An evaluation of semantic artefacts diversity in environmental sciences*

**The study has been accepted for publication in a peer-reviewed journal and should be published soon under the title "Assessing semantic interoperability in environmental sciences: variety of approaches and semantic artefacts"*

Cristina Di Muri[1], Martina Pulieri[2,3], Davide Raho[1], Alexandra N. Muresan[1], **Andrea Tarallo**[1*], Jessica Titocci[1], Enrica Nestola[1], Alberto Basset[2,4], Sabrina Mazzoni[2] and Ilaria Rosati[1,3]

[1]*Italian National Research Council (CNR), Research Institute on Terrestrial Ecosystems (IRET), Lecce, 73100, Italy*
[2]*University of Salento, Department of Biological and Environmental Sciences and Technologies, Lecce, 73100, Italy*
[3]*LifeWatch Italy, Lecce, 73100, Italy*
[4]*LifeWatch ERIC, Lecce, 73100, Italy*

FOIS 2024

JOWO WORKSHOP: FAIR PRINCIPLES FOR ONTOLOGIES AND METADATA IN KNOWLEDGE MANAGEMENT

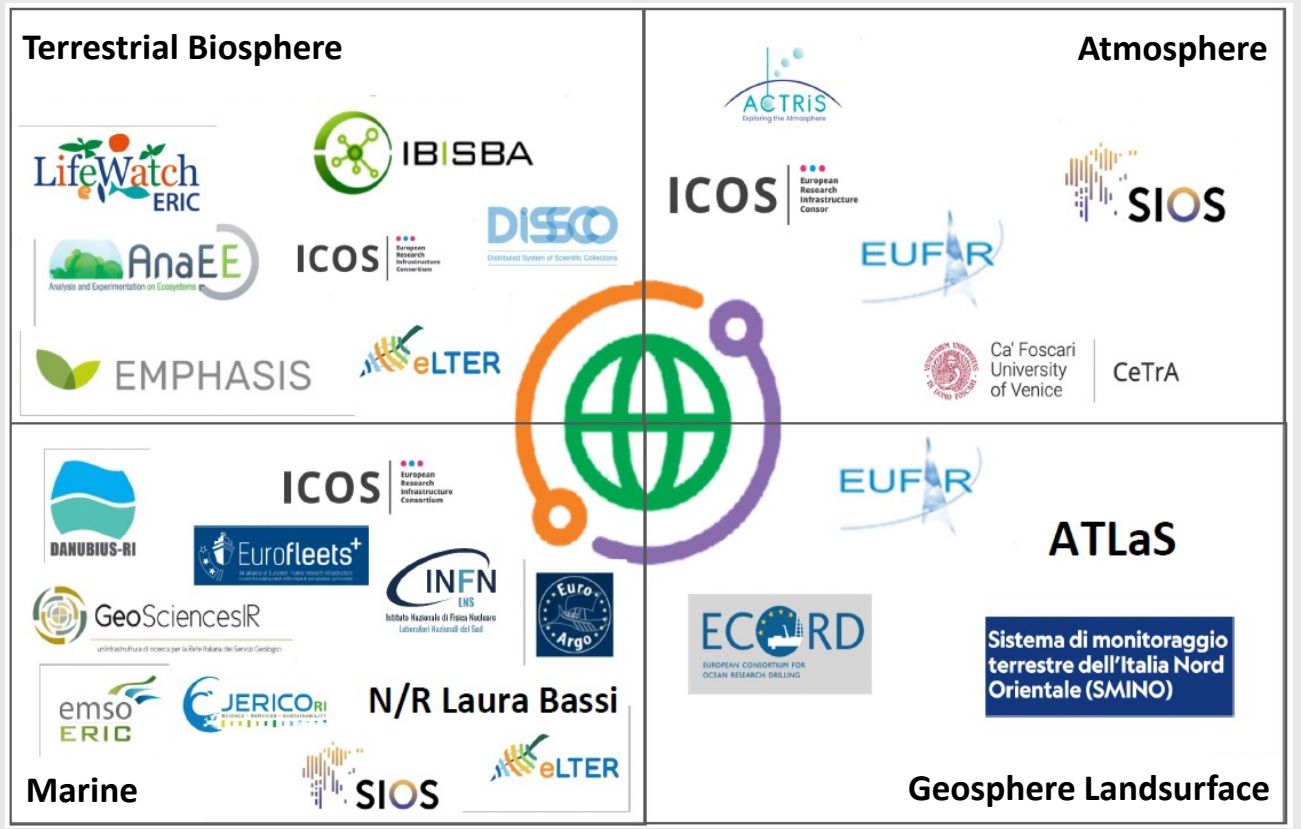
14TH INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS

15-19 JULY 2024 (ENSCHEDE, NETHERLANDS)

Context: ITINERIS
Italian Integrated Environmental Research Infrastructures System

ITINERIS https://itineris.cnr.it/

- 22 Research Infrastructures;
- 8 Work Packages;
- Scope : Italian Hub development;
- WP 2 : Access to facilities, FAIR data and services;
- Activity 2.4 : Semantic interoperability.



F A I R

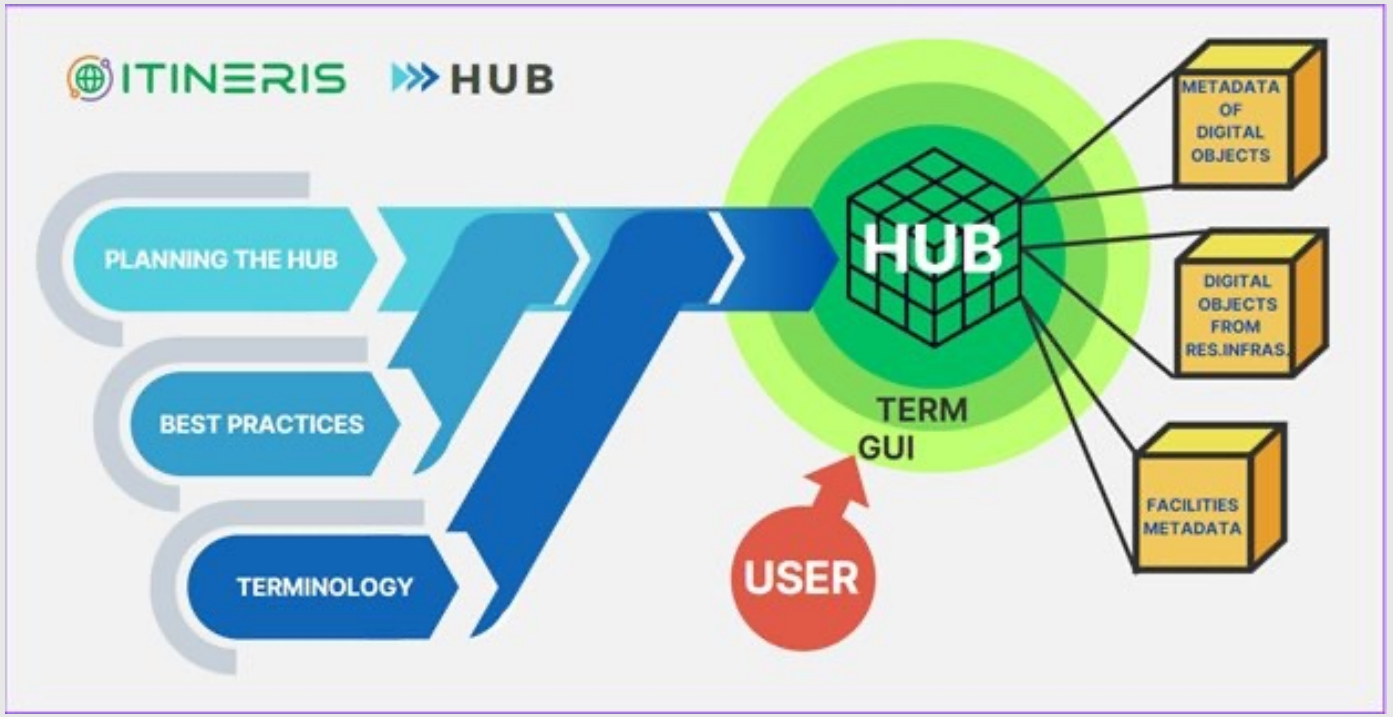
Findable Accessible Interoperable Reusable

