

Towards FAIR metadata standards in FAIRsharing.org

Alejandra Gonzalez-Beltran, PhD

Inaugural Workshop Metadata 4 Machines (M4M)
GO-FAIR initiative

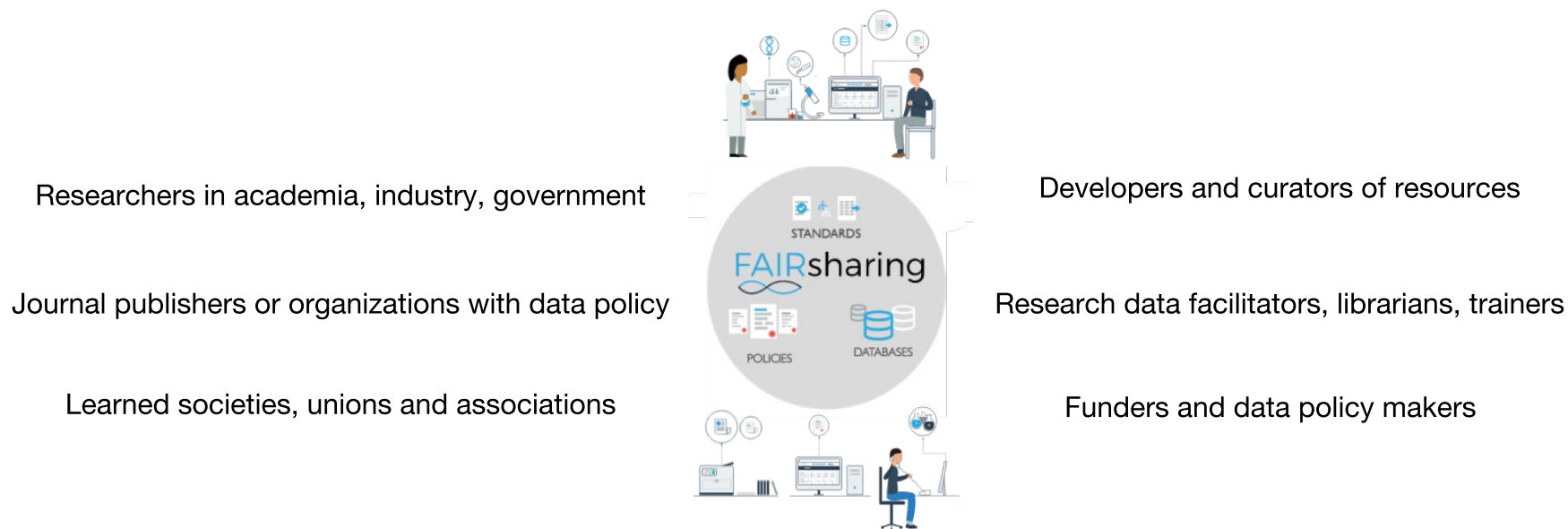
15th October 2018, Leiden, Netherlands



FAIRsharing.org

A curated, informative and educational resource on data and metadata *standards*,
inter-related to *databases* and data *policies*

We guide **consumers** to find, select and use these resources with confidence,
and **producers** to make their resource more discoverable, more widely adopted and cited



Pre-print at: <https://doi.org/10.1101/245183>

Authored by 68 stakeholders in academia, industry, funding agencies,
standards organisations, infrastructure providers and scholarly publishers

also part of:



FAIRsharing and the FAIRmetrics

- FAIRsharing works with the FAIR metrics WG to serve as:
 - **Registry** to describe digital assets, such as databases/repositories, standards, policies, enhancing their discoverability (schema.org) and citability (DOIs)
 - **Look up service** for identifier schemas and standards (phase 1: now)
 - **Validation service** against metadata standards (phase 2: planned)

Comment | [OPEN](#) | Published: 26 June 2018

A design framework and exemplar metrics for FAIRness

Mark D. Wilkinson , Susanna-Assunta Sansone , Erik Schultes, Peter Doorn, Luiz Olavo Bonino da Silva Santos & Michel Dumontier 

