



BIODT
biodiversitydigitaltwin

School & Hackaton

March 3rd - March 7th
Officine Cantelmo, Lecce IT

FAIR and Open Science: from general principles to practical implementations

Julian Lopez Gordillo - Naturalis Biodiversity Center

 <https://orcid.org/0000-0003-0401-5122>



Funded by
the European Union

Background

- BSc in Robotics Engineering - Universidad de Sevilla
- MSc in Artificial Intelligence - University of Groningen

Data Specialist @

[Naturalis Biodiversity Center](#)

(Leiden, Netherlands)



Working on open science, FAIRness, computational reproducibility...



Julian Lopez Gordillo

 [/julian-lopez-gordillo/](#)

 [juliancervos](#)

 [0000-0003-0401-5122](#)

Introduction to open science

- The 4 R's
- The reproducibility crisis
- The FAIR principles
- Metadata
- Machine-actionability
- Persistent Identifiers



BIODT

biodiversitydigitaltwin

School & Hackaton



...FAIR?

...Metadata?

...Reproducibility?

...Reusability? Replicability? Repeatability?

Reproducibility → different team, same experimental setup

- The experiment is done with different equipment, different environment, same code/algorithm.

Repeatability → same team, same experimental setup

- The experiment is done by the same team, same environment (software/hardware), same code/algorithm.

Replicability → different team, different experimental setup

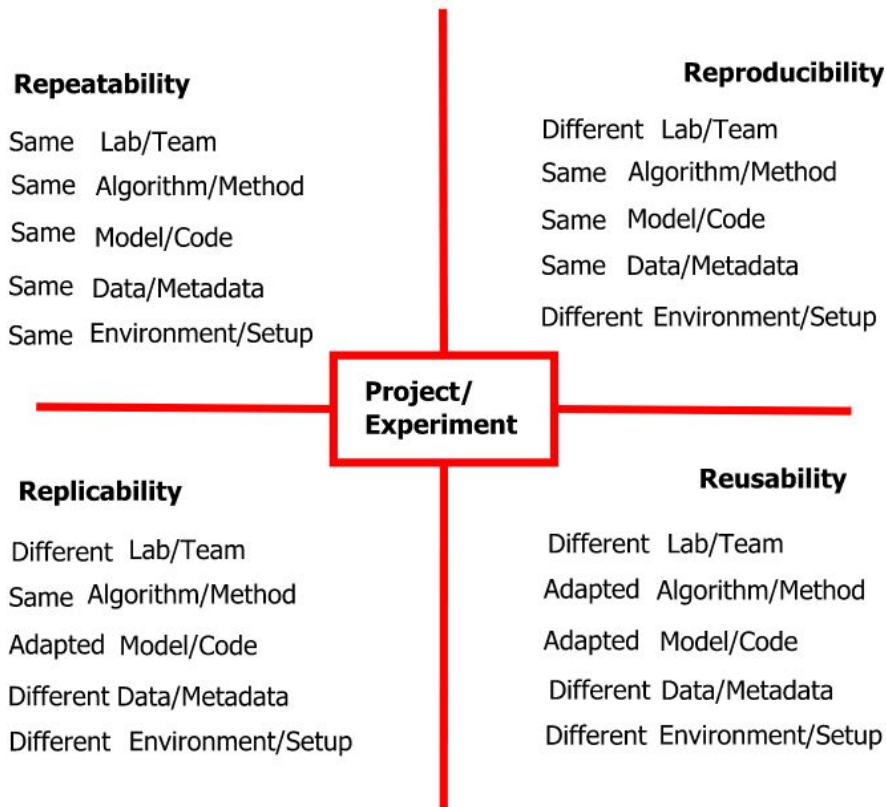
- The experiment is done with different equipment, different environment, different code, same algorithm.

Reusability → different equipment, different and partial experimental configuration

- The experiment is carried out with different equipment, different environments, different codes, and the algorithm partially implemented.

Data Science

Project/Experiment	Repeatability	Reproducibility	Replicability	Reusability
Reproducibility Degree	Lab/Team1 R1	Lab/Team2 R2	Lab/Team3 R3	Lab/Team4 R4
Algorithm/Method				Adapted
Model/Code				Adapted
Data/Metadata				Different
Environment/Setup				Different



Experimental reproducibility: similar input (data) + similar experimental protocol → similar results

Statistical reproducibility: same input (data) + same analysis → same conclusions (independently from (random) sampling variability)

Computational reproducibility: similar input (data) + same code/software + same software environment → exact same bit-wise results

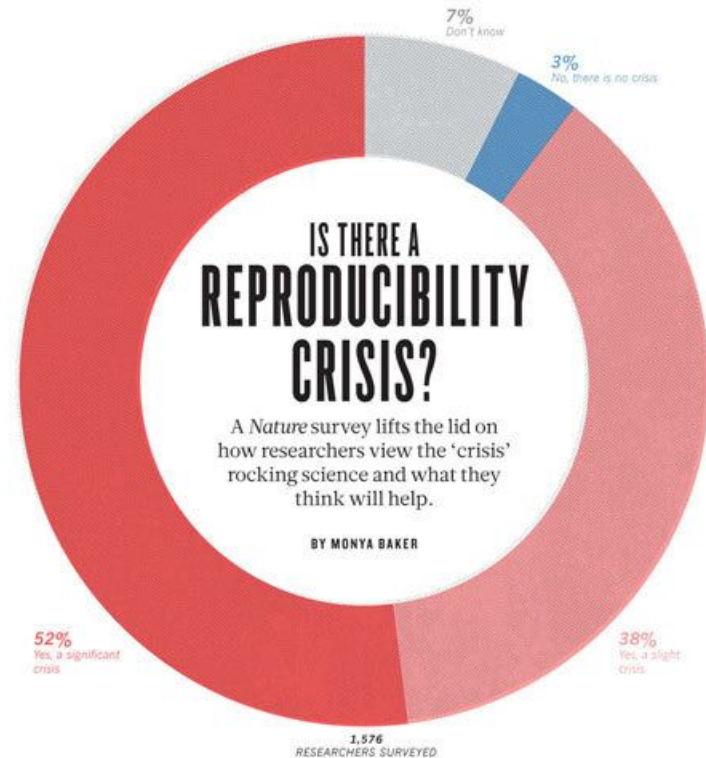
News Feature | [Published: 25 May 2016](#)

1,500 scientists lift the lid on reproducibility

[Monya Baker](#)

[Nature](#) 533, 452–454 (2016) | [Cite this article](#)

158k Accesses | 2032 Citations | 5173 Altmetric | [Metrics](#)