Aim of the study

FAIR -> Data providers should use existing SA to describe their data.

The integration of Terminology Service (TS) in ITINERIS Hub

- The TS will query various catalogues of Semantic Artefacts (SA) in the Environmental Science domain.
- This promote reuse over creation: especially crucial for Research Infrastructures in the development phase.

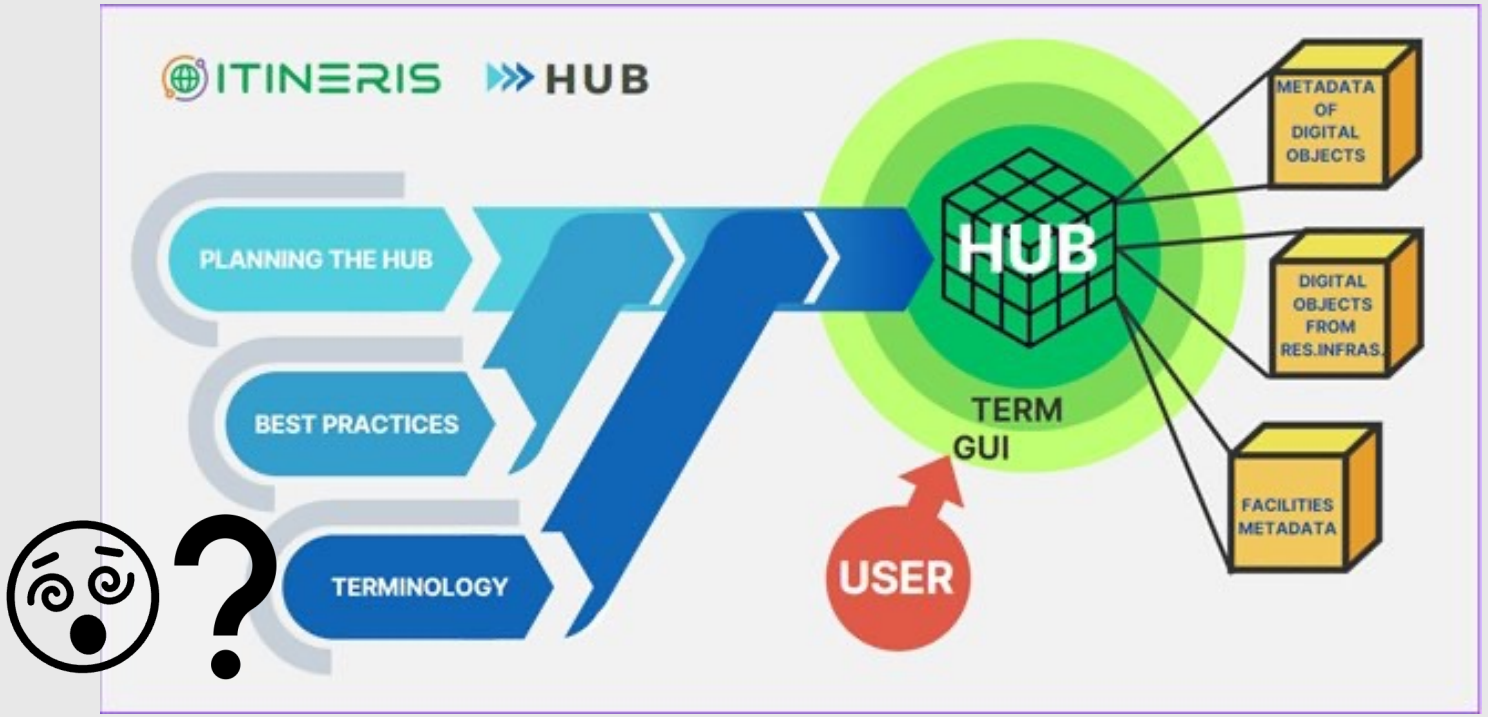


Aim of the study

FAIR -> Data providers should use existing SA to describe their data:

The integration of Terminology Service (TS) in ITINERIS Hub

- The TS will query various catalogues of Semantic Artefacts (SA) in the Environmental Science domain.
- This promote reuse over creation: especially crucial for Research Infrastructures in the development phase.