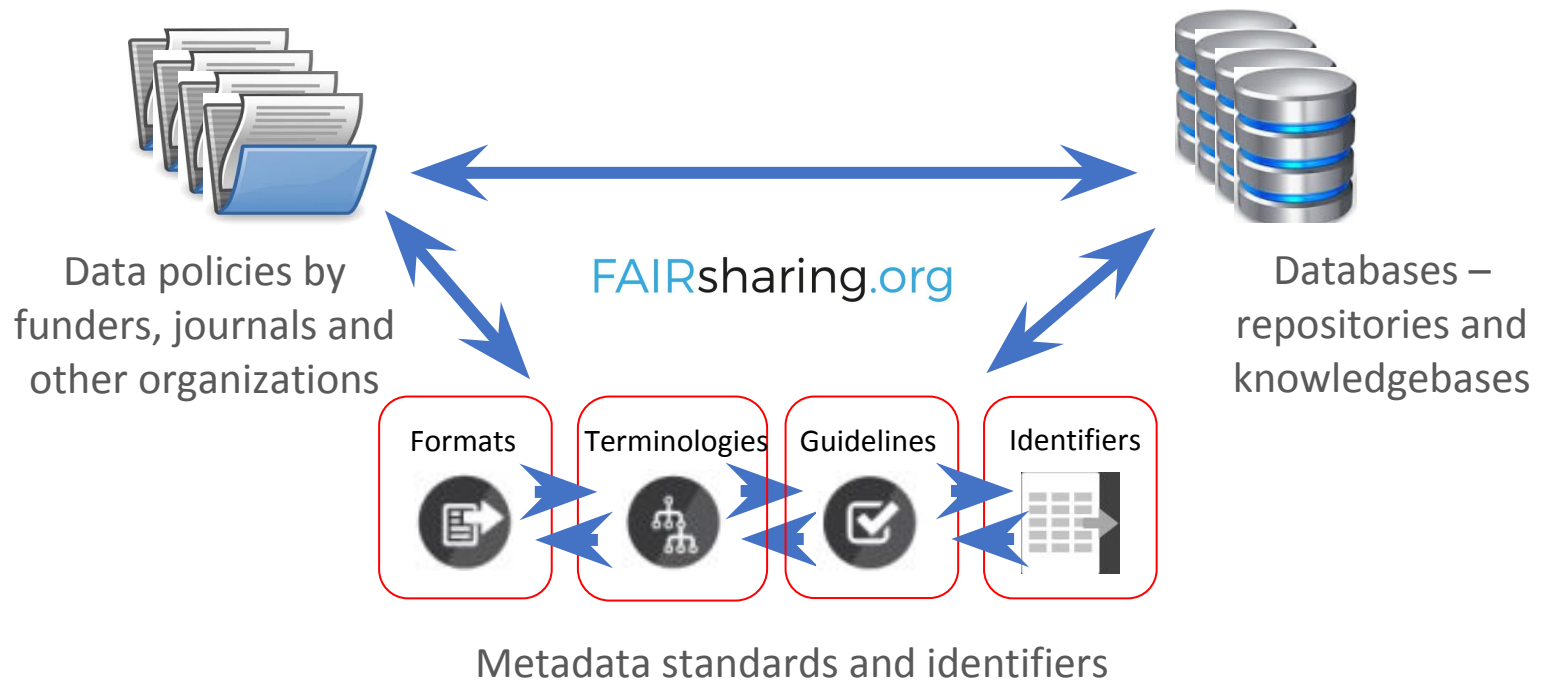
Scientific Data **5**, Article number: 180118 (2018) | [Download Citation](#) 

FAIRmetrics.org

also part of:

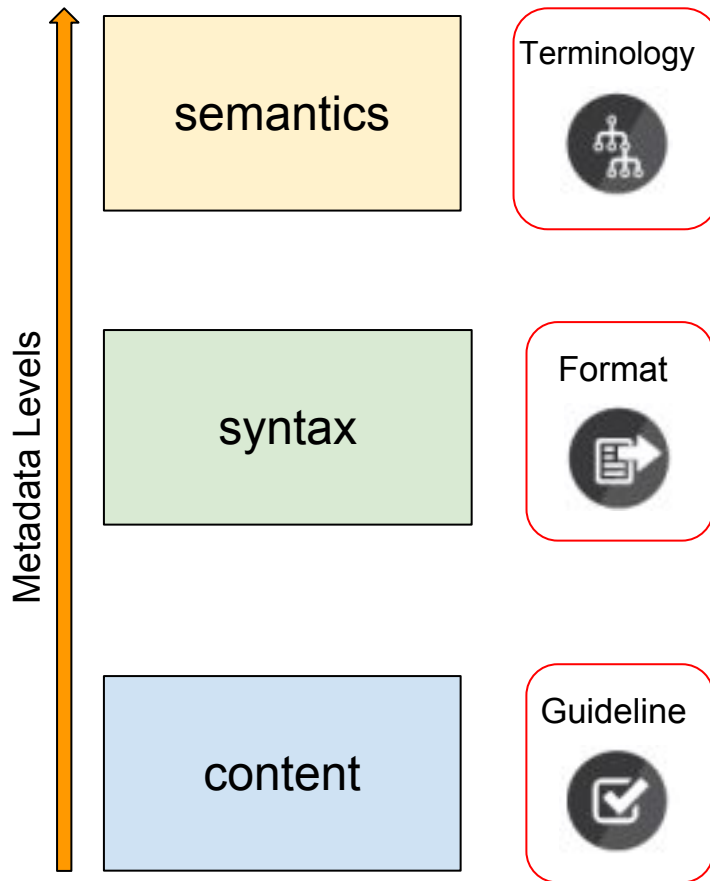


Mapping a complex and evolving landscape...



Metadata standards in practice

Databases/
Repositories



Metadata standards in practice

Databases/
Repositories



Metadata Levels

semantics

Terminology



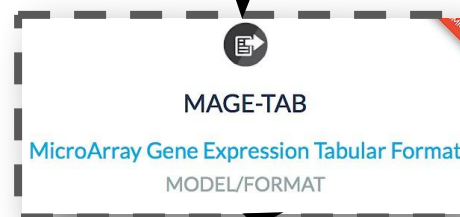
syntax

Format



content

Guideline



Reporting guidelines

a.k.a. Minimum Information Checklists

https://fairsharing.org/standards/?q=&selected_facets=type_exact:reporting%20guideline

- There are **151** reporting guidelines registered in FAIRsharing.org
 - The majority of them are described through manuscripts / documents (usually PDF or Word files)
 - Only **some** of those reporting guidelines have been described in a machine readable and verifiable way (e.g XML schema, YAML, JSON schema) e.g. **MIBiG, MIAPTE, MICEE, MIDE, MIABIS, MIACA, MIACME, AIRR**
 - Some guidelines have been implemented in different formats and those need to be taken into account

Towards **FAIR** metadata standards

Machine actionable metadata standard components:

- Facilitate the creation of compliant metadata through **automated templates**
- Automate verification of compliance with the guidelines through **automated validation** (a.k.a. **FAIR metric**)
- **Automated** generation of **documentation** about the guidelines
- **Automated syntactic** and **semantic comparison** between reporting guidelines