Replication crisis

Article [Talk](#)

From Wikipedia, the free encyclopedia

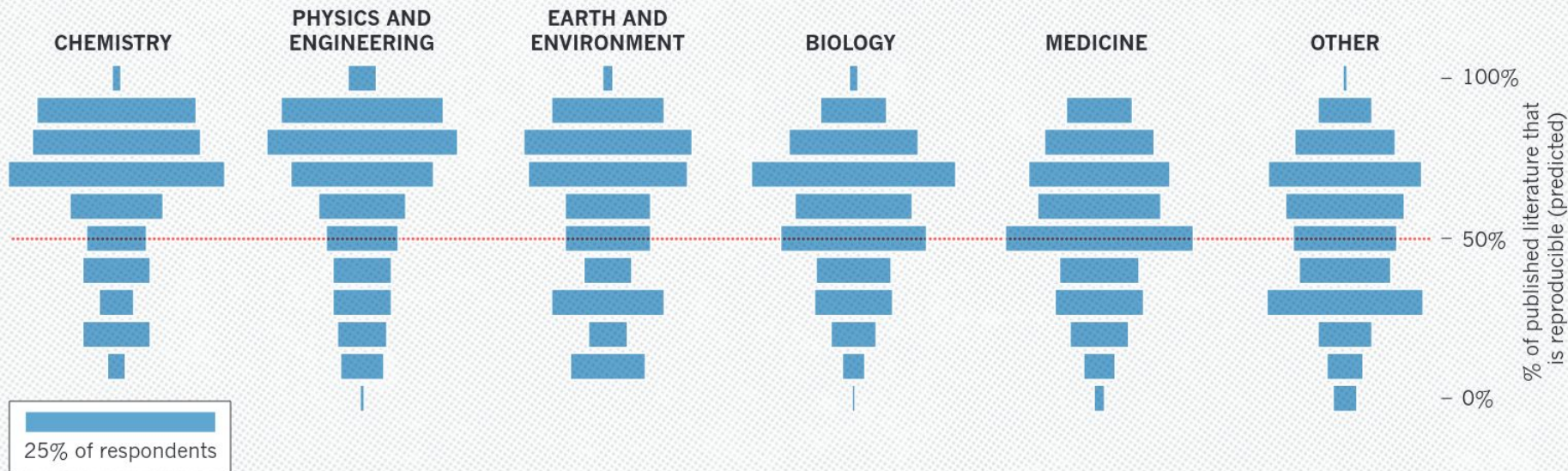
The **replication crisis** (also called the **replicability crisis** and the **reproducibility crisis**) is an ongoing [methodological](#) crisis in which the results of many scientific studies are difficult or impossible to [reproduce](#). Because the reproducibility of empirical results is an essential part of the [scientific method](#),^[2] such failures undermine the credibility of theories building on them and potentially call into question substantial parts of scientific knowledge.

A 'CRISIS' IN NUMBERS

Nature surveyed 1,576 scientists online to get their thoughts on reproducibility in their field and in science in general. See go.nature.com/2vjr4y for more charts and access to the full data.

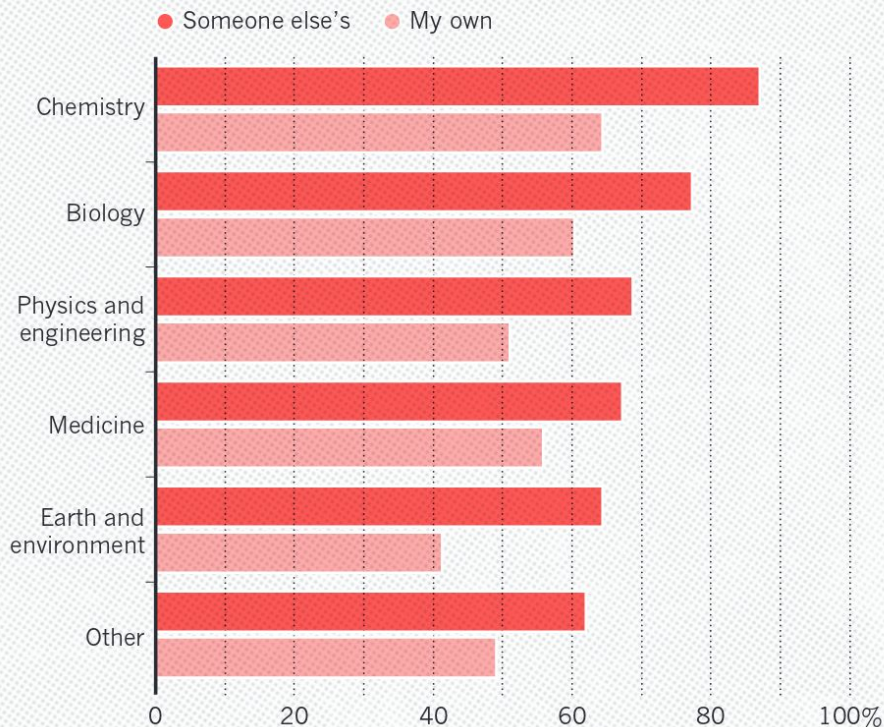
HOW MUCH PUBLISHED WORK IN YOUR FIELD IS REPRODUCIBLE?

Physicists and chemists were most confident in the literature.



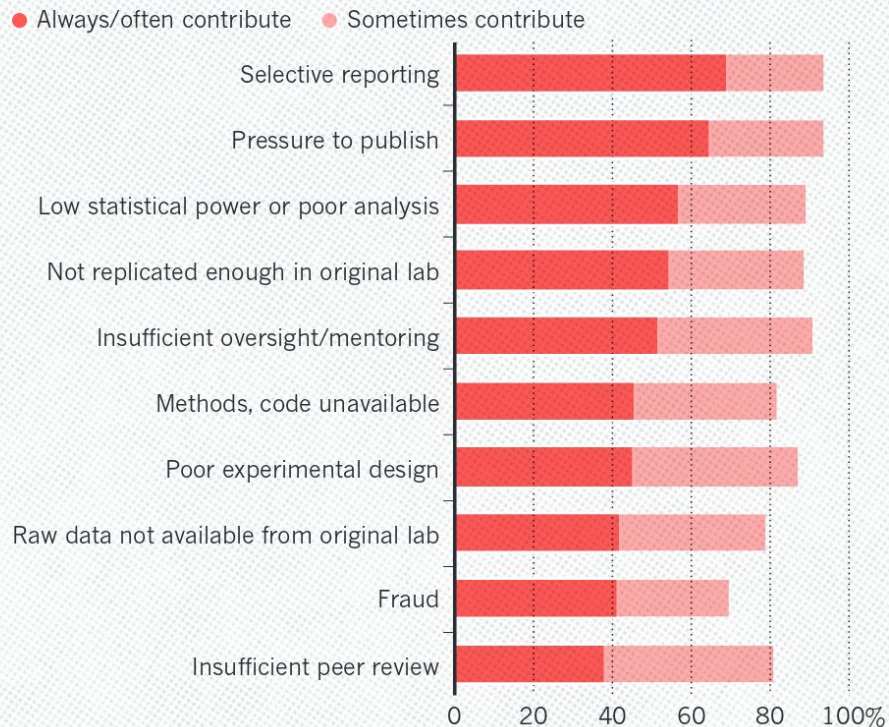
HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.