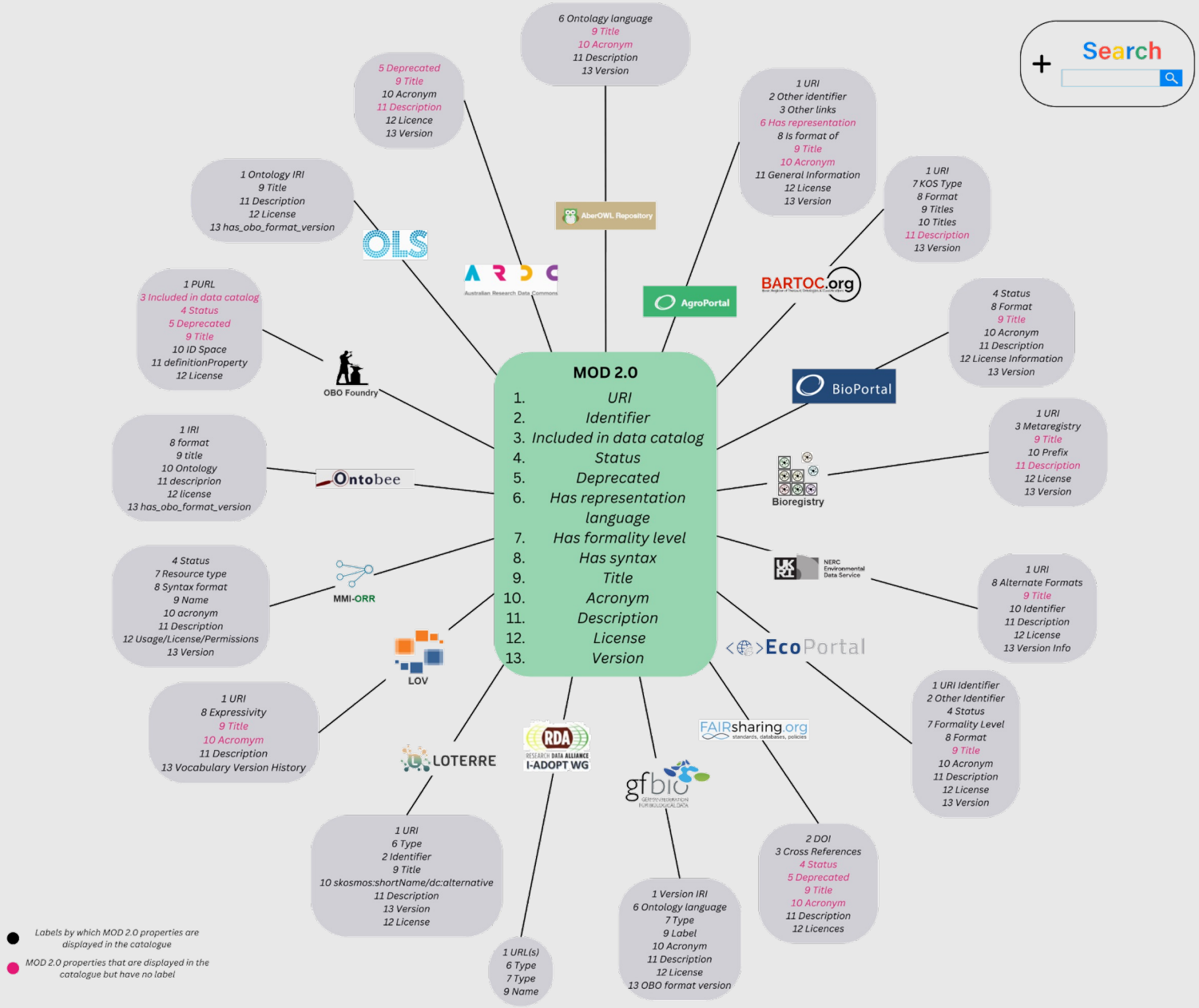


Collection of Semantic Artefacts:

- SA collection and cataloguing
- Aiming to create a comprehensive collection for easy access and reuse https://osf.io/axy3s/
- FAIR Analyses

- Selection of SA for ITINERIS
- FAIRification of SA not yet included in SA catalogs

Methods 1/4 - Collection and harmonisation

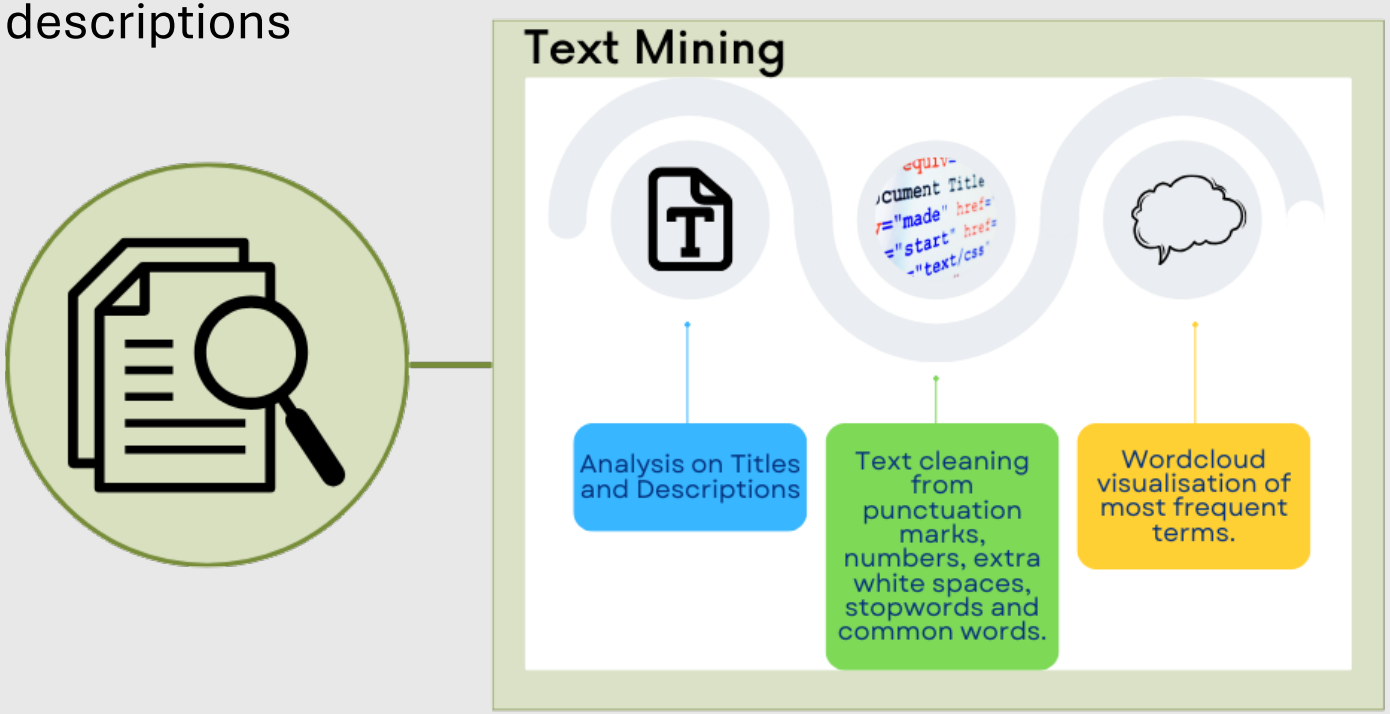


Methods 2/4 - Classification

Two methods used for SA classification in ITINERIS environmental domains

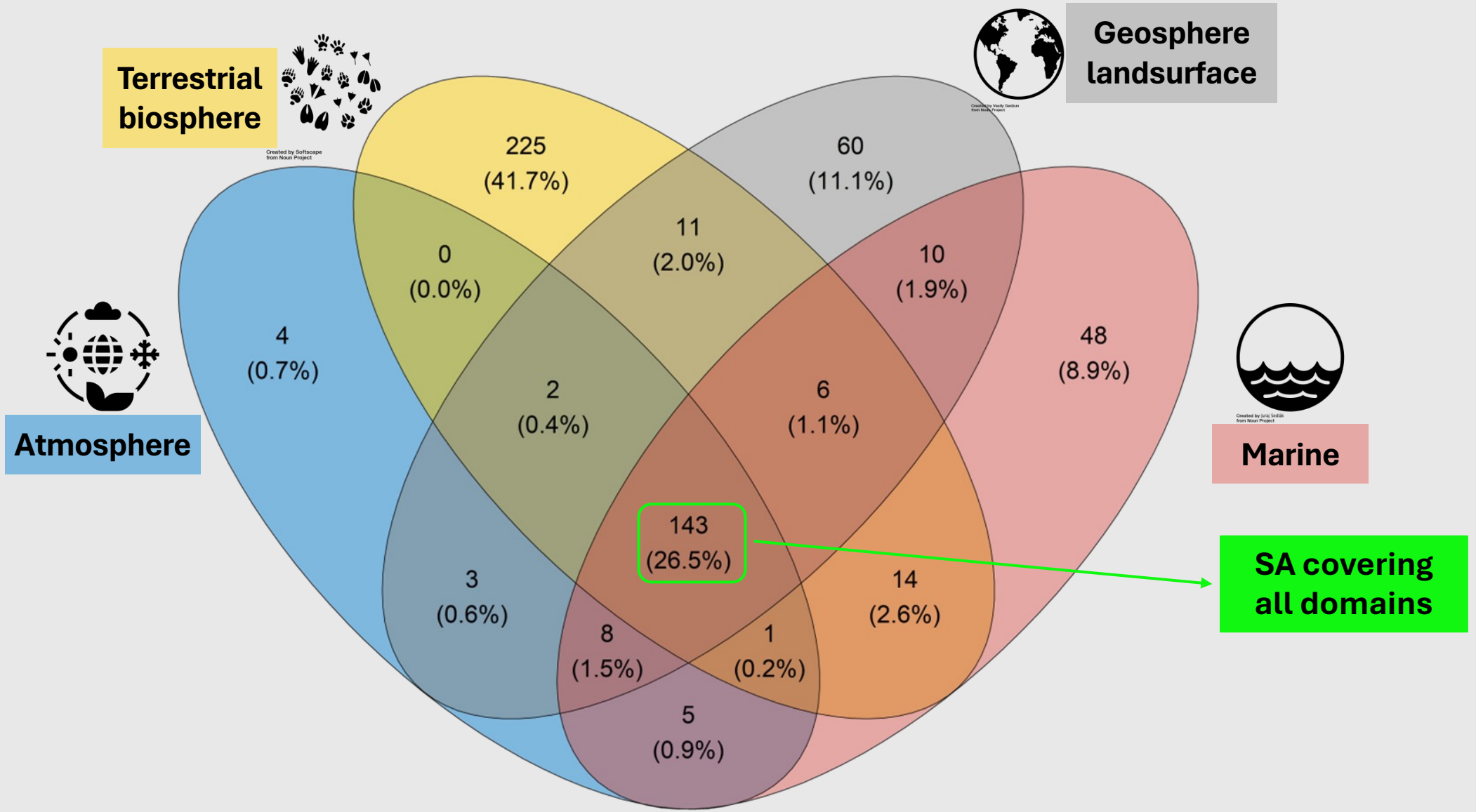
1. Categorisation based on terms/concepts/classes