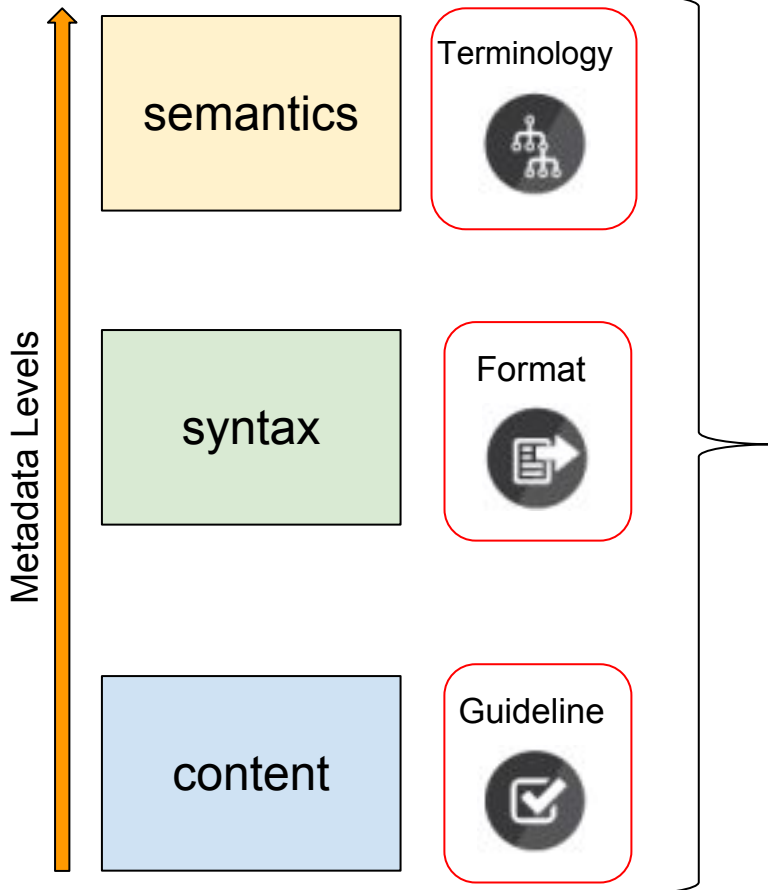
Reusable (modular) metadata standard components:

- **Automatic identification** of common **reusable components** and differences
- **Automatic identification** of **dependencies** between guidelines
- Automated creation of **templates** that comply with a two or more reporting guidelines (**automated merge**)



