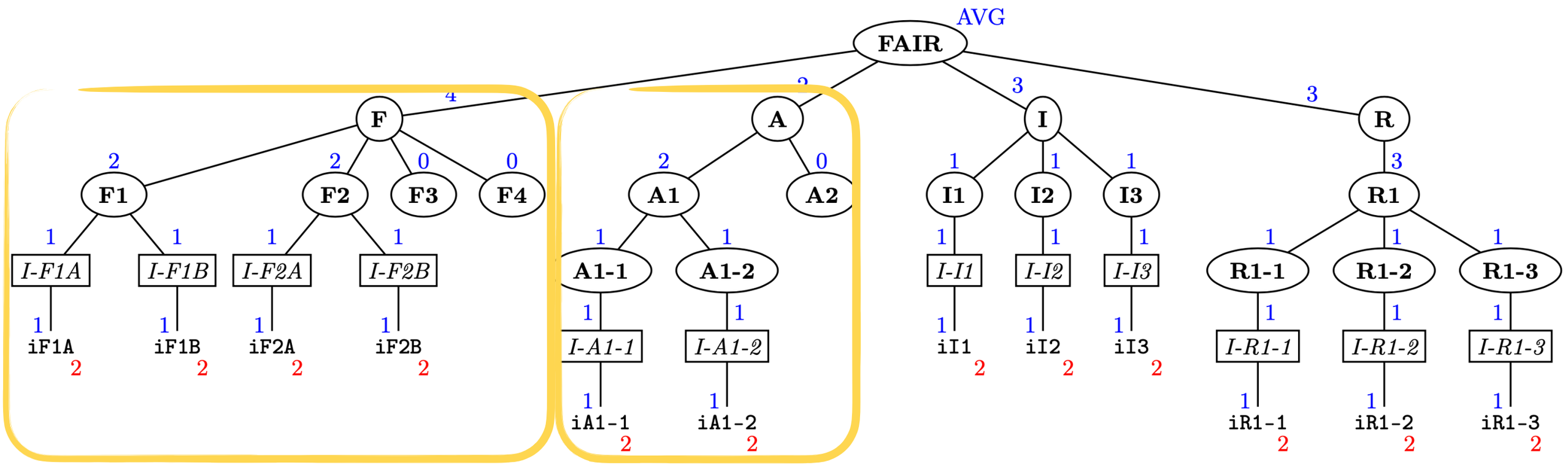
Idea: a principle is considered as covered if it exists at least one implementation for the principle or, its sub-principles.



F3 & F4 are never evaluated → F has 50% 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).



imp(A) = 2*2 / 12*2 = 1/6

imp(F) = 4*2 / 12*2 = 2/6

→ 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

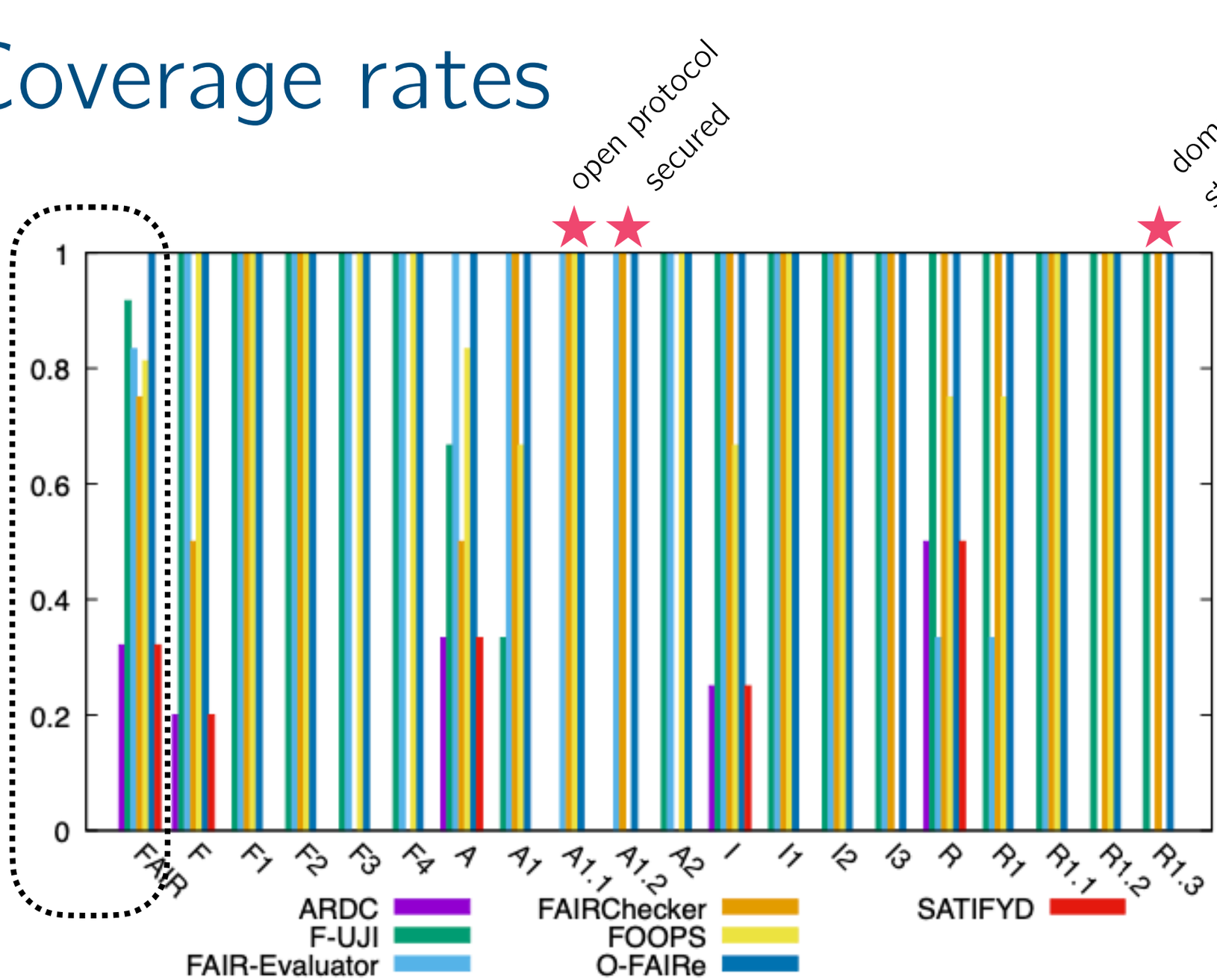
- 10 selected web resources for FAIR assessments, covering
  - datasets descriptions
  - ontologies
  - online courses
  - bioinformatics software
  - RDF files