2. Text mining analysis on titles and descriptions



Results: Classification - domain coverage

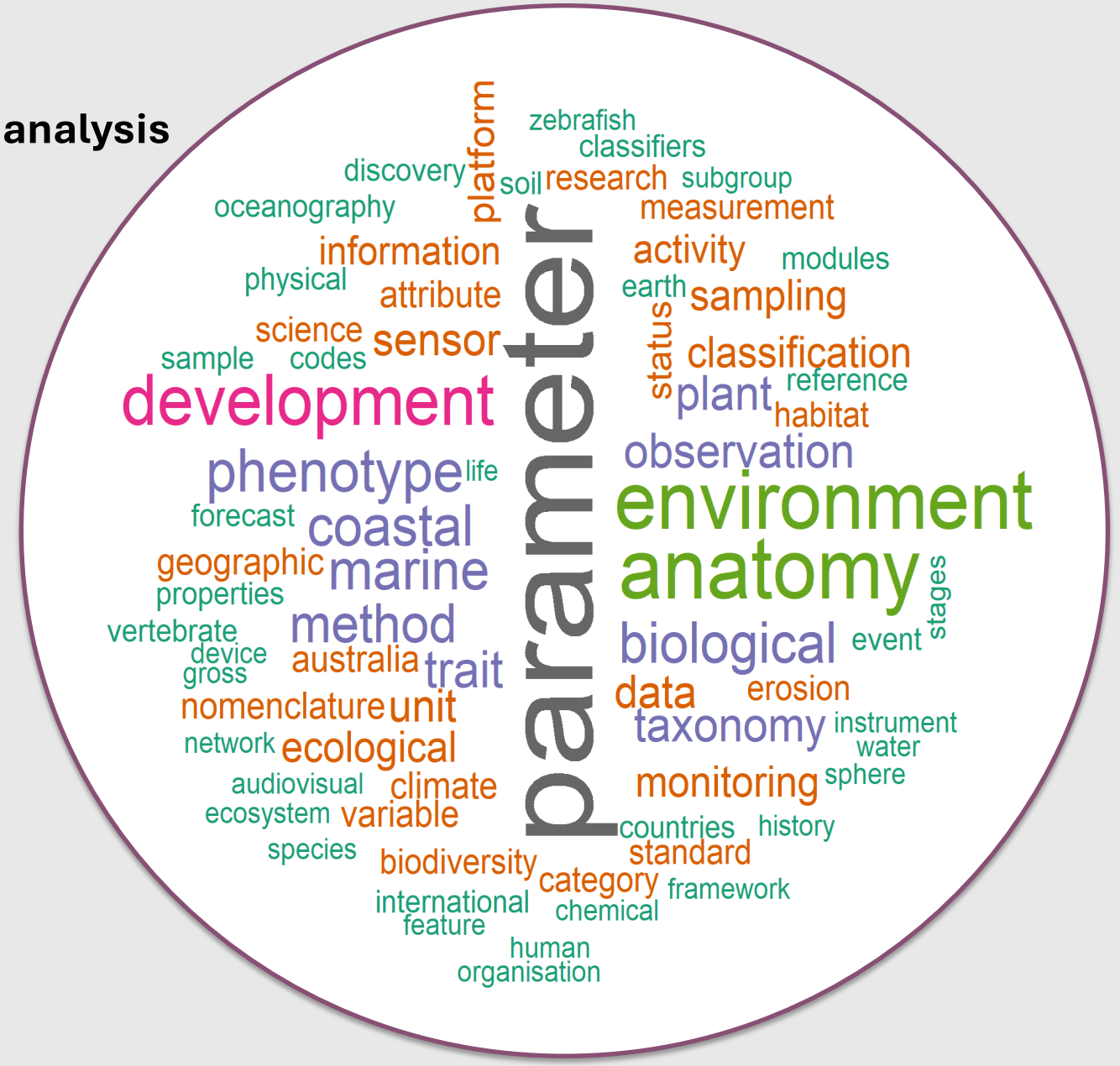
540 SA collected and classified according to ITINERIS environmental domains



Results: Classification - topics coverage

Most frequent words resulting from text mining analysis

- Parameter (freq = 43)
- Anatomy (freq = 24)
- Environment (freq = 23
- Development (freq = 19)
- Phenotype (freq = 15)
- Biological, Coastal & Marine (freq = 14)
- Method (freq = 13)
- Trait (freq = 12)
- Observation, Plant & Taxonomy (freq = 11)
- Data, Sensor & Unit (freq = 10)