Complex and uncontrolled software dependencies → multitude of libraries, frameworks, versions... *“dependency hell”*

Low writing quality → overloaded academic jargon, language ambiguities

Compilation and infrastructure setup → compiling source code, debugging it, configuring the running platform...

Float point operations → float point operations can give slightly different numerical results in different systems

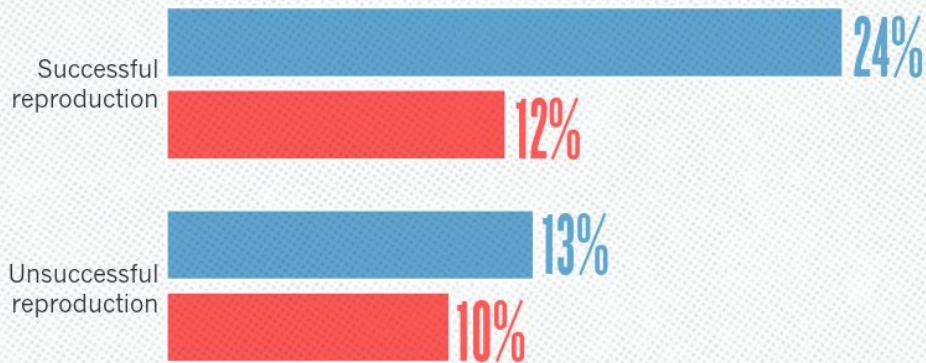
Adapted operating systems → replicated environments require installing the exact version and its respective dependencies to obtain the equivalent result

Lack of academic reward → *“publish or perish”, etc...*

HAVE YOU EVER TRIED TO PUBLISH A REPRODUCTION ATTEMPT?

Although only a small proportion of respondents tried to publish replication attempts, many had their papers accepted.

● Published ● Failed to publish

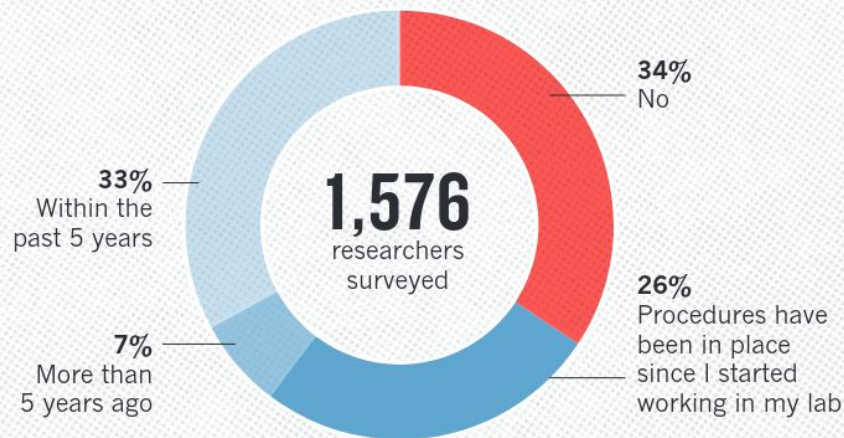


Number of respondents from each discipline:

Biology **703**, Chemistry **106**, Earth and environmental **95**, Medicine **203**, Physics and engineering **236**, Other **233**

HAVE YOU ESTABLISHED PROCEDURES FOR REPRODUCIBILITY?

Among the most popular strategies was having different lab members redo experiments.



FUTURE PERFECT

SCIENCE

Science has been The Reproducibility

decade.

Bad papers are still

By K

Trust, Software

Dana Crane | February

Editorial | [Open access](#)

No raw data the reproducibility

[Tsuyoshi Miyakawa](#) 

[Molecular Brain](#) **13**, Article number: 24 (2020) | [Cite this article](#)

62k Accesses | **108** Citations | **2113** Altmetric | [Metrics](#)

CAREER NEWS | 20 January 2025

Perspective | [Open access](#) | [Published: 25 July 2023](#)

The replication crisis has led to positive structural, procedural, and community changes

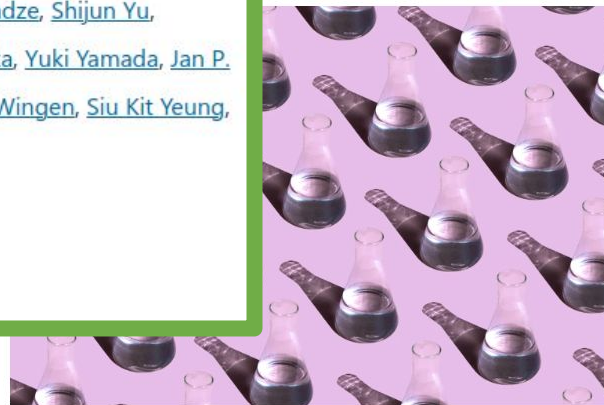
[Max Korbmacher](#), [Flavio Azevedo](#) , [Charlotte R. Pennington](#), [Helena Hartmann](#), [Madeleine Pownall](#), [Kathleen Schmidt](#), [Mahmoud Elsherif](#), [Nate Breznau](#), [Olly Robertson](#), [Tamara Kalandadze](#), [Shijun Yu](#), [Bradley J. Baker](#), [Aoife O'Mahony](#), [Jørgen Ø. -S. Olsnes](#), [John J. Shaw](#), [Biljana Gjoneska](#), [Yuki Yamada](#), [Jan P. Röer](#), [Jennifer Murphy](#), [Shilaan Alzahawi](#), [Sandra Grinschgl](#), [Catia M. Oliveira](#), [Tobias Wingen](#), [Siu Kit Yeung](#), ... [Thomas Evans](#) [+ Show authors](#)

[Communications Psychology](#) **1**, Article number: 3 (2023) | [Cite this article](#)

7736 Accesses | **50** Altmetric | [Metrics](#)

sh' culture blamed for reproducibility crisis

Researchers also flagged small sample sizes and issues of reproducibility problems.