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

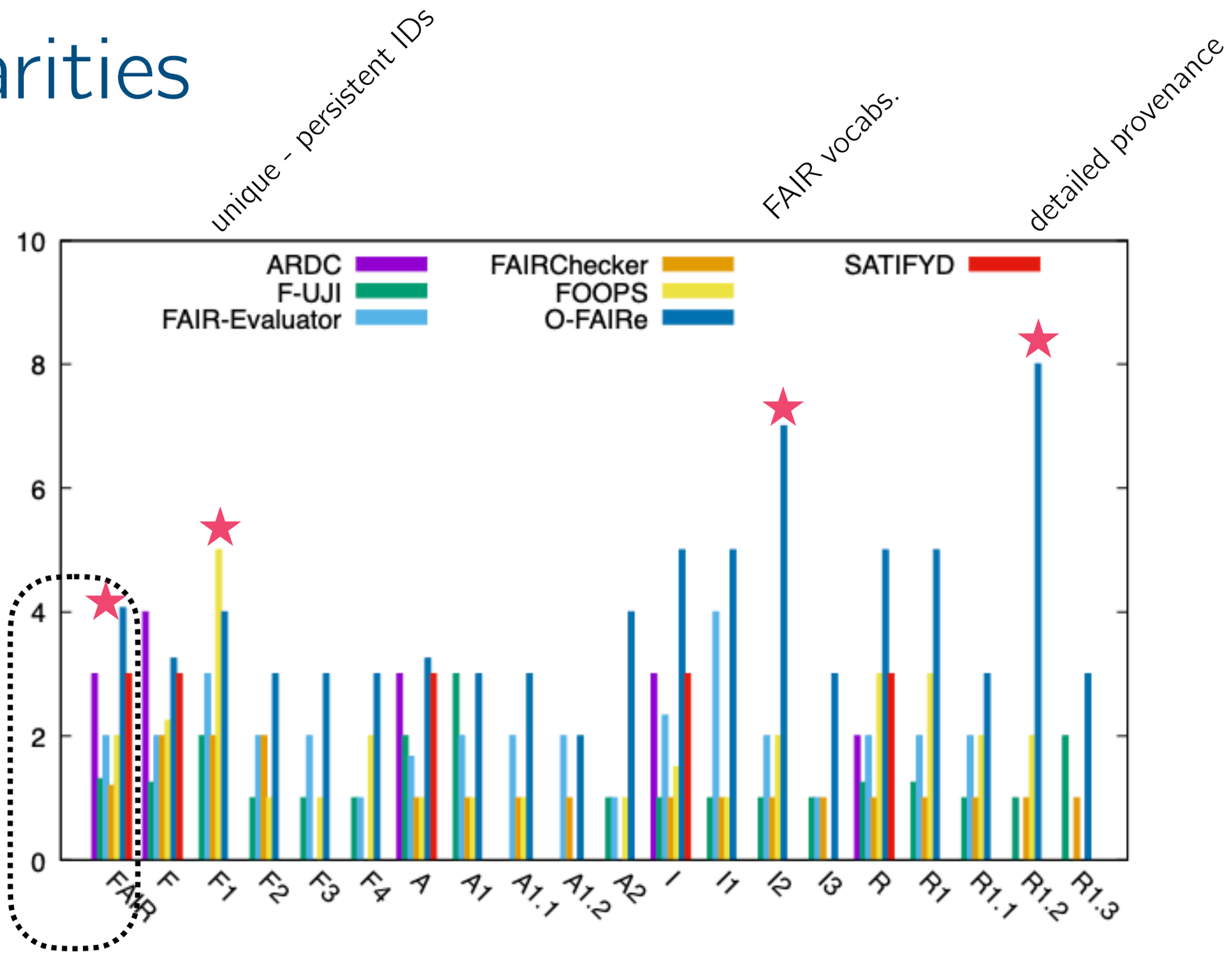
$\longrightarrow$ technical issues ?

