# RO-Crate

http://www.researchobject.org/ro-crate/

Carole Goble, The University of Manchester, UK

Stian Soiland-Reyes, The University of Manchester, UK & The University of Amsterdam, The Netherlands

Herbert Van de Sompel, DANS, The Netherlands & Ghent University , Belgium
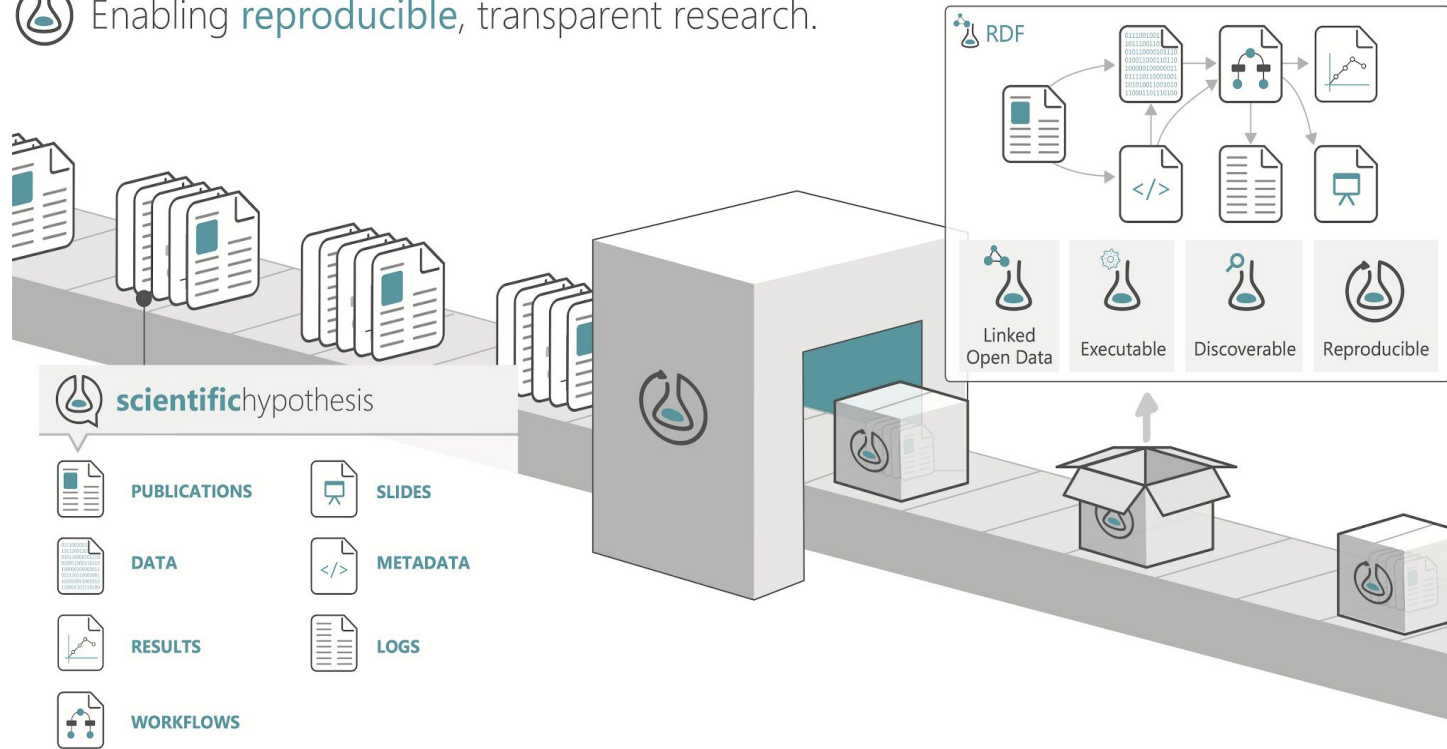
Mark Wilkinson, CBGP UPM-INIA/CSIC, Spain

# Packaging Entities with Machine actionable Metadata



Lightweight, developer friendly, infrastructure independent, extensible packaging approach.

Uses off the shelf web methods.

Aggregate files and/or any URI-addressable content, with contextual information into a machine actionable, metadata rich, structured archive
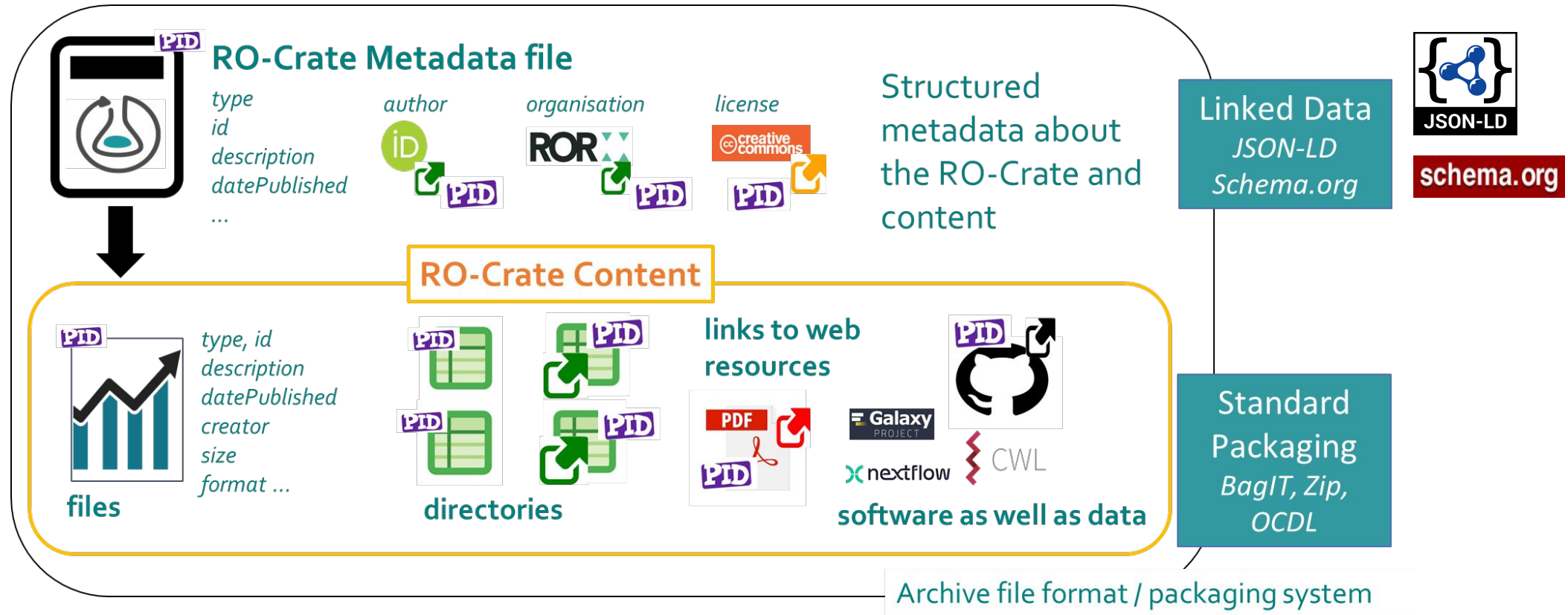
Human readable, search engine accessible.