Credit: Getty

*“More than **70%** of researchers have tried and failed to reproduce another scientist’s experiments, and more than **50%** have failed to reproduce their own experiments.”*

Data from study [available here](#). Questionnaire used in the study [available here](#).

Discussion

Have you failed to reproduce an experiment? (e.g. relevant paper in your field, a colleague’s results)

Have you struggled to reproduce your own work? (e.g. after a few weeks/months not working on something)



Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

Metadata:

“Data that provides information about other data”

What are other examples of metadata you can think of?

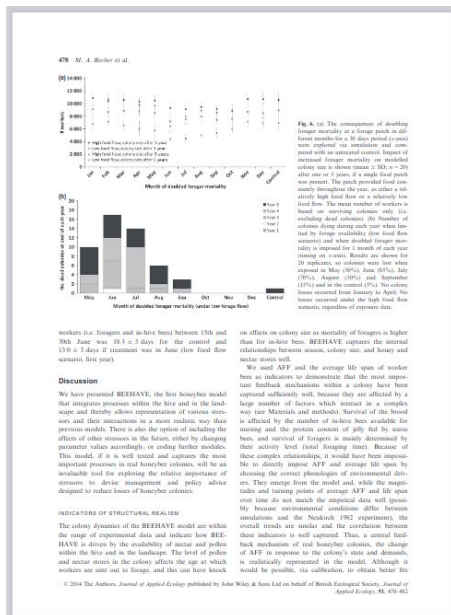
Basic	Permissions	Open With	Image
Image Type	jpeg (JPEG)		
Width	1744 pixels		
Height	1744 pixels		
Camera Brand	FUJIFILM		
Camera Model	X-T4		
Exposure Time	1/1600 s		
Exposure Programme	Manual		
Aperture Value	F1.4		
ISO Speed Rating	160		
Flash Fired	Fired		
Metering Mode	Multi-segment		
Focal Length	35.0 mm		
Software	Shotwell 0.30.14		

```
{
  "@context": [
    "http://doi.org/10.5063/schema/codemeta-2.0",
    "http://w3id.org/ro/crate/1.2-DRAFT/context"
  ],
  "@graph": [
    {
      "@type": "CreativeWork",
      "@id": "ro-crate-metadata.json",
      "about": {"@id": "."},
      "conformsTo": {"@id": "http://w3id.org/ro/crate/1.2-DRAFT"},
      "author": {"@id": "http://orcid.org/0000-0003-0401-5122"}
    },
    {
      "@id": ".",
      "@type": ["Dataset", "SoftwareApplication"],
      "name": "BEEHAVE",
      "description": "BEEHAVE is a computer model to simulate the develop",
      "version": "1.0.0",
      "dateCreated": "2014-03-04",
      "datePublished": "2014-03-04",
      "dateModified": "2014-03-04",
      "license": {"@id": "http://www.gnu.org/licenses/gpl-3.0-standalone"},
      "author": [
        {"@id": "http://orcid.org/0000-0003-0791-7164"},
        {"@id": "http://orcid.org/0000-0002-3221-9512"},
        {"@id": "http://orcid.org/0000-0003-2338-4636"}
      ],
      "publisher": {"@id": "http://orcid.org/0000-0003-0791-7164"},
    }
  ]
}
```

Human-readable

Machine-readable?

Machine-actionable



“Machine-Actionability refers to the information that is consistently structured so that machines can be programmed against such a structure.”

```
{
  "@context": [
    "http://doi.org/10.5063/schema/codemeta-2.0",
    "http://w3id.org/ro/crate/1.2-DRAFT/context"
  ],
  "@graph": [
    {
      "@type": "CreativeWork",
      "@id": "ro-crate-metadata.json",
      "about": { "@id": "." },
      "conformsTo": { "@id": "http://w3id.org/ro/crate/1.2-DRAFT" },
      "author": { "@id": "http://orcid.org/0000-0003-0401-5122" },
    },
    {
      "@id": ".",
      "@type": [ "Dataset", "SoftwareApplication" ],
      "name": "BEEHAVE",
      "description": "BEEHAVE is a computer model to simulate the development of a honeybee colony.",
      "version": "1.0.0",
      "dateCreated": "2014-03-04",
      "datePublished": "2014-03-04",
      "dateModified": "2014-03-04",
      "license": { "@id": "http://www.gnu.org/licenses/gpl-3.0-standalone" },
      "author": [
        { "@id": "http://orcid.org/0000-0003-0791-7164" },
        { "@id": "http://orcid.org/0000-0002-3221-9512" },
        { "@id": "http://orcid.org/0000-0003-2338-4636" }
      ],
      "publisher": { "@id": "http://orcid.org/0000-0003-0791-7164" },
    }
  ]
}
```

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards



“A persistent identifier (PID) is a long-lasting reference to a document, file, web page, or other object.”

GUPRI: Globally Unique, Persistent, Resolvable Identifier



Characteristics	Definition	General rationale/impact on data integration	Specific example of a possible ramification due to non-adherence
Unambiguous	One <code>Local ID</code> must be associated to no more than one entity locally. One <code>URI</code> must be associated to no more than one entity globally.	Avoids collisions that result in integrating on the wrong entity.	A physician uses a wrong clinical guideline and makes a wrongful diagnosis because the info button within the clinical information system is linked to the wrong document.
Unique	One entity should ideally be identified by no more than one <code>URI</code> .	(1) Eliminates the cost of maintaining public mappings between equivalent identifiers (2) Avoids false negatives if data integrators do not leverage or know about a mapping.	A researcher fails to make a pathway discovery because she does not realize that http://mydb.org/1234567 and http://mydb.org/q?=1234567 are in fact the same.
Stable (identifier)	The <code>URI</code> , and by extension the <code>local ID</code> , should wherever possible stay the same over time.	Avoids <code>link rot</code> .	A researcher is unable to reproduce an experiment because the link to a record is dead.
Stable (entity)	Identifier must NOT be reassigned to an altogether different entity, though the original entity may evolve provided a change history is documented.	Avoids integrating on the wrong entity.	Because a new chemical gets an old ID, a chemist uses the wrong chemical in a reaction.
Version-documented	If the entity's definition or essential metadata changes substantially, (Lesson 7) the <code>identifier</code> should, wherever possible be versioned and/or change history documented.	Avoids integrating on the wrong entity state (specified through version).	A given experiment is not reproducible because the specific build version of a gene sequence was not specified.
Persistent	The identifier must NOT be deleted (but may be deprecated).	Avoids <code>link rot</code> .	Information about a gene model is completely lost.
Web-resolvable	The <code>URI</code> must be resolvable to a web address where the data or information about the entry can be accessed.	Avoids the unnecessary proliferation of resolvable identifiers issued by third parties (for entities that are not resolvable and/or not identified in their native context)	A dozen different third-party providers mint identifiers for entities that are not actually under their control. Harmonization between these off-brand identifiers is painful.