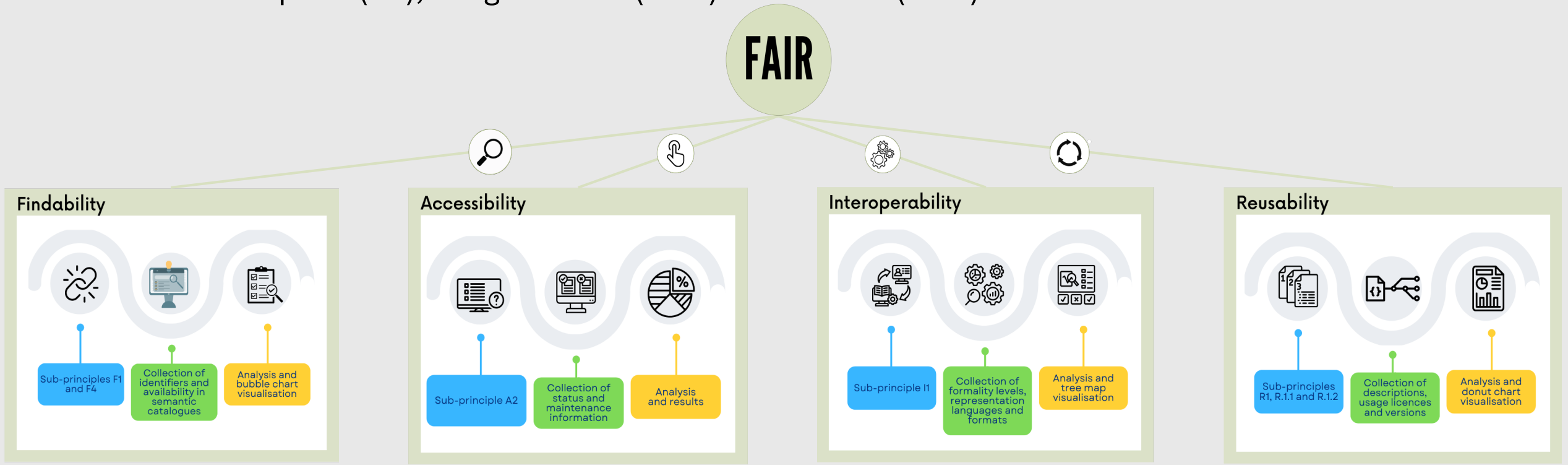


Methods 3/4 - FAIRNESS analysis

FAIRNESS analysis based on the collected metadata properties*

**Amdouni, E., & Jonquet, C. (2021). FAIR or FAIRer? An integrated quantitative FAIRness assessment grid for semantic resources and ontologies. In Research Conference on Metadata and Semantics Research.*

1. URI and identifiers (F1)
2. Inclusion in semantic catalogues (F4)
3. Status, version & maintenance (A2)
4. Formality level, language and format (I1)
5. Description (R1), usage licence (R1.1) and version (R1.2)



Results: FAIRNESS analysis – Findability (F1, F4)