**FAIR metadata
representation**

Support for
metadata integration

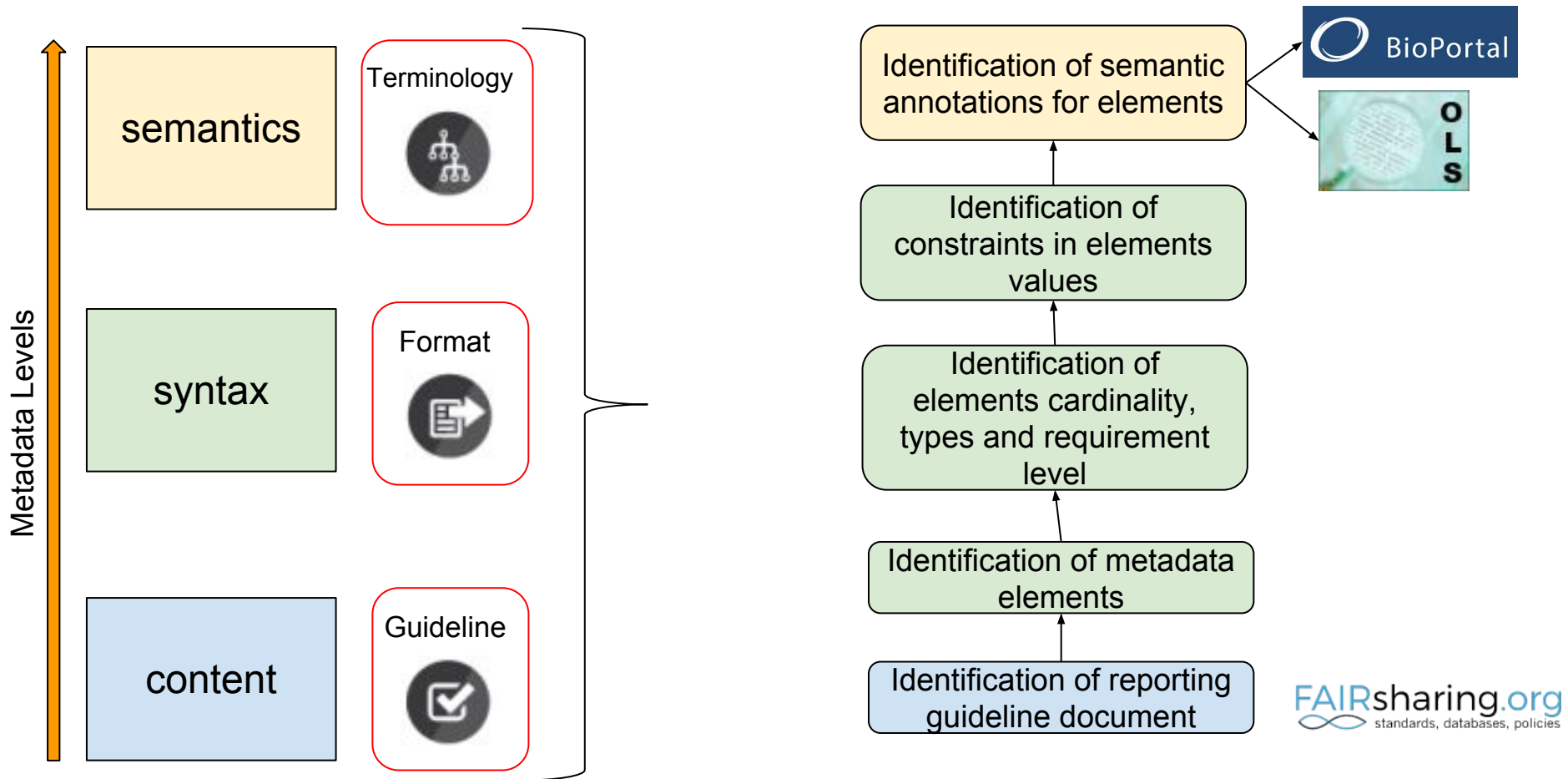


**Actionable
and modular**
representation of
content, syntax and
semantics

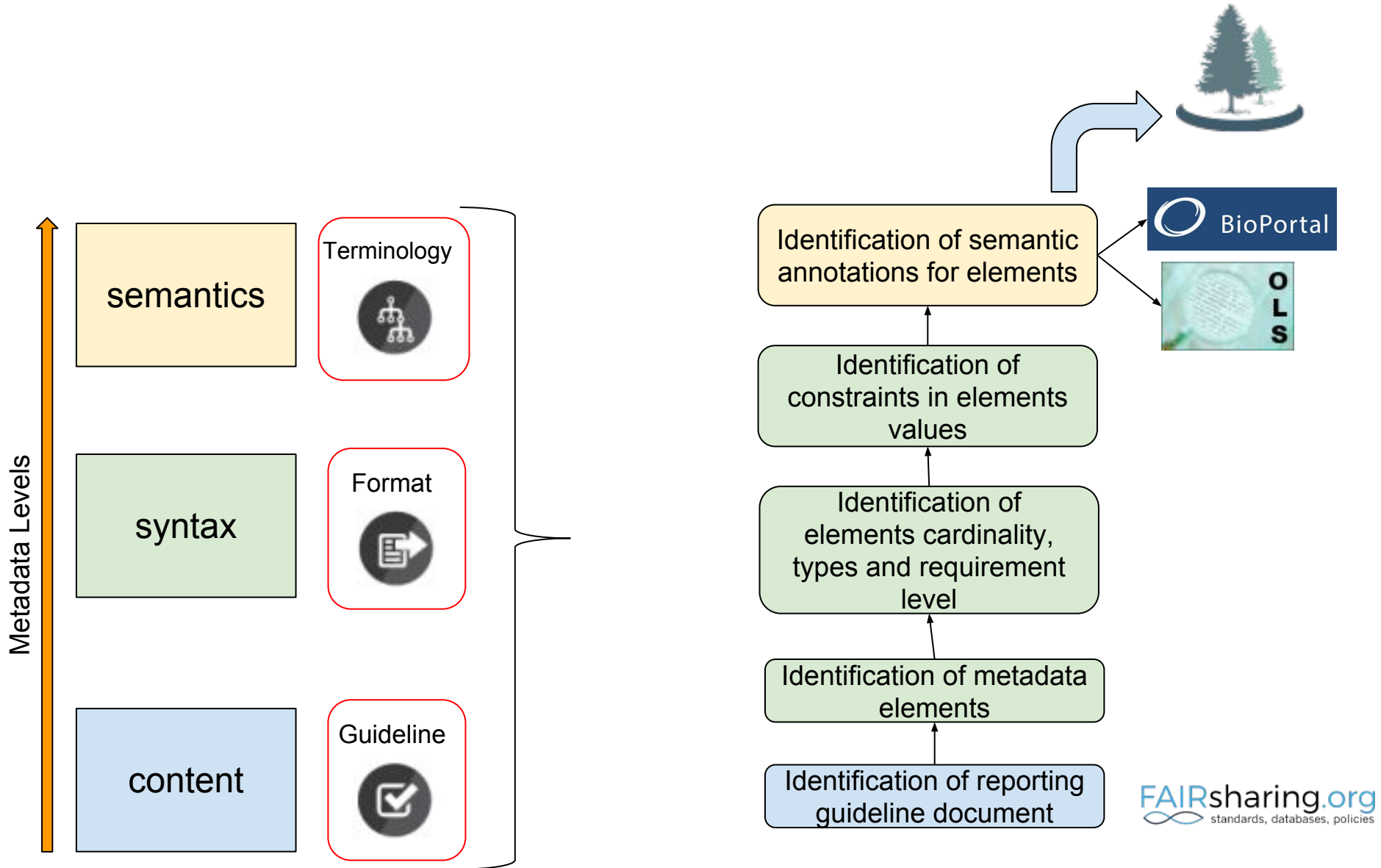
automated ...