http://www.researchobject.org/ro-crate/

# Object Exchange & Archive Format

## Structured self-describing, machine readable, metadata objects

Wraps **data, metadata, software** and **references** in single package
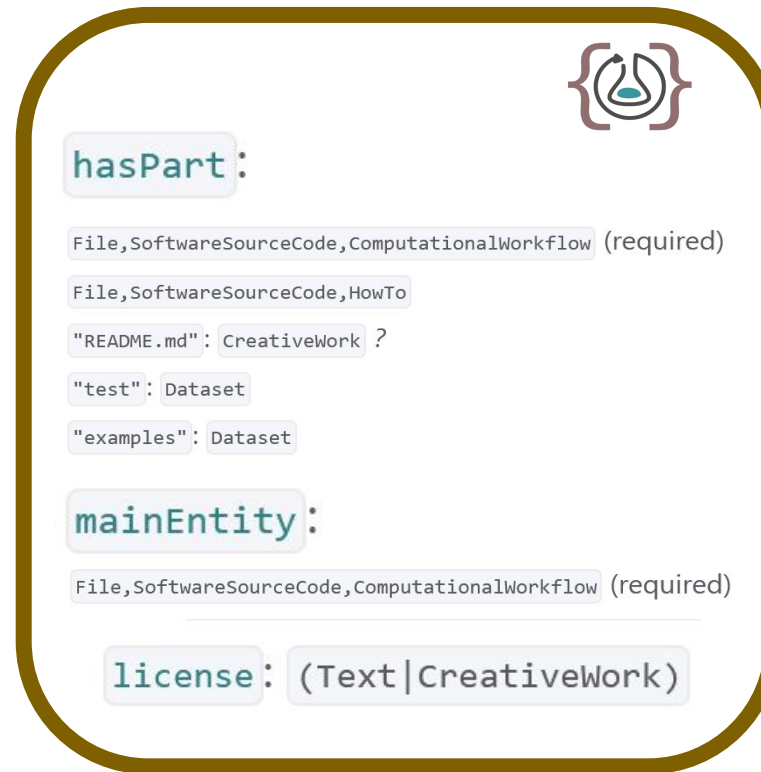
**RO-Crate**



**RO-Crate Metadata file**

type
id
description
datePublished
...

author

organisation

license

Structured metadata about the RO-Crate and content

**RO-Crate Content**

type, id
description
datePublished
creator
size
format ...

files

directories

links to web resources

software as well as data

Linked Data
*JSON-LD Schema.org*

JSON-LD

schema.org

Standard Packaging
*BagIT, Zip, OCDL*

Archive file format / packaging system

Packaging research artefacts with RO-Crate
*Data Science* https://doi.org/10.3233/DS-210053

**RO-Crate Specification 1.1**
https://w3id.org/ro/crate/1.1

# RO-Crate Profiles:
# Community defined Content Checklists

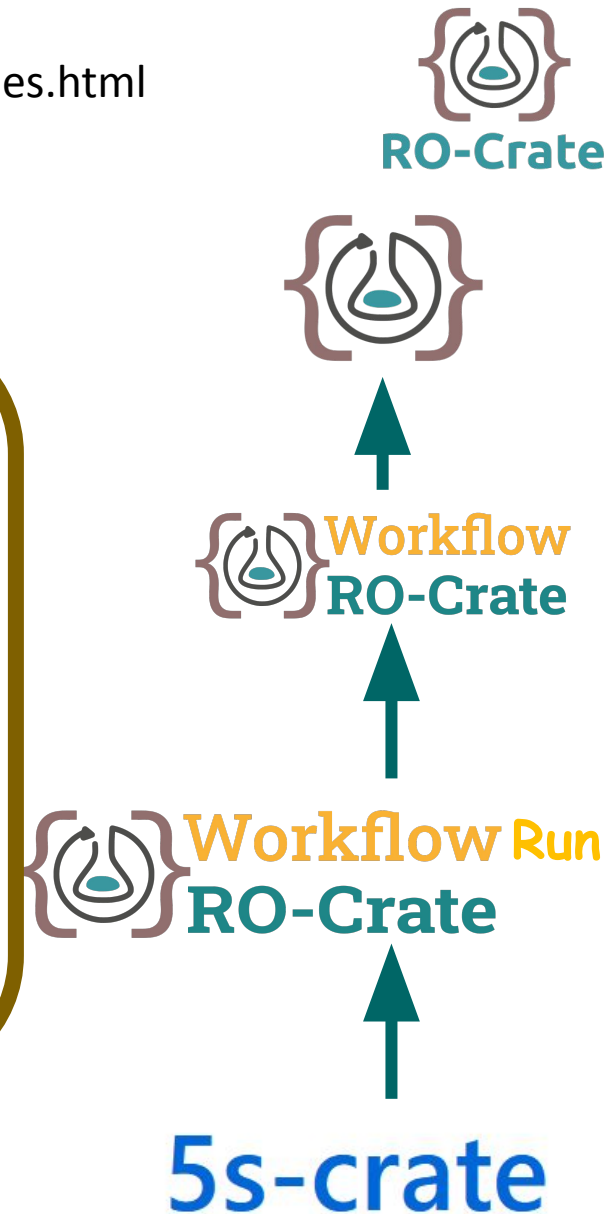https://www.researchobject.org/ro-crate/profiles.html