Convertible	The <code>local ID</code> and its <code>URI</code> counterpart must be inter-convertible by applying the <code>URI pattern</code> to the <code>local ID</code> . Note that in some communities (e.g., ontologies), the <code>local ID</code> is often a <code>CURIE</code> by default.	Avoids the need for special handling of edge cases when integrating data at scale.	Data integrators spend time cleaning identifiers and handling edge-cases instead of doing science.
Defined	The total set of assignable identifiers for the database must be describable through a formal pattern (regular expression).	Facilitates validation and extraction from scientific text, thus the pattern should be as tightly specified as possible (see Lesson 3).	Identifiers cannot be validated and a provider may find it hard to assess their impact in the literature.
Web-friendly	The <code>local ID</code> should wherever possible be of a format that does not need special handling when used in <code>URL</code> and common exchange formats (e.g., <code>XML</code>).	Avoids potential points of failure due to malformed <code>URL</code> , <code>XML</code> , etc.	Use of the identifier produces malformed <code>XML</code> and/or requires special detection and encoding.
Free to assign	The <code>identifier</code> should ideally be assigned at no cost to individuals depositing data in a repository.	Lowers barriers for data generators to deposit data.	Data generators become reluctant to deposit data in order to minimize costs.
Open access and use	The <code>identifier</code> and its <code>label</code> should be able to be transparently referenced and actioned (e.g., in a public index or search) anywhere by anyone and for any reason. Restrictions on associated data may apply but are not recommended.	Enables integration on the basis of scientific merit, rather than on the restrictions of the license.	When there are license restrictions on the identifier and/or label (not just the content) it thwarts meaningful reuse and redistribution of whole datasets.
Documented	The identifier scheme should be documented.	Encourages consistent use of existing identifiers by others and reduces the number of ways identifiers are represented.	Inconsistent informal approaches to referencing are difficult to harmonize post-hoc. By extension, impact is harder to assess.



BIODT

biodiversitydigitaltwin

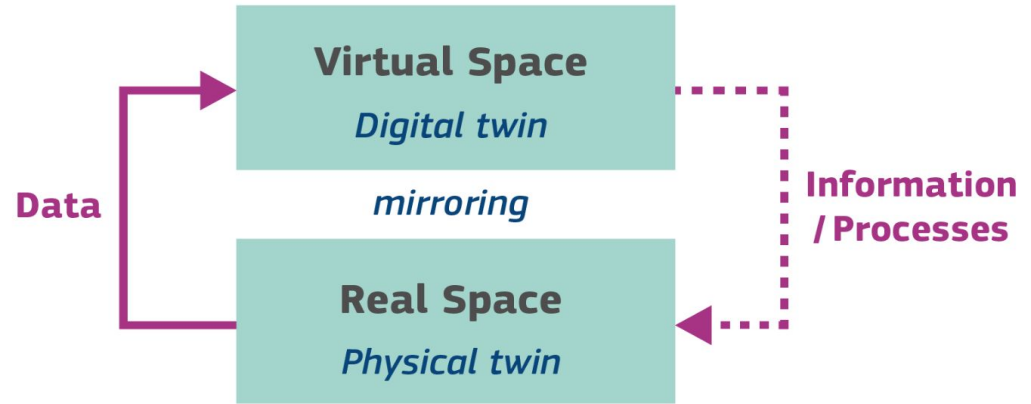
School & Hackaton

Overview of approach taken in BioDT:

- FAIR Digital Objects
- Research Object Crate framework
- JSON-LD
- Metadata profiles and community standards



“A digital twin is a virtual representation of real-world entities and processes, synchronized at a specified frequency and fidelity.”



The Biodiversity Digital Twin prototype will provide **advanced models for simulation and prediction** capabilities, through practical use cases addressing critical issues related to global biodiversity dynamics.

Use Cases

Species response to environmental change



- Biodiversity dynamics
- Ecosystem services

Genetically detected biodiversity



- Crop wild relatives and genetic resources for food security
- DNA detected biodiversity, poorly known habitats

Dynamics and threats from and for species of policy concern



- Invasive species
- Endangered species

Species interactions with each other and with humans



- Disease outbreaks
- Pollinators

Research Infrastructures

GBIF



The Global Biodiversity Information Facility (GBIF) is an international network and data infrastructure providing open access to biodiversity data.

LifeWatch ERIC



LifeWatch ERIC is the e-Science European infrastructure for biodiversity & ecosystem research.

eLTER



The Integrated European Long-Term Ecosystem (eLTER) focuses on critical zone and socio-ecological research.

DiSSCo



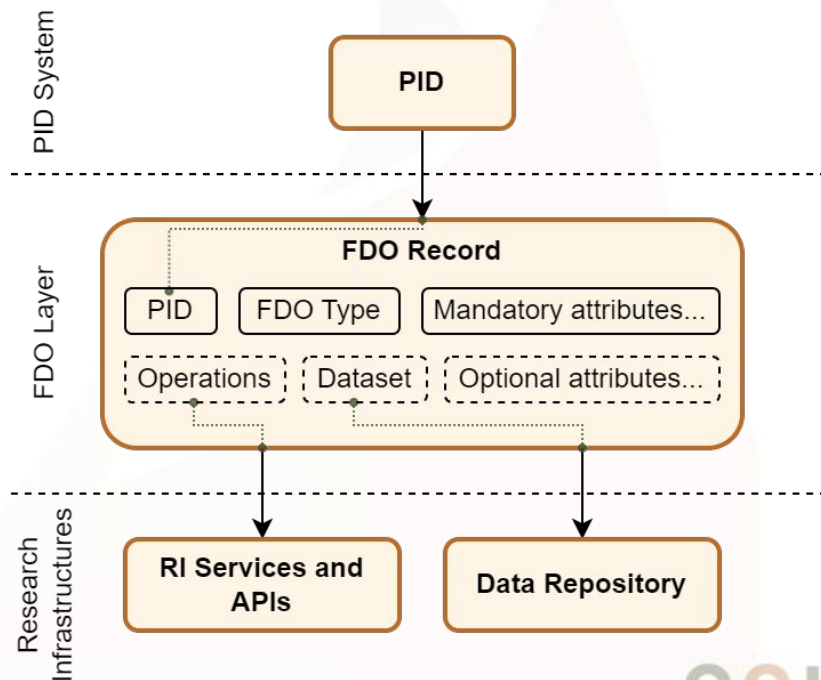
The Distributed System of Scientific Collections (DiSSCo) is a Research Infrastructure (RI) for Natural Science Collections.