- ...templates**
- ...validation**
- ...documentation**
- ...comparison**
- ...identification of common components**
- ...identification of dependencies**
- ...merge (for compliance with various standards)**

Standards FAIRification Method



Standards FAIRification Method



Gonzalez-Beltran, A - FAIR metadata standards - October 2018

Automating standards documentation

<https://github.com/FAIRsharing/JSONschema-documenter>



MIACA (Minimum Information about a Cellular Assay) schema

Schema Metadata

- **\$schema** : <http://json-schema.org/draft-04/schema>
- **id** : <https://w3id.org/mircat/miaca/schema/schema.json>
- **title**: MIACA (Minimum Information about a Cellular Assay) schema
- **description**: JSON-schema representing MIACA reporting guideline.
- **type**: object
- **provenance**:
 - **url** <http://w3id.org/mircat/miaca/provenance.json>

Schema Fields

@context

Description: The JSON-LD context
Expected types (any number of types from below):

- **string**
- **object**
- **array**

@id

Description: The JSON-LD identifier
Expected type : **uri**

@type

Description: The JSON-LD type
Expected type : **uri**

project

Cardinality: (0 ... 1)
Expected type:

project_schema

MIACA (Minimum Information about a Cellular Assay) project schema

Schema Metadata

Automating the **syntactic** and **semantic** comparison between checklists



<https://github.com/FAIRsharing/JSONschema-compare-and-view>

	MIACA	MIACME
sdo:CreativeWork	MIACA (Minimum Information about a Cellular Assay) schema	MIACME (Minimum Information about a Cell Migration Experiments) schema
sdo:Person	MIACA Source - corresponding to a person and their organization	MIACME person schema. A human being.
sdo:addressCountry	country	×
sdo:email	email	email
sdo:identifier	ID	identifier
sdo:name	name	×
sdo:address	address	address
sdo:affiliation	institution	affiliation
sdo:lastName	×	lastName
sdo:faxNumber	×	fax
sdo:telephone	×	phone
sdo:givenName	×	firstName
sdo:Project	MIACA (Minimum Information about a Cellular Assay) project schema	MIACME investigation schema
sdo:identifier	ID	identifier
sdo:description	projectDescription	description
sdo:member	source	×
sdo:instrument	instrument	×
sdo:contactPoint	×	contacts
sdo:name	×	title

Take-home messages

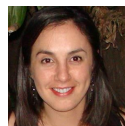
- FAIRsharing.org is an important resource for helping in the analysis and convergence of metadata standards to implement the FAIR principles
- FAIRsharing is part of GO-FAIR and RDA
- FAIRsharing works in support of the development of FAIR metrics
 - Look up service
 - Validation service
- **FAIRification** of metadata standards
 - **Actionable** and **modular** representation of reporting guidelines, formats and their semantic requirements and constraints
 - Automated templates, validation, documentation, comparison, identification of commonalities and dependencies, merging