**RO-Crate**

A set of conventions, types and properties that minimally require and expect to be present in that subset of RO-Crates.
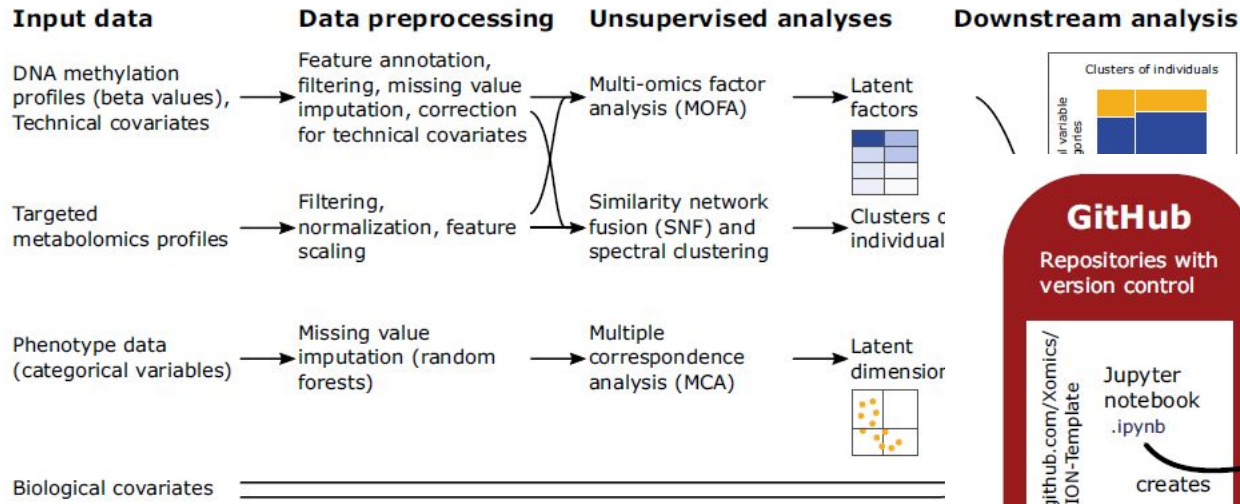
- **Duck typing** for creation, consumption, rendering

- **Classification** for finding and comprehension.

- **Profile Crate** for further defining RO-Crates profile resources

- Key to **extensibility and diversity**



```
hasPart:

File,SoftwareSourceCode,ComputationalWorkflow (required)
File,SoftwareSourceCode,HowTo
"README.md": CreativeWork ?
"test": Dataset
"examples": Dataset

mainEntity:

File,SoftwareSourceCode,ComputationalWorkflow (required)

license: (Text|CreativeWork)
```

**Workflow RO-Crate**

**Workflow Run RO-Crate**

**5s-crate**

https://www.researchobject.org/ro-crate/1.2-DRAFT/profiles
https://www.researchobject.org/ro-crate/1.1/workflows.html
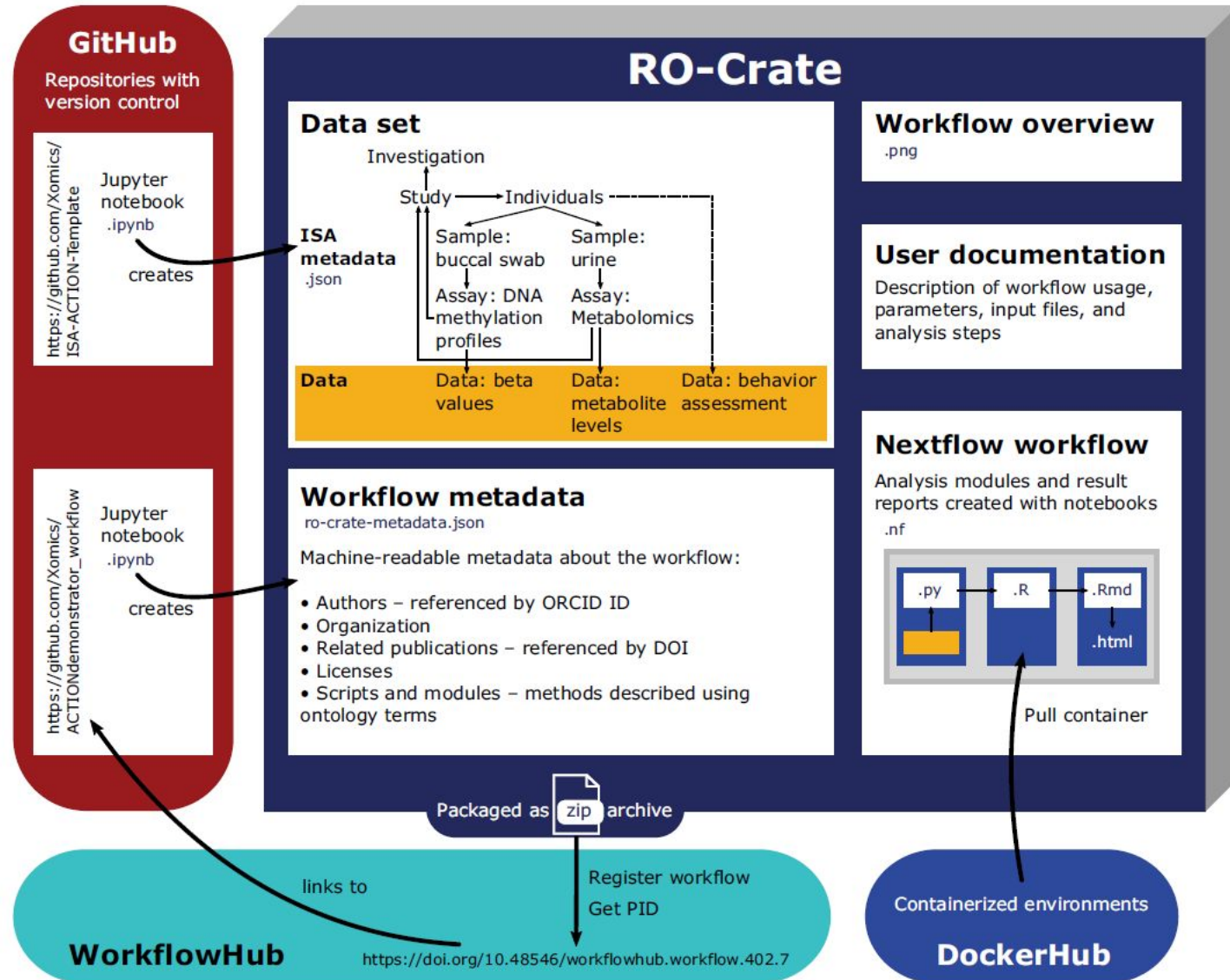https://www.researchobject.org/workflow-run-crate/