Objectives

1. Build and deploy a pre-operational BioDT for addressing biodiversity dynamics
2. Support the interoperability of data and services through the integration of the BioDT with research infrastructure platforms and workflows
3. Ensure interoperability of BioDT with [Destination Earth](#) and [the European Data Infrastructure](#)

Task 5.1 - Consolidate RI FAIR Data designs into a FAIR Digital Object (FDO) layer:

1. Provide a **reference** for the design of digital objects within BioDT such that they are **coherent** and can interact with each other as part of the larger system.
2. Harmonise the description of digital objects to **external resources provided by different RIs** and the other data and service providers.

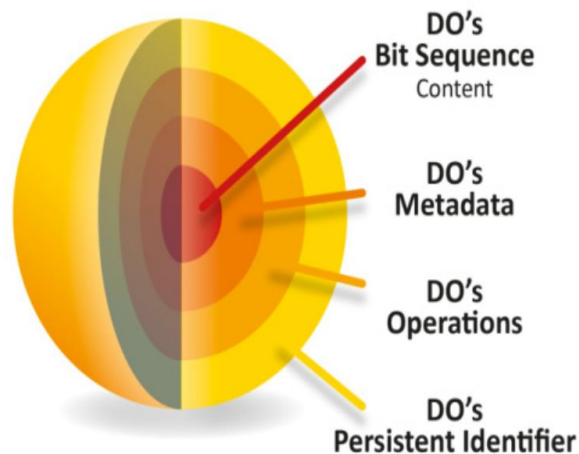


Centred around structured metadata

An abstraction for data, software, workflows, models...

Goal: model for guiding FAIR implementations

- Technology-independent
- Specifications on minimum requirements
- Many possible implementations
→ FDO Configurations...



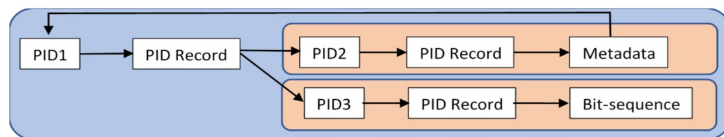
*A FAIR Digital Object is a unit composed of data and/or metadata regulated by structures or schemas, and with an assigned globally unique and persistent identifier (PID), which is findable, accessible, interoperable and reusable **both by humans and computers** for the reliable interpretation and processing of the data represented by the object.*

The FDO specifications are still in development, and some details are only defined loosely

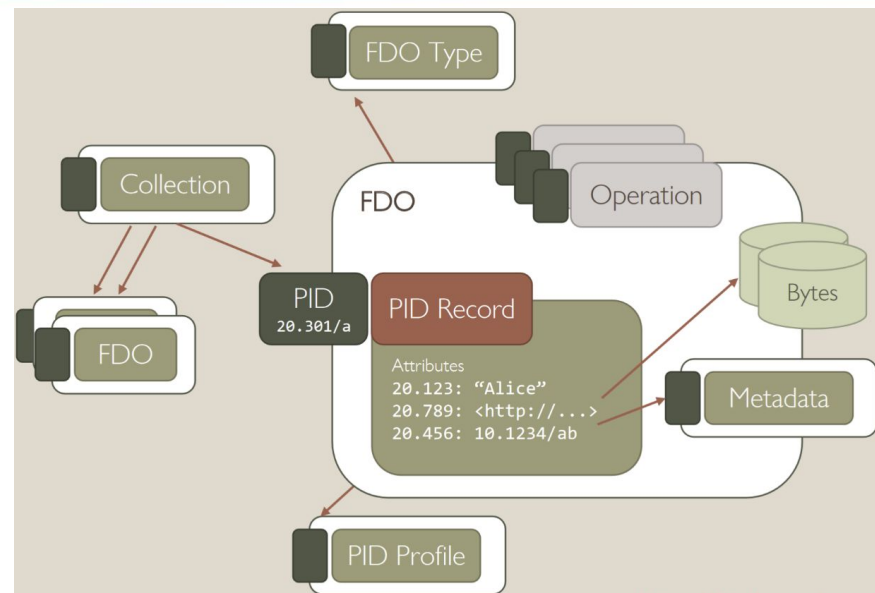
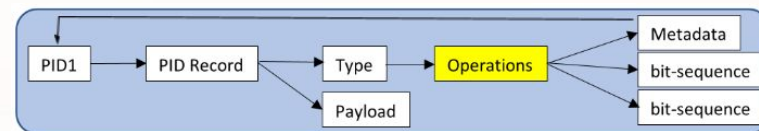
One of the fundamental aspects of an FDO is to have a (globally unique, resolvable) Persistent Identifier, or **PID**

There should be some **structured metadata** associated with that PID (typically, the PID Record, also called FDO Record lately)

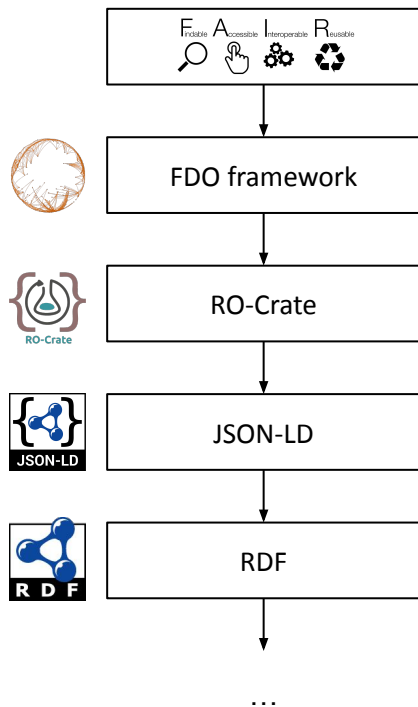
From there, there are multiple ways of complying with the FDO specs. For example, different ways for having extended metadata for a certain FDO (PID1):