F1: identifiers

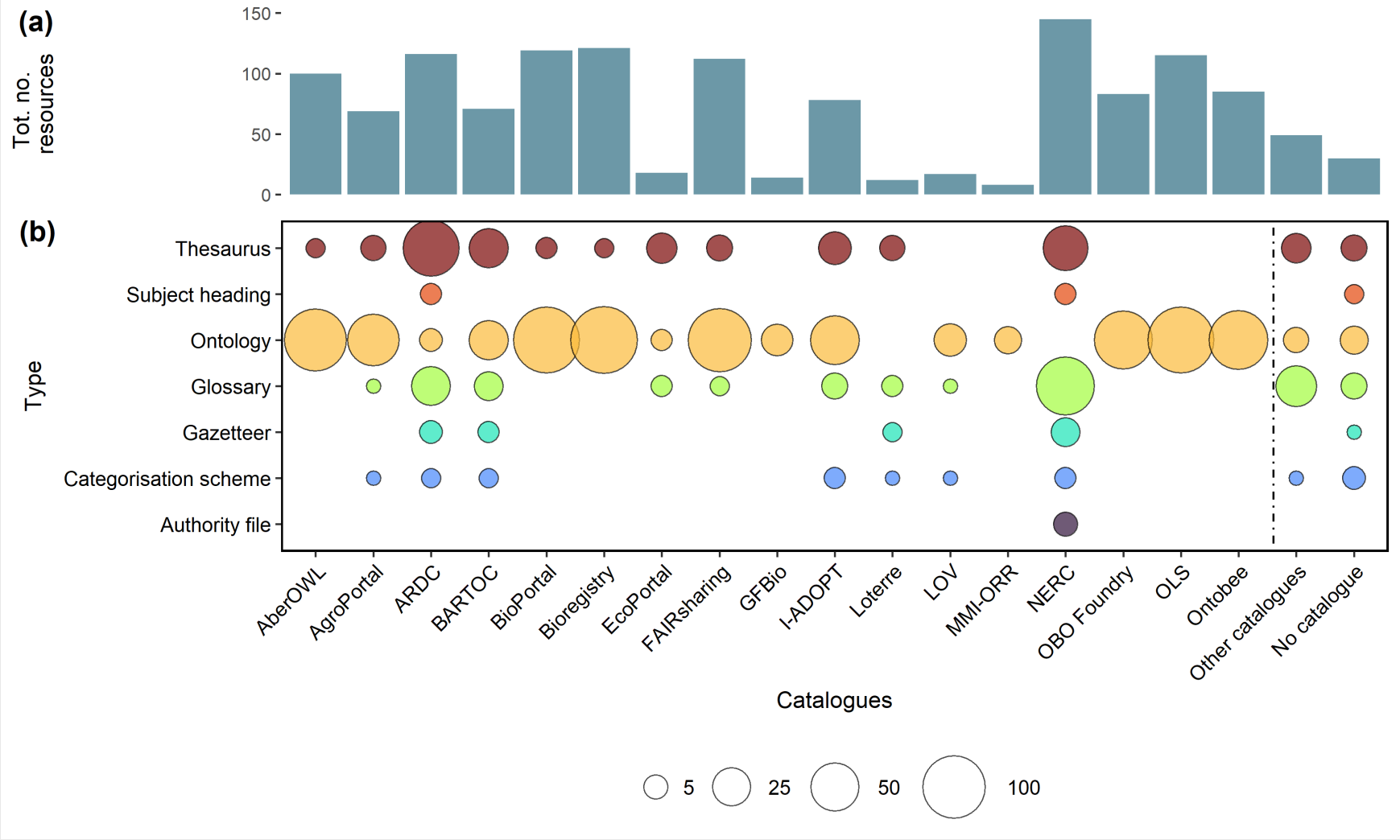
⇩

18,5% with DOIs

F4: Distribution of SA across catalogues

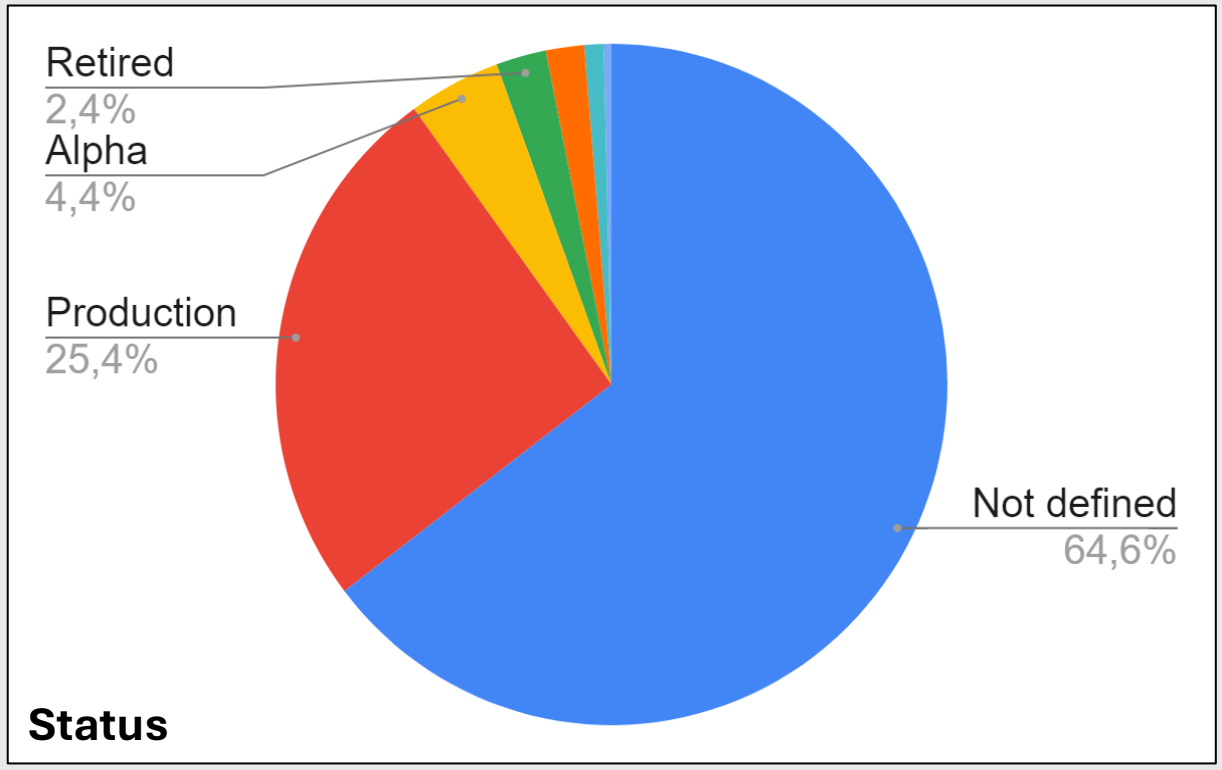
⇩

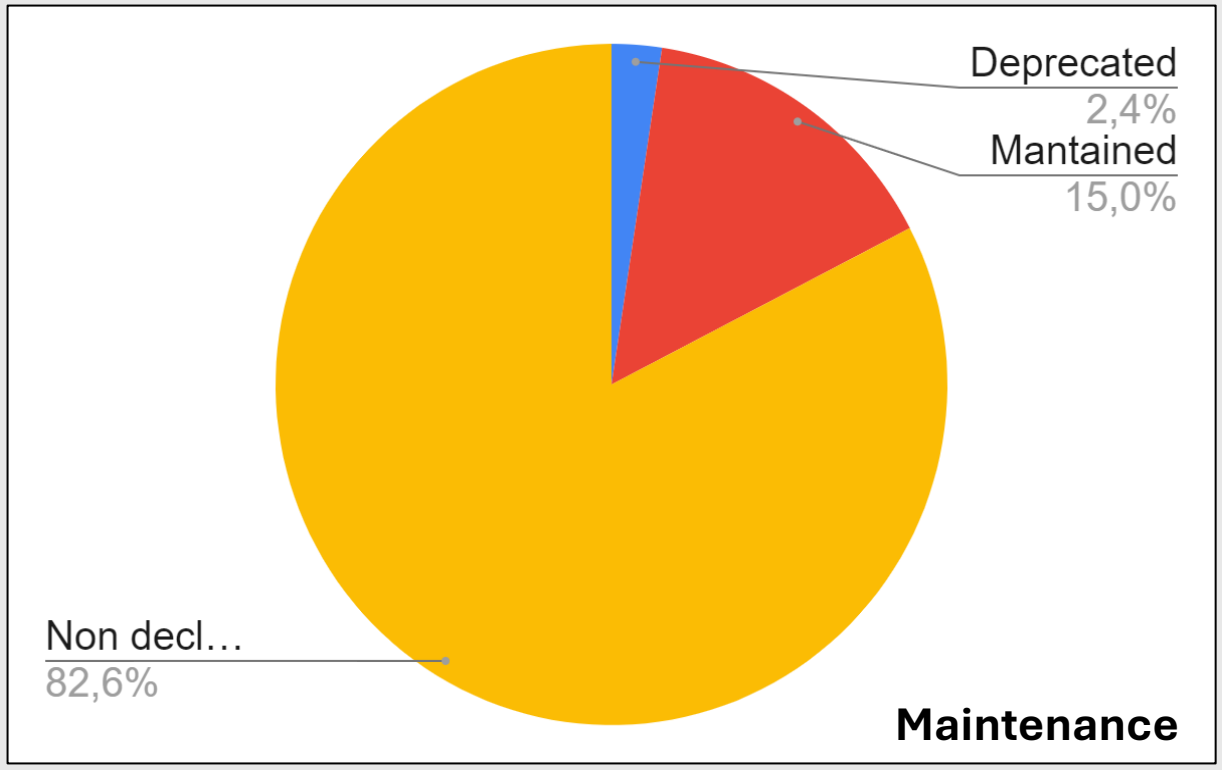
30 SA not available in semantic catalogues



Results: FAIRNESS analysis – Accessibility (A2)

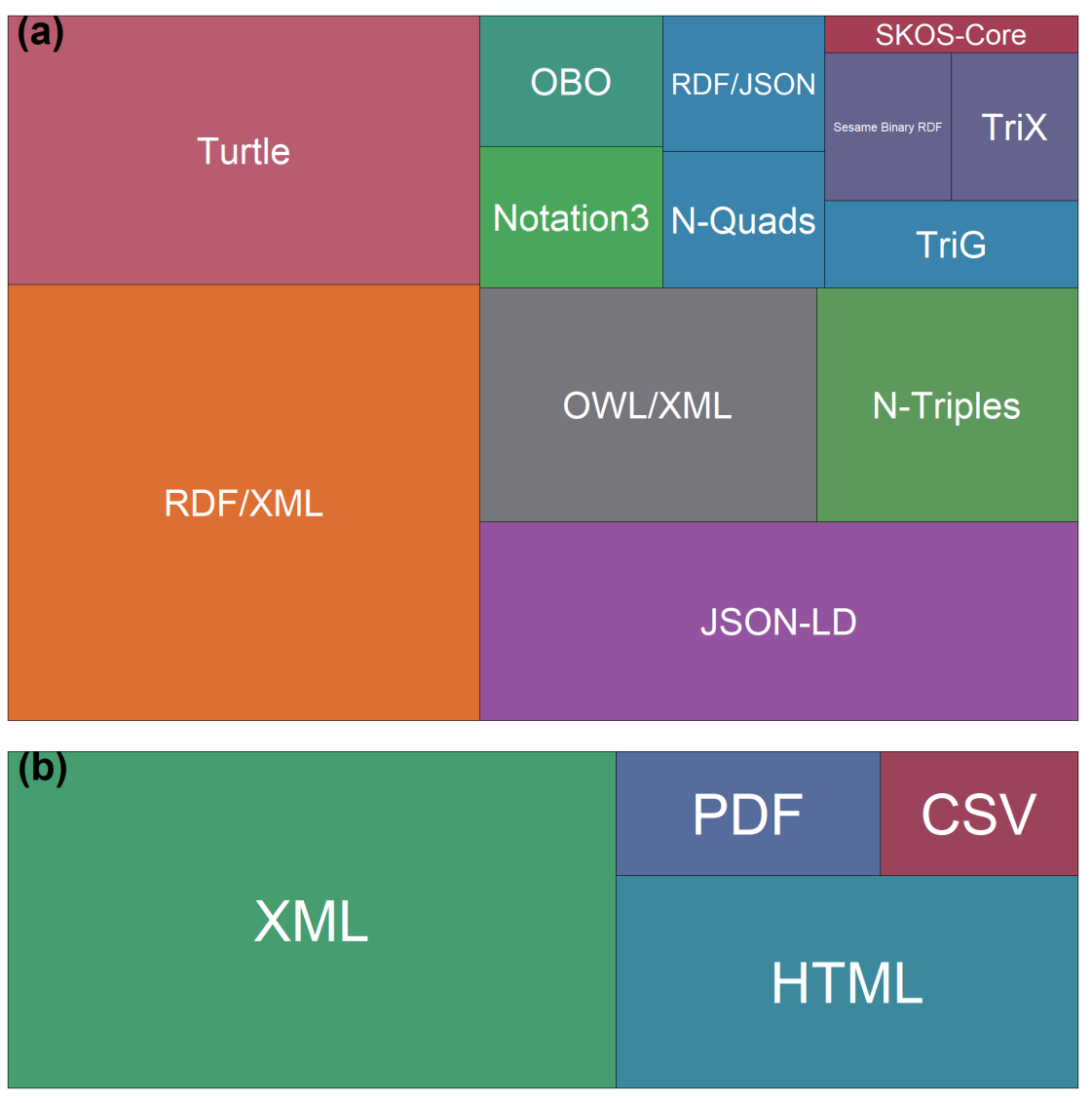
349 SA the status was not defined, whereas, in all other instances, the status was specified as being in "Production" (N = 137), "Alpha" (N = 24), "Retired" (N = 13), "Beta" (N = 10), "Inactive" (N = 5), and "Uncertain" (N = 2). Moreover, 13 SA were declared as no longer maintained (i.e., deprecated) and 81 as maintained. In 446 instances, the maintenance was not declared.