Recording provenance of workflow runs with RO-Crate, https://arxiv.org/abs/2312.07852
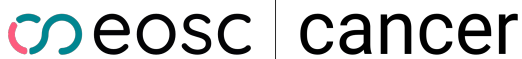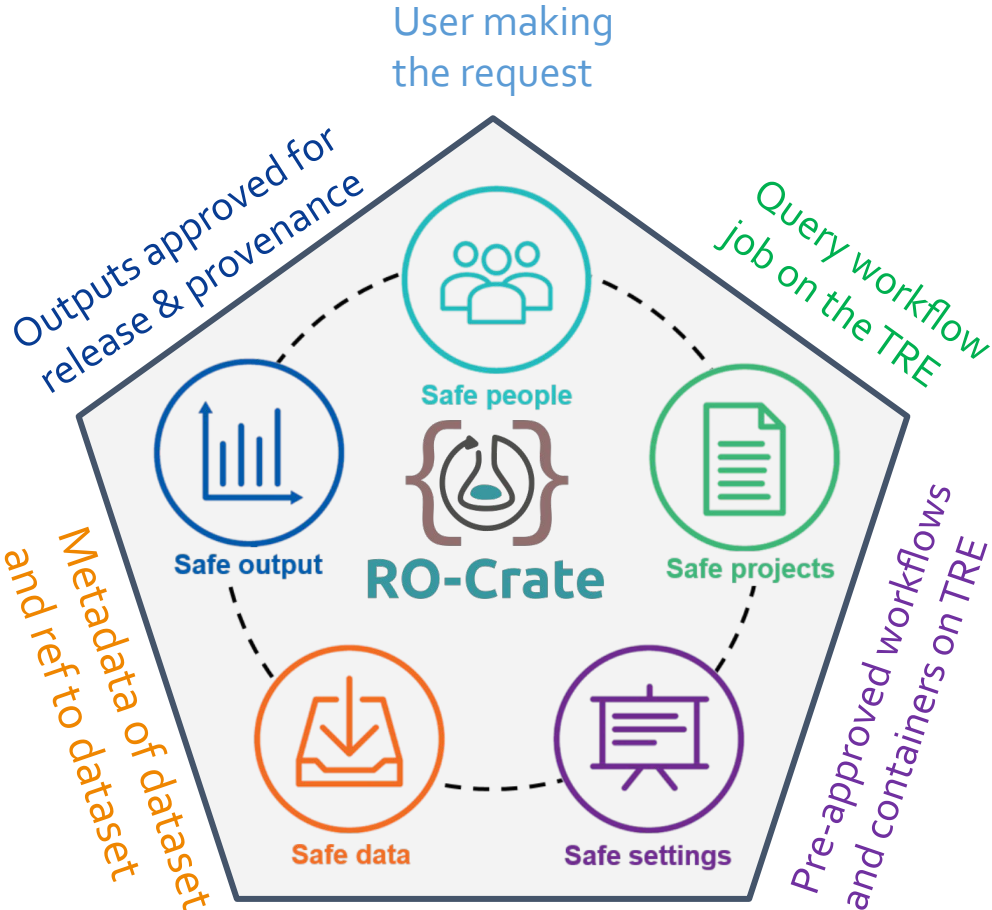
elixir

Anna Niehues, Casper de Visser, et al A multi-omics data analysis workflow packaged as a FAIR Digital Object, *GigaScience*, Volume 13, 2024, giad115, https://doi.org/10.1093/gigascience/giad115

# Metadata actionability for federated analytics across Trusted Research Environments: Five Safes RO-Crate



https://trefx.uk/5s-crate/
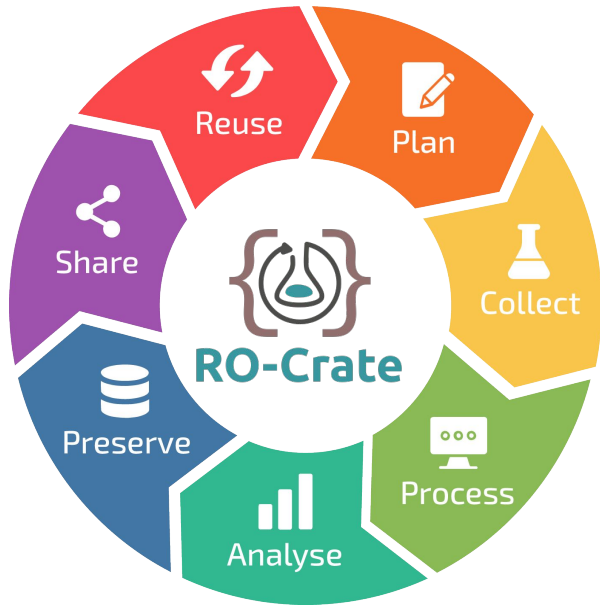
# RO-Crate In Use: active open community