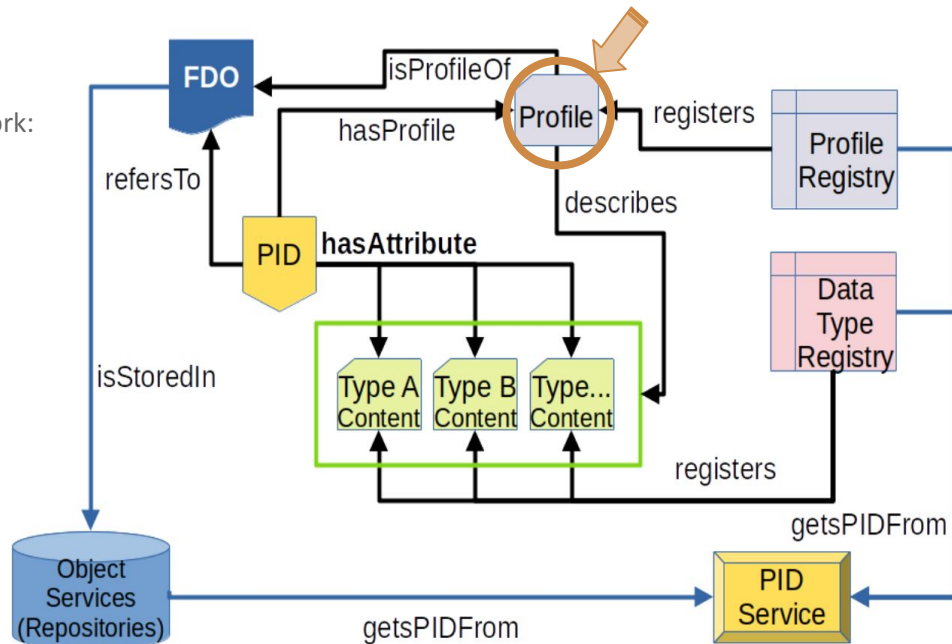
Extended metadata as a separate FDO (PID2) linked on the PID1 Record



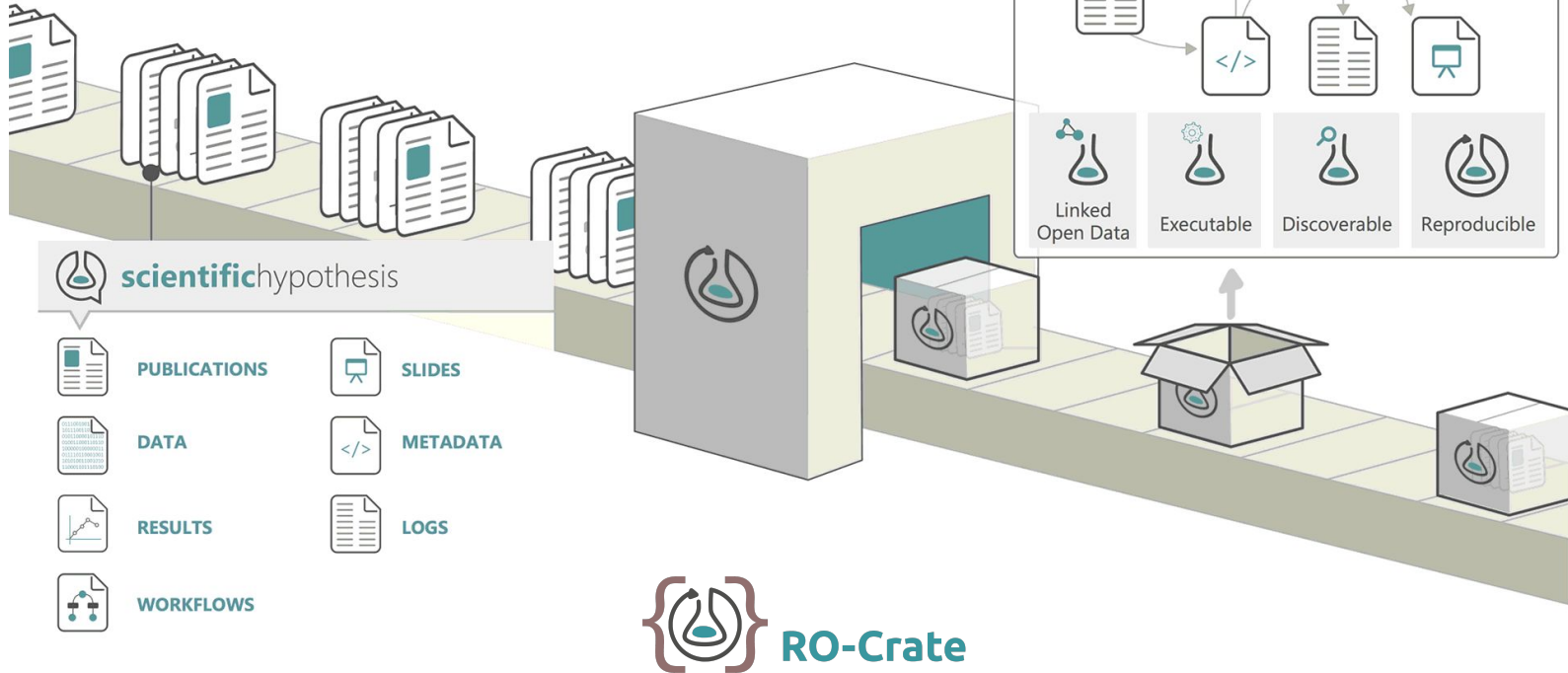
Extended metadata retrievable via an operation as defined in the FDO's Type (linked on the PID1 Record).



Implementation framework:



 Enabling **reproducible**, transparent research.





JSON for Linking Data

Data is messy and disconnected. JSON-LD organizes and connects it, creating a better Web.

<https://json-ld.org/>

⇔ Linked Data

Linked Data empowers people that publish and use information on the Web. It is a way to create a network of standards-based, machine-readable data across Web sites. It allows an application to start at one piece of Linked Data, and follow embedded links to other pieces of Linked Data that are hosted on different sites across the Web.

A Simple Example

```
{
  "@context": "http://json-ld.org/contexts/person.jsonld",
  "@id": "http://dbpedia.org/resource/John_Lennon",
  "name": "John Lennon",
  "born": "1940-10-09",
  "spouse": "http://dbpedia.org/resource/Cynthia_Lennon"
}
```

{ } JSON-LD

JSON-LD is a lightweight Linked Data format. It is easy for humans to read and write. It is based on the already successful JSON format and provides a way to help JSON data interoperate at Web-scale. JSON-LD is an ideal data format for programming environments, REST Web services, and unstructured databases such as CouchDB and MongoDB.