Results: FAIRNESS analysis – Interoperability (I1)

Serialisation formats



MACHINE ACTIONABLE

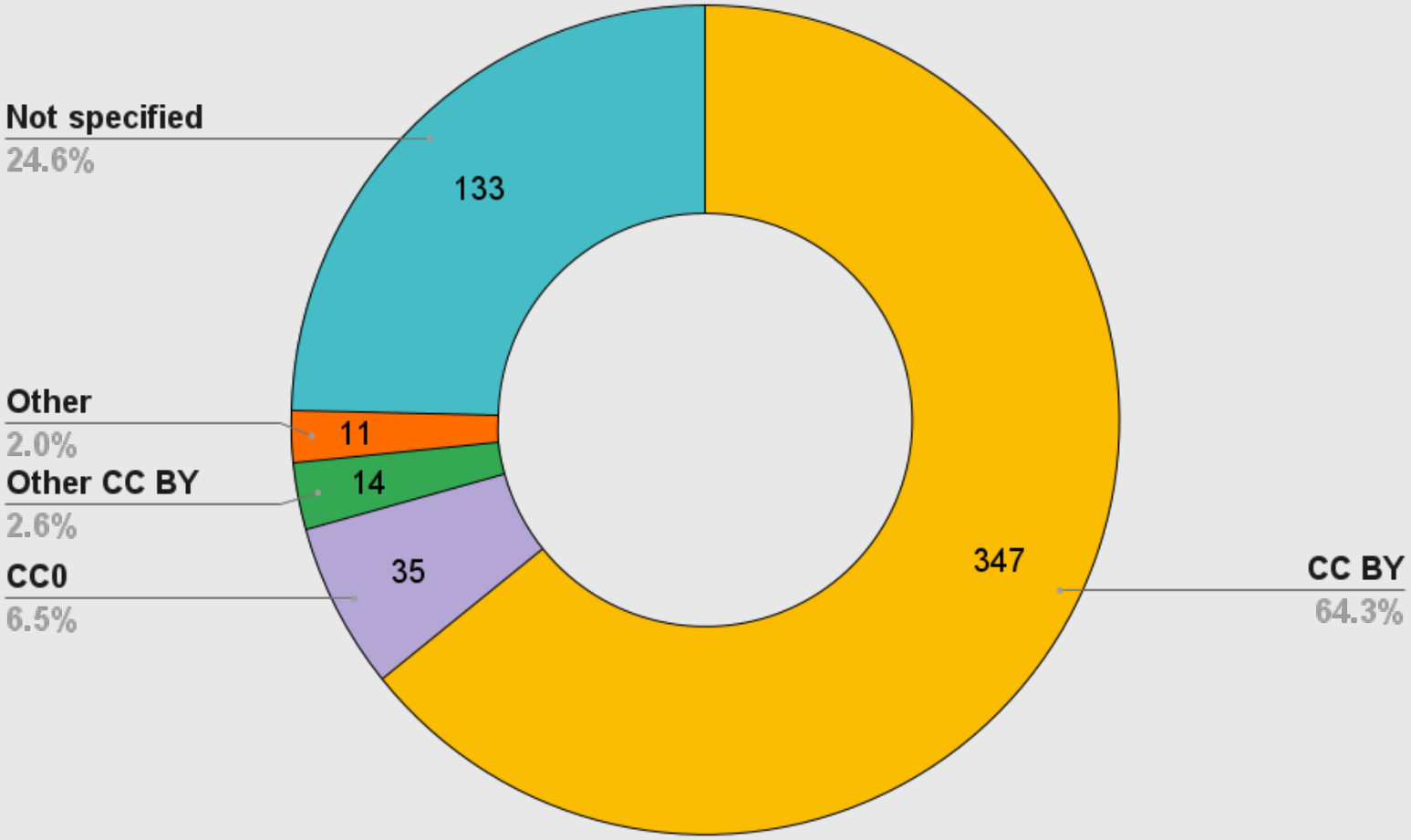
497 SA built using standard languages (OWL, SKOS, and RDFS) and serialisation formats.

MACHINE READABLE

43 resources built without standard languages.

Results: FAIRNESS analysis – Reusability (R1, R1.1, R1.2)