https://www.researchobject.org/ro-crate/in-use/

Repositories

Exchange

Sharing & (Live) Publication
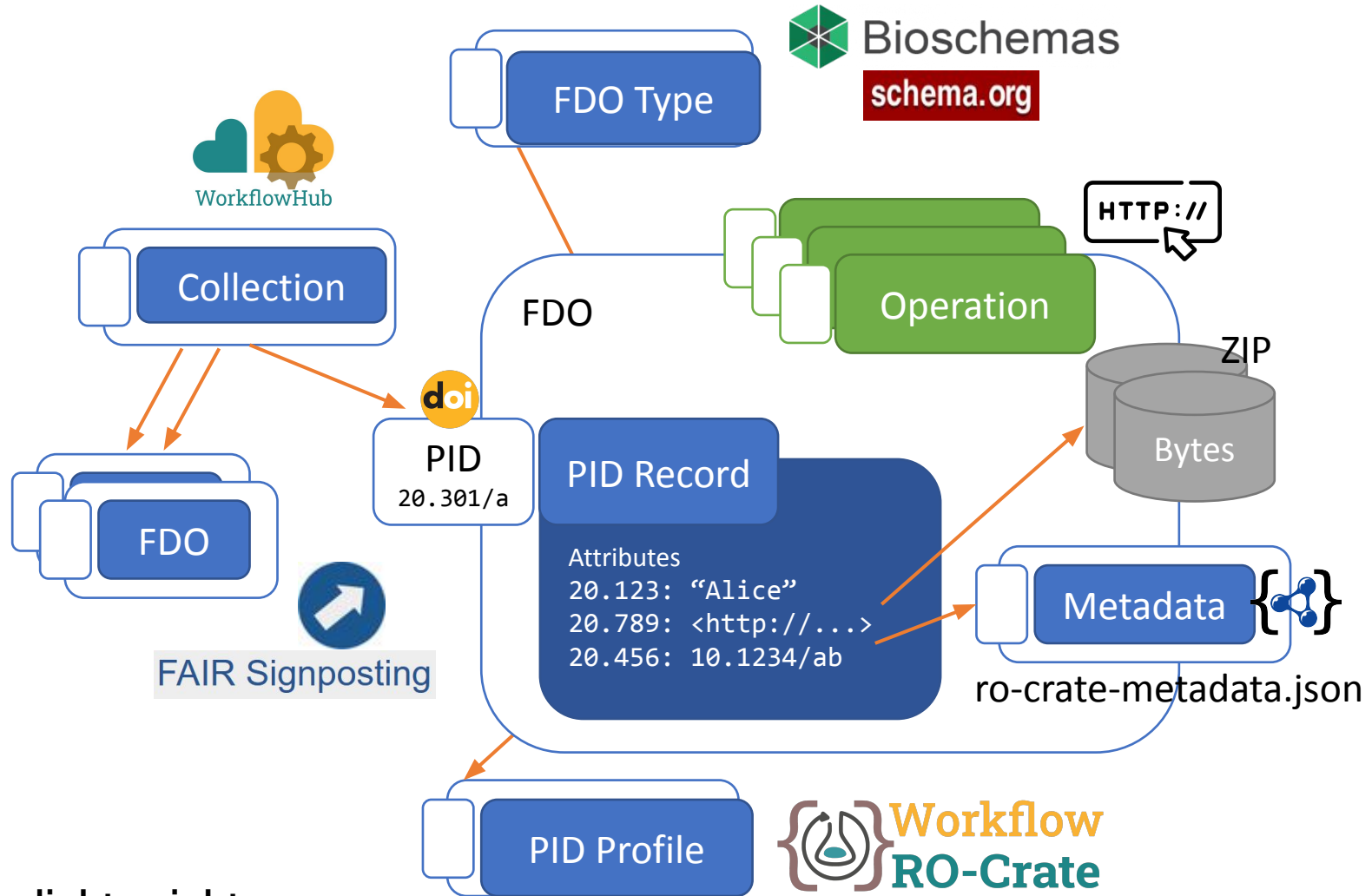
Reproducibility & Provenance
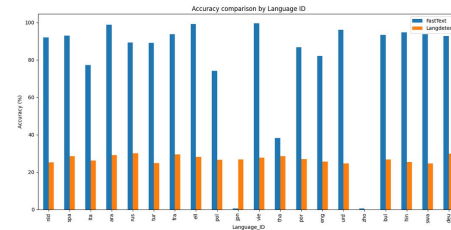
# RO-Crate is an implementation of FDO

Implementing FDO web with current web standards

**FAIR Signposting – to look up**

**RO-Crate – to describe object**

- Types from schema.org
- Metadata files
- Directives to profiles the FDO conforms to
- Community APIs (like GA4GH)



Soiland-Reyes, Sefton, et al (2022): **Creating lightweight FAIR Digital Objects with RO-Crate**. *Research Ideas and Outcomes*, 1st Intl Conf on FAIR Digital Objects

https://doi.org/10.3390/publications8020021

# FAIR Signposting
## https://signposting.org/FAIR/

Carole Goble, The University of Manchester, UK

Stian Soiland-Reyes, The University of Manchester, UK & The University of Amsterdam, The Netherlands