$\longrightarrow$ interpretation ?

{O-FAIRe} $\longrightarrow$ the only one with a full coverage (1)

Low coverage for questionnaires (no sub-principles)

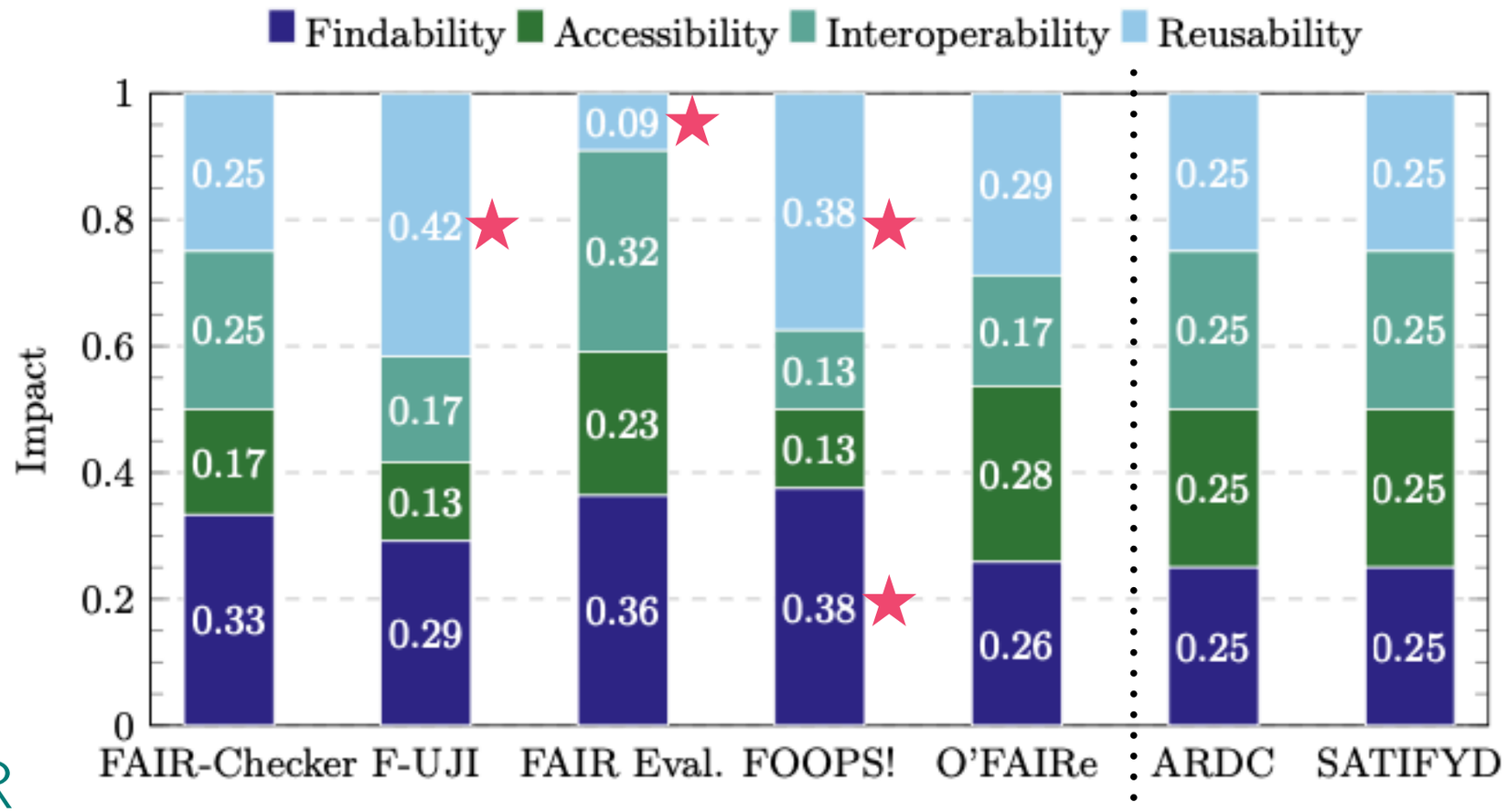
Granularities



- O-FAIRe (targeting ontologies) has the most fine-grained evaluation (e.g. I2 and R1.2) $\rightarrow$ great diversity in metadata for ontologies
- FOOPS deeply investigate F1 (identifiers)

Impacts

Are all principle equally contributing to the global FAIR assesment score ?
→ "No ..."



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 specifities of FAIR measurement tools
- We introduced a generic **model** for representing FAIR mesures 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

Jennien Andersen

Philippe Lamare

Sylvie Cazalens

contact

alban.gaignard@univ-nantes.fr

Funding:

ANR AAPG2019 DeKaloG

French institute for Bioinformatics (Elixir-FR)