*Our team works to enabling reproducible research
and FAIR science, driving science and discoveries*



Susanna-Assunta
Sansone, PhD
Associate Professor, Associate Director



Alejandra
Gonzalez-Beltran, PhD
Research Lecturer



Philippe
Rocca-Serra, PhD
Senior Research Lecturer



Allyson
Lister, PhD
Knowledge Engineer



Peter
McQuilton, PhD
Knowledge Engineer



Milo
Thurston, DPhD
Research Software Engineer



Massimiliano
Izzo, PhD
Research Software Engineer



Delphine
Dauga, PhD
Biocurator Contractor



Ramon
Granell
*Research Software and
Knowledge Engineer*



Melanie
Adekale, PhD
Biocurator Contractor



Dominique
Batista
*Research Software and
Knowledge Engineer*



For this specific ‘metadata atomization
work’, we acknowledge two of our projects:

- NIH CEDAR: <https://metadatacenter.org>
- NIH Data Commons Pilot:
<https://commonfund.nih.gov/commons>

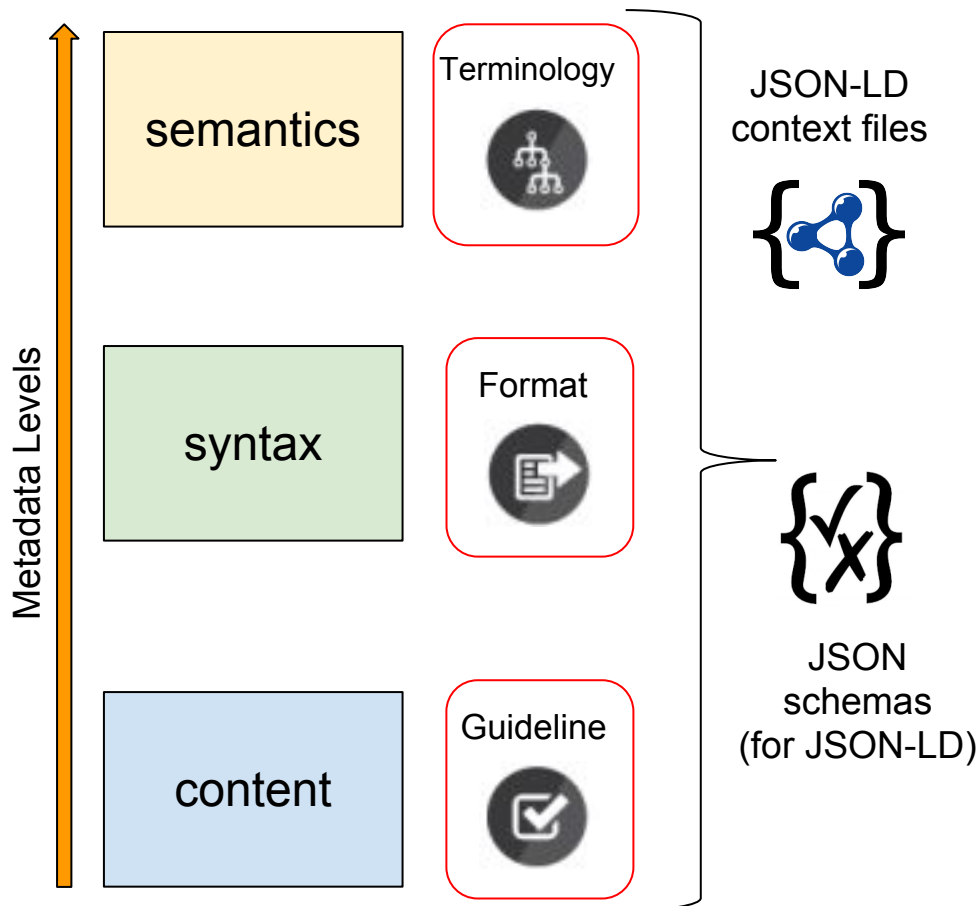
FAIRsharing.org

also part of





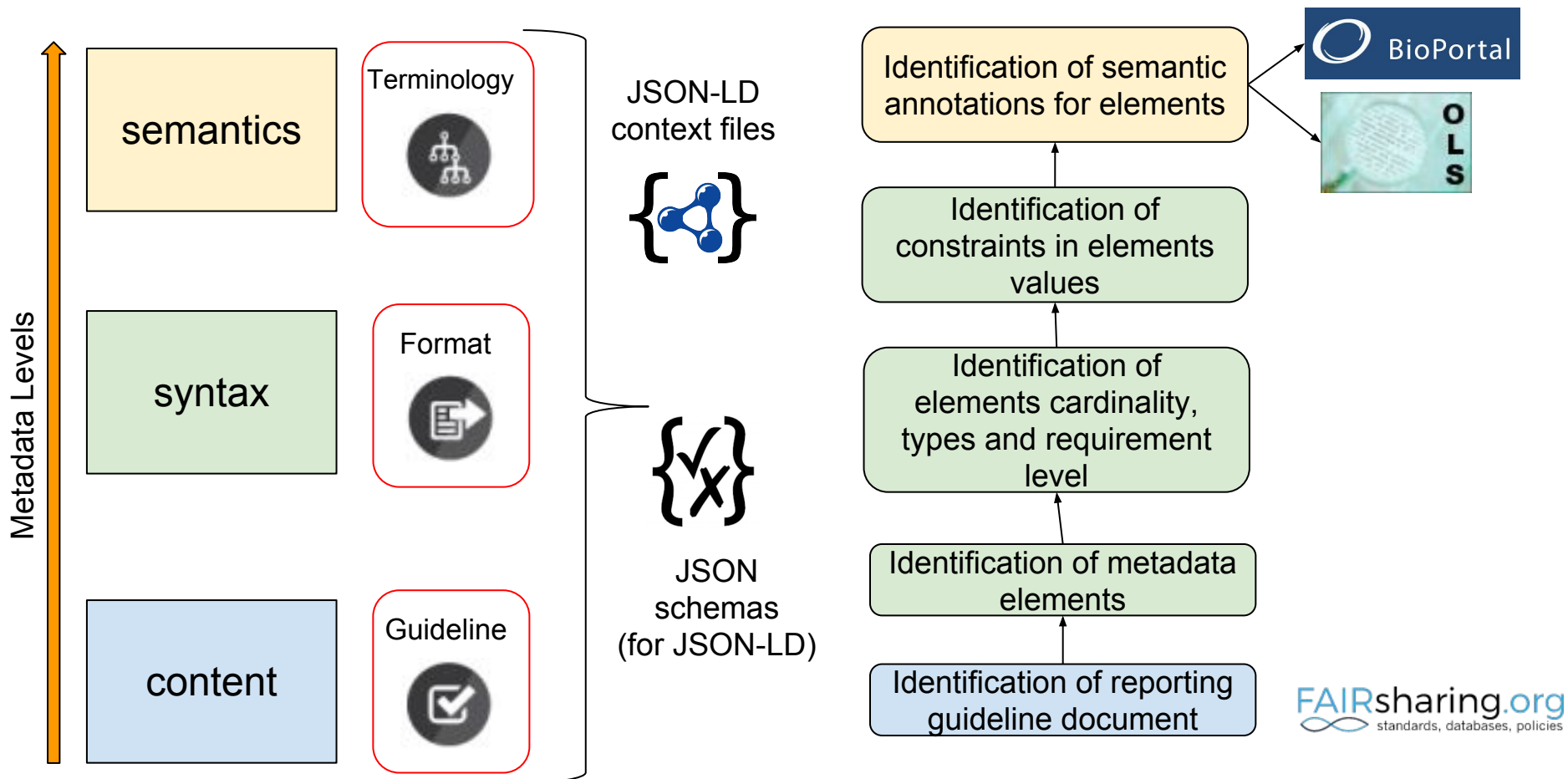
Support for
metadata integration



automated ...

- ...templates**
- ...validation**
- ...documentation**
- ...comparison**
- ...identification of common components**
- ...identification of dependencies**
- ...merge (for compliance with various standards)**

Standards FAIRification Method



Standards FAIRification Method