Herbert Van de Sompel, DANS, The Netherlands & Ghent University , Belgium

Mark Wilkinson, CBGP UPM-INIA/CSIC, Spain

# FAIR Signposting: Guiding machine agents through metadata space



Wilkinson, M., Sansone, S.-A., Grootveld, M., Dennis, R., Hecker, D., Huber, R., Soiland-Reyes, S., Van de Sompel, H., Czerniak, A., Thurston, M., Lister, A., & Gaignard, A. (2024). Report on FAIR Signposting and its Uptake by the Community. Zenodo. https://doi.org/10.5281/zenodo.10490289
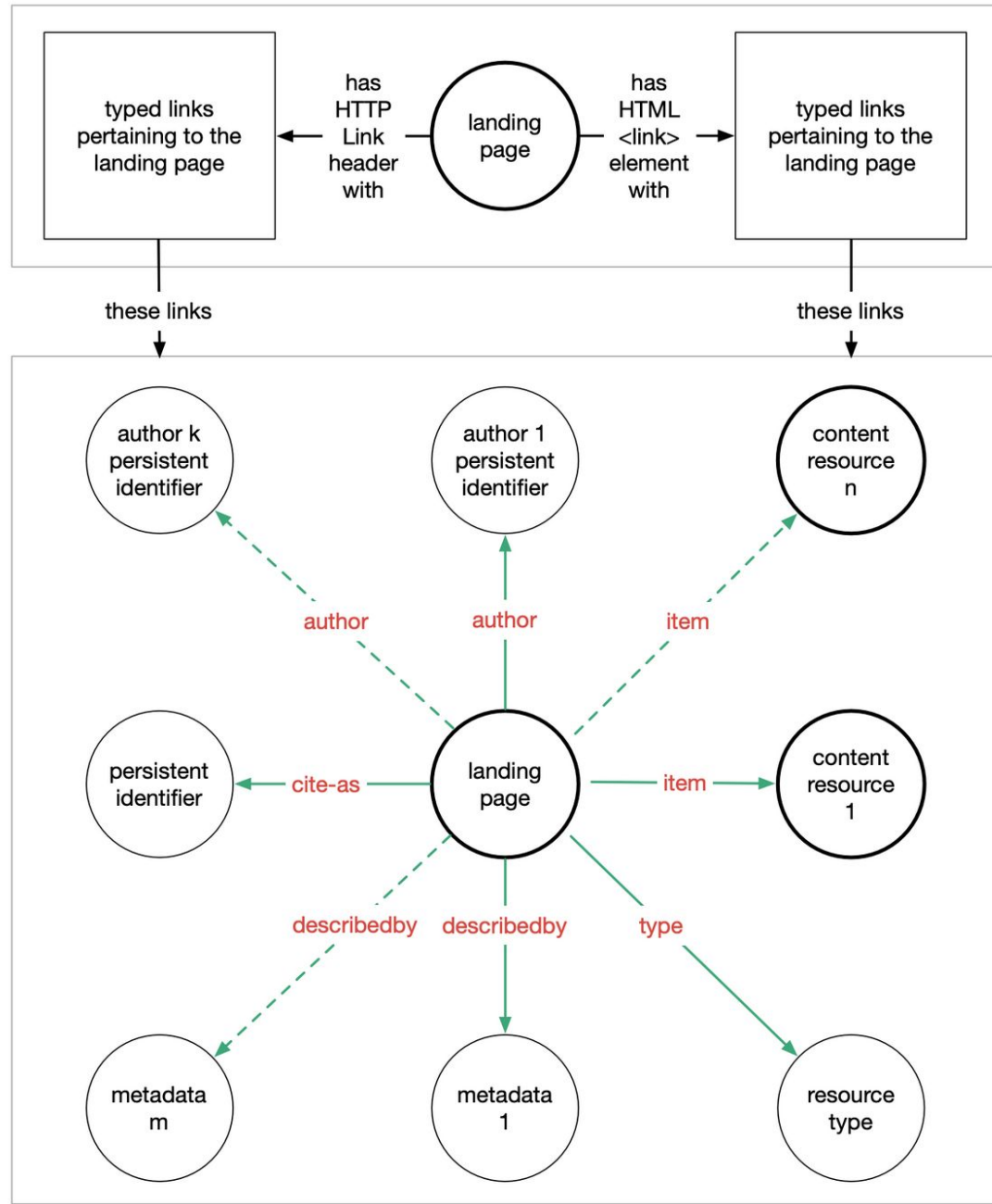
Lightweight, developer friendly, infrastructure independent

Uses off the shelf web methods typed links and landing pages.

Make machine navigation explicit to automated agents so they can locate three essential FAIR elements when they arrive at an object's landing page:

- Its GUID identifier
- Its data records
- Its corresponding metadata records

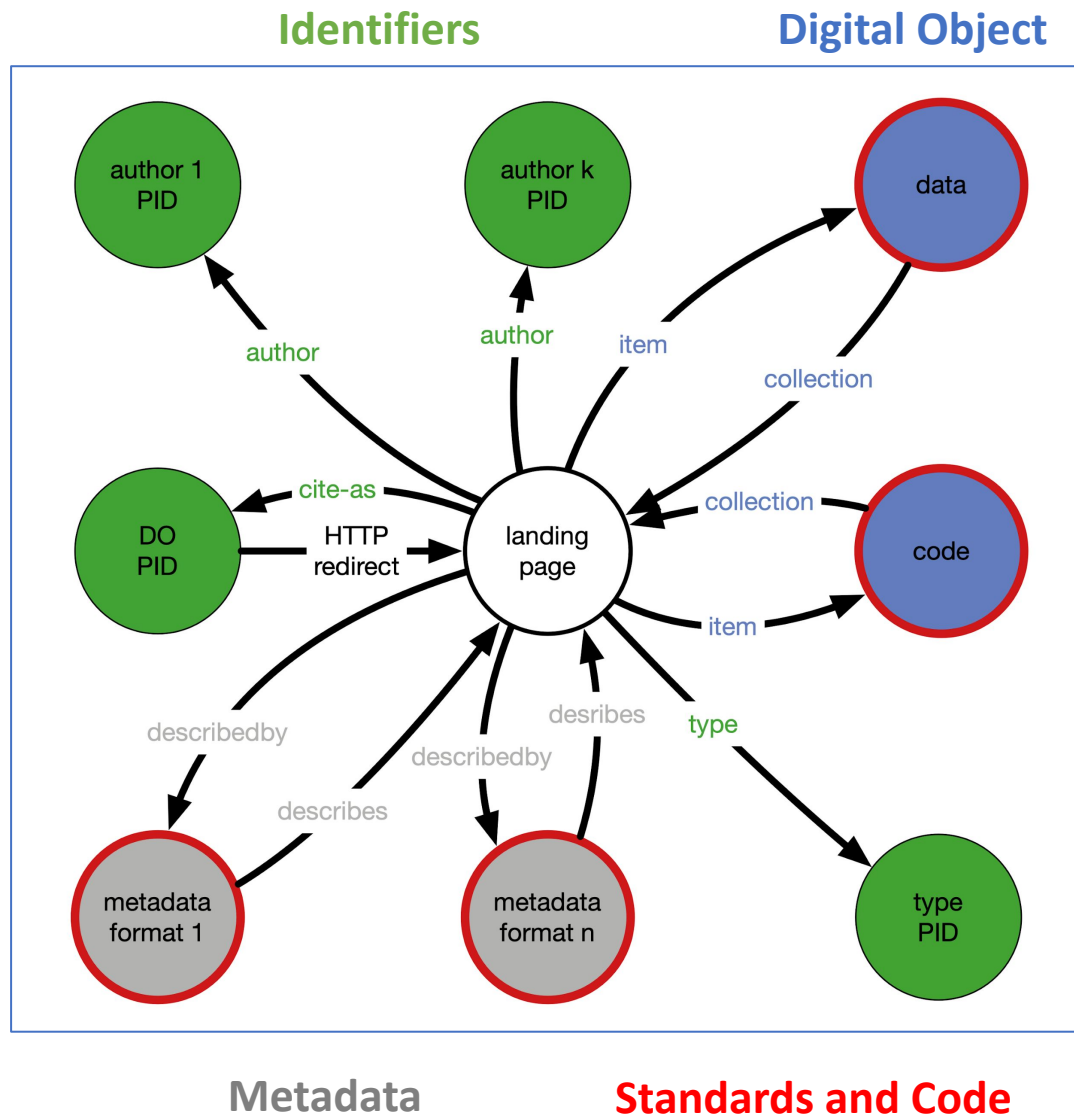# A FAIR Signposting landing page

The HTTP Link Headers, or the HTML Link elements, contain an explicit  set of typed links pointing to the three primary information features of a digital object:

- The landing page canonical GUID, e.g. its DOI (*cite-as links*)
- Its  data records (*item links*)
- Its metadata records (*describedby links*)

# FAIR Signposting Implementation for FDO

FDO PID Record is mainly a navigable data structure

Actual details described by RO-Crate profiles using Schema.org

The persistent identifier (DO PID) redirects to a landing page, from which signposting represents a lightweight FDO Record adding

- *type* (PID in controlled vocabulary to classify the object)

- *describes* (metadata)

- *item* (downloads of the object)

- *author* (e.g. PIDs to ORCID)

- *cite-as* back to the PID.

- Inverse links (e.g., *describes, collection*) allow the FDO to be identified from these constituent resources, e.g. a search engine.

Zenodo. https://doi.org/10.5281/zenodo.10490289
Adapted from https://doi.org/10.5281/zenodo.7977333

https://research.manchester.ac.uk/en/publications/fair-signposting-exposing-the-topology-of-digital-objects-on-the-

# FAIR Signposting in Use

- Easy to implement on existing services

- Re-uses **existing Web standards**: HTTP Link: header or HTML `<link>` tags – no extra magic!

- Takes the guesswork out of navigating metadata to operate FAIR

- Benchmarks for FAIR Signposting
- CKAN Signposting extension
- COAR Notify Protocol
- DANS Data Stations
- Dataverse FAIR Signposting
- Digital Repository Ireland
- DSpace 7 FAIR Signposting
- DSpace-CRIS
- eurac research Environmental Data Platform
- Open Journal System FAIR Signposting
- Pangaea
- Python client for FAIR Signposting
- Radar (Research Data Repository) FAIR Signposting
- RIOXX v3 metadata format
- Ruby Web Link parsing gem
- UCD Digital Library
- UNT Digital Library
- WorkflowHub

https://signposting.org/adopters/

# Generalist Repository Ecosystem Initiative



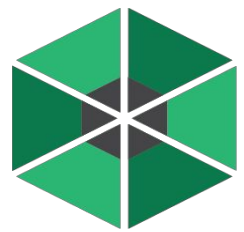Committed to implementing FAIR Signposting

**Take Home:** Practical, adoptable machine actionability enhancing existing infrastructure
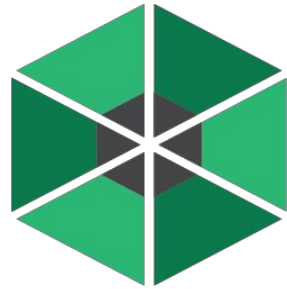
RO-Crate 

+

FAIR Signposting 

 Bioschemas

Combination allows to programmatically and unambiguously navigate from PID to packaged content and start "operations" on the content (E.g. using Jupyter notebooks).

"Good enough machine actionability"

Developer-ware

Community driven

# Bioschemas
## https://bioschemas.org/

Carole Goble, The University of Manchester, UK

Stian Soiland-Reyes, The University of Manchester, UK & The University of Amsterdam, The Netherlands

Leyla Jael Castro, ZB MED Information Centre for Life Sciences, Germany
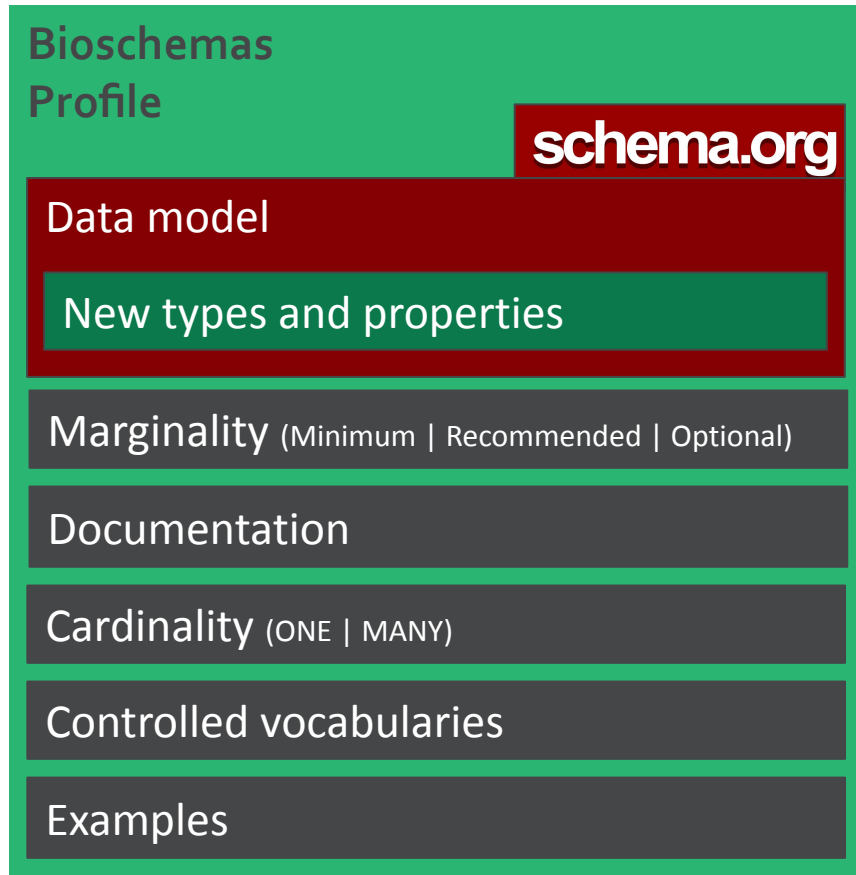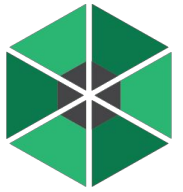
FAIR Machine Actionability
Lorentz Symposium, The Road to FAIR and Equitable Science, 22-26 January 2024, Leiden, the Netherlands