R1.1: Usage licences



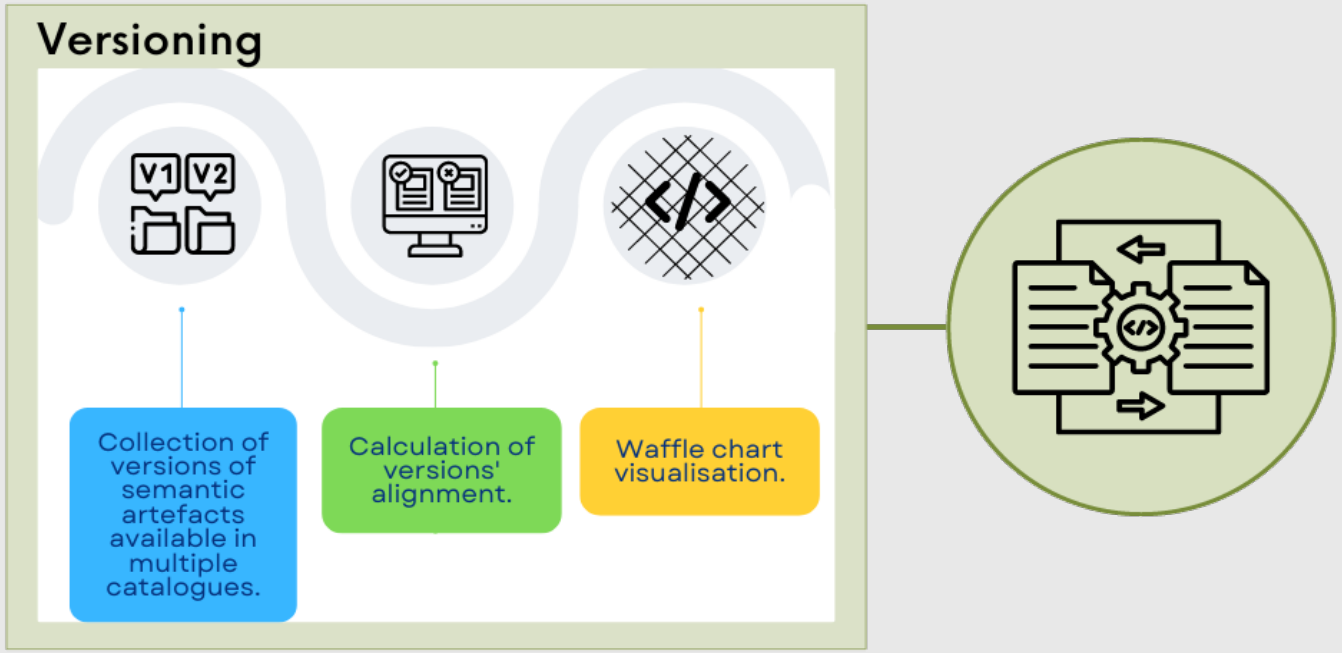
R1: Descriptions

8.9% without description

R1.2: Versions

22% missing version information and different formats were used (e.g. semantic versioning, dates, alphanumeric strings)

Methods 4/4 - Versioning analysis

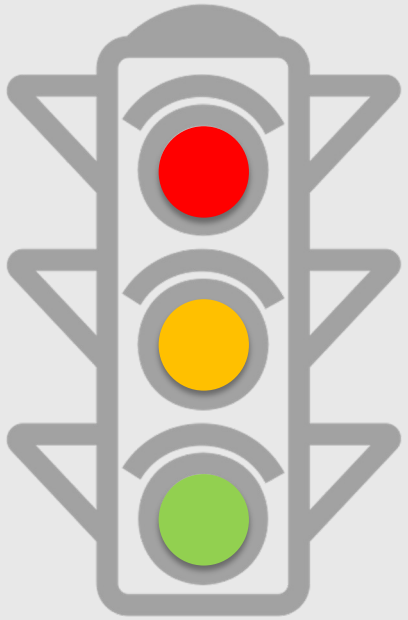


Version information was extracted for each SA and catalogue in which the resource was available

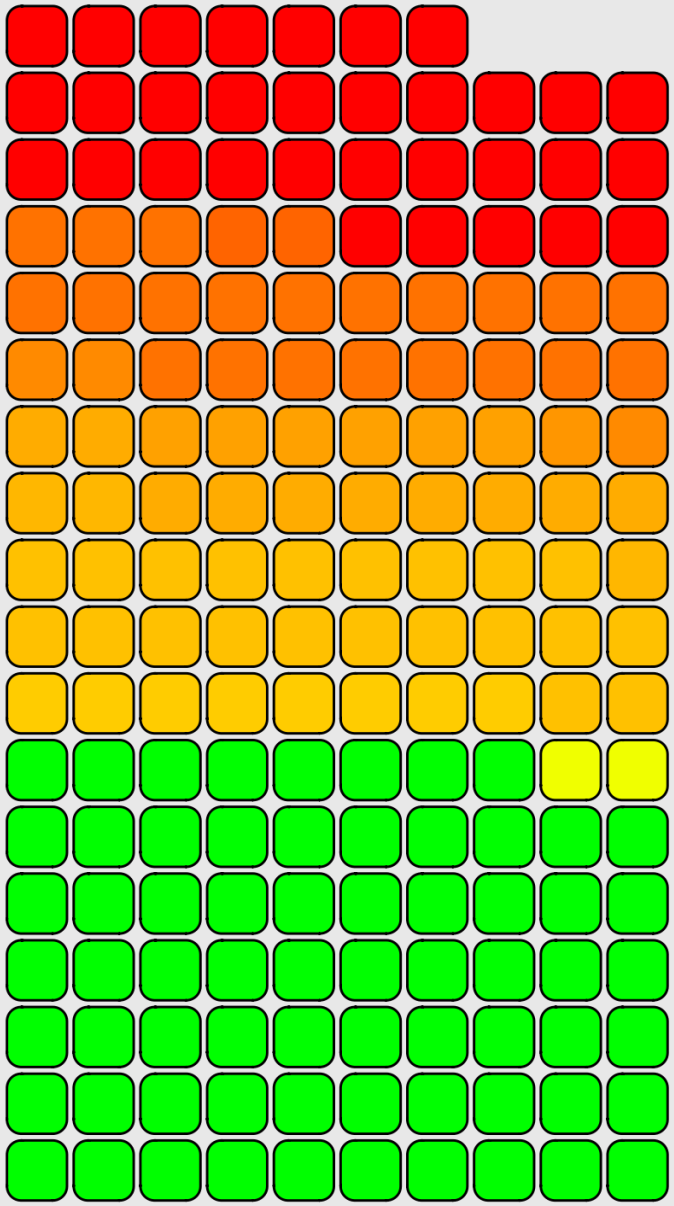
Only SA shared between at least two catalogues were considered (N = 177)

$$\text{Version misalignment} = \frac{\text{N of unique values}}{\text{N of catalogues}}$$

Results : Versioning analysis



Created by Darwin Mulya
from Noun Project



61.8% of SA
With versions not aligned among catalogues.

1 **Complete Misaligment**
0.75
0.67
0.6
0.57
0.5
0.4
0.38
0.33
0.29
0.25
0 **Complete Aligment**

Conclusion and future perspectives

- Despite the numerous recommendations, still some SA do not adhere to FAIR principles.
- Alignment is required between the SA available across multiple catalogues - not only versioning! PIDs, different acronyms, etc.

The terminology service will possibly offer a solution to some of these issues (eg versioning discrimination, mapping between terms)

Further analyses are needed: mapping between resources (presence of duplications without mapping, etc.), the analysis of the syntax of identifiers and their resolvability, and other aspects of FAIRness that we have not considered.

Thank you :)

andrea.tarallo@cnr.it

The publication has been funded by EU - Next Generation EU Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures” - Project IR0000032 – ITINERIS - Italian Integrated Environmental Research Infrastructures System - CUP B53C22002150006. The authors acknowledge the Research Infrastructures participating in the ITINERIS project with their Italian nodes: ACTRIS, ANAEE, ATLaS, CeTRA, DANUBIUS, DISSCO, e-LTER, ECORD, EMPHASIS, EMSO ,EUFAR ,Euro-Argo, EuroFleets, Geoscience, IBISBA, ICOS, JERICO, LIFEWATCH, LNS, N/R Laura Bassi, SIOS, SMINO. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor European Commission can be held responsible for them.

FOIS 2024

CNR IRET

ITINERIS

JOWO WORKSHOP: FAIR PRINCIPLES FOR ONTOLOGIES AND METADATA IN KNOWLEDGE MANAGEMENT

14TH INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS

15-19 JULY 2024 (ENSCHEDE, NETHERLANDS)