# Bioschemas: machine actionable metadata using web technology
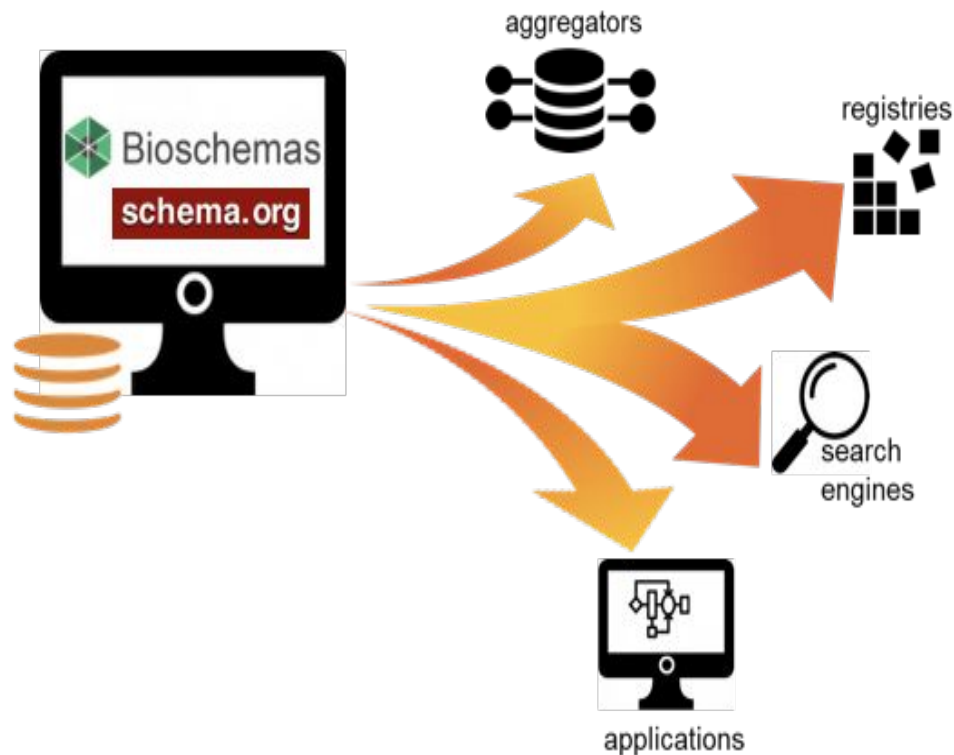


- Lightweight, developer friendly, infrastructure independent structured metadata markup for Web resources

- Uses off the shelf web metadata markup approach used by search engines; Google dataset search

- Improve findability and carry metadata description of for wider FAIRness – e.g. describing RO-Crate Profiles

- Developed in a community process, driven by the biosciences and related disciplines but mostly independent of bioscience

- Part of a movement to using schema.org for metadata in the Sciences & feed Knowledge Graphs

# Bioschemas: machine actionable metadata using web technology



Bioschemas Profile

schema.org

Data model

New types and properties

Marginality (Minimum | Recommended | Optional)

Documentation

Cardinality (ONE | MANY)

Controlled vocabularies

Examples

- A set of opinionated profiles of schema.org

- "Just enough" metadata
  - Dataset from 91 properties to 5 minimal and 8 recommended



**Findable**
- F1, F3. Promote - use of unique identifiers
- F2. Provide - rich metadata specifications
- F4. Promote - use of registries

**Accessible**
- A1. Use - HTTP(S)
- A2. Promote - use of registries

**Interoperable**
- I1. Use - JSON-LD
- I2. Use - schema.org
- I3. Provide - metadata specifications linking objects to each other

**Reusable**
- R1. Provide - metadata specifications (minimum, recommended, optional)
- R1.1. Promote - use of licenses
- R1.2 Promote - provenance and attribution
- R.1.3 Provide - community standards

# Bioschemas: machine actionable metadata using web technology



- **General-purpose** profiles
  - Dataset, Training Material, Course, ComputationalWorkflow, ComputationalTool

- **Domain-specific** profiles
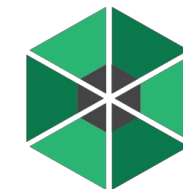  - ChemicalSubstance, Gene, MolecularEntity, Protein, ProteinStructure, Sample, Taxon
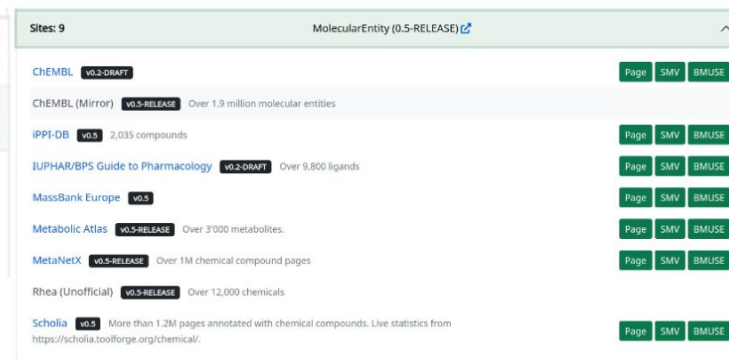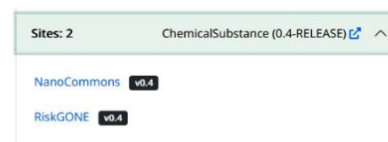
# Bioschemas in Use

**>67 ELIXIR resources** and beyond covering **>180** profile deployments

Standard mechanisms for Training and Tools

Used by RO-Crates



Active Community