<https://www.w3.org/TR/json-ld/>

```
{ "@context": "https://w3id.org/ro/crate/1.1/context",
  "@graph": [
```

```
{ "@type": "CreativeWork",
  "@id": "ro-crate-metadata.json",
  "conformsTo": { "@id": "https://w3id.org/ro/crate/1.1" },
  "about": { "@id": "." }
}
```

```
{ "@id": ".",
  "identifier": "https://doi.org/10.5281/zenodo.1009240",
  "@type": "Dataset",
```

```
  "hasPart": [
    { "@id": "cp7glop.ai" },
    { "@id": "lots_of_little_files/" },
    { "@id": "communities-2018.csv" },
    { "@id": "https://doi.org/10.4225/59/59672c09f4a4b" },
    { "@id": "SciDataCon-Presentations/AAA_Pilot_Abstract.html" }
  ],
},
```

```
"author": { "@id": "https://orcid.org/0000-0002-8367-6908" },
"publisher": { "@id": "https://ror.org/03f0f6041" },
"citation": { "@id": "https://doi.org/10.1109/TCYB.2014.2386282" },
"name": "Presentation of user survey 2018"
```

```
},
{"@id": "cp7glop.ai",
 "@type": "File",
 "name": "Diagram showing trend to increase",
...
},
...
```

JSON-LD

preamble

RO-Crate **metadata file**
descriptor



RO-Crate **root**
dataset

..collection of **Data**
entities

..described w/ **contextual**
entities

Flat list of metadata per entity



```
{
  "@id": "figure.png",
  "@type": ["File", "ImageObject"],
  "name": "XXL-CT-scan of an XXL Tyrannosaurus rex skull",
  "identifier": "https://doi.org/10.5281/zenodo.3479743",
  "author": {"@id": "https://orcid.org/0000-0002-8367-6908"},
  "encodingFormat": "image/png"
}
```



```
{
  "@id": "https://orcid.org/0000-0002-8367-6908",
  "@type": "Person",
  "affiliation": { "@id": "https://ror.org/03f0f6041" },
  "name": "J. Xuan"
}
```



```
{
  "@id": "https://ror.org/03f0f6041",
  "@type": "Organization",
  "name": "University of Technology Sydney",
  "url": "https://www.uts.edu.au/"
}
```

The collection describes each object
(briefly)

Data and **Contextual** entities
described *within* RO-Crate Metadata File

Base vocabulary & types: **schema.org**

Cross-references to further contextual
entities

RO-Crate **principle**:
Reuse existing PIDs and URLs

..but always **describe entities** which lack a
human-readable resolution

```

1  {
2    "@context": ["https://w3id.org/ro/crate/1.1"],
3    "@graph": [
4      {
5        "@id": "ro-crate-metadata.json",
6        "@type": "CreativeWork",
7        "about": { "@id": "./" },
8        "conformsTo": { "@id": "https://w3id.org/ro/crate/1.1" },
9        "author": { "@id": "http://orcid.org/0000-0003-0401-5122" }
10     },
11     {
12       "@id": "./",
13       "@type": ["Dataset", "SoftwareApplication"],
14       "identifier": "https://workflowhub.eu/workflows/810?version=1",
15       "name": "Cultural Ecosystem Services pDT - Biodiversity Model",
16       "description": "Species distribution models for species that p",
17       "conformsTo": "https://biomt.github.io/biomt-fair/metadata_pro",
18       "dateCreated": "2023-06-29",
19       "license": { "@id": "https://spdx.org/licenses/MIT.html" },
20       "author": [
21         { "@id": "https://orcid.org/0000-0001-6755-9456" },
22         { "@id": "https://orcid.org/0000-0003-2428-272X" },
23         { "@id": "https://orcid.org/0009-0003-5290-786X" },
24         { "@id": "https://orcid.org/0000-0002-4180-9338" },
25         { "@id": "https://orcid.org/0000-0003-0401-5122" }
26       ],
27       "codeRepository": "https://github.com/BioDT/uc-ces/tree/main/bi",
28       "contributors": [
29         { "@id": "https://orcid.org/0000-0001-6755-9456" },
30         { "@id": "https://orcid.org/0009-0003-5290-786X" }
31       ],
32       "softwareVersion": "0.1.0",

```

```

33     {
34       "@id": "https://sws.geonames.org/10104113/",
35       "@type": "Place",
36       "name": "Cairngorms National Park",
37       "description": "The Cairngorms is part of an international
38         located in the Scottish Highlands, 127 miles north of
39         Angus and Perth and Kinross. Nearly half of the land is
40         of international importance for nature, and is protected
41         Special Scientific Interest within the Park.",
42       "geo": {
43         "@type": "GeoCoordinates",
44         "latitude": "57.04395",
45         "longitude": "-3.60334",
46         "name": "Latitude: 57.04395 Longitude: -3.60334"
47       }
48     },
49     {
50       "@id": "https://orcid.org/0000-0001-6755-9456",
51       "@type": "Person",
52       "name": "Simon Rolph",
53       "email": "simrol@ceh.ac.uk",
54       "affiliation": { "@id": "https://ror.org/00pggkr55" }
55     },
56     {
57       "@id": "https://orcid.org/0000-0003-2428-272X",
58       "@type": "Person",
59       "name": "Chris Andrews",
60       "affiliation": { "@id": "https://ror.org/00pggkr55" }
61     },
62     {
63       "@id": "https://orcid.org/0009-0003-5290-786X",

```

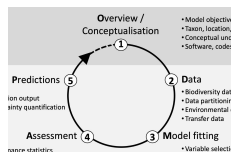
Model

conformsTo
license
@id / identifier
@type / additionalType
name
description
dateCreated
creator
keywords
 codeRepository
 contributor
 softwareVersion
 programmingLanguage
 softwareRequirements
 supportingData

 | **FAIR-IMPACT**



Hugging Face



Dataset

Kernel Attributes

url
 variableMeasured
 measurementTechnique
 spatialCoverage
 temporalCoverage
 encodingFormat
 contentSize
 version
 publisher

Workflow

Kernel Attributes

input
 output
 version
 programmingLanguage
 url
 sdPublisher

Mapping Set

Kernel Attributes

mapping_set_id
 mappings
 subject_id
 predicate_id
 object_id
 mapping_justification
 curation_rule
 mapping_set_version
 mapping_provider



**Biodiversity
Information
Standards**
 TDWG



BIODIVERSITY
BUILDING
BLOCKS FOR
POLICY



Bioschemas



COMMON
WORKFLOW
LANGUAGE



WorkflowHub

Workflow Run
 RO-Crate



SSSOM
 SIMPLE STANDARD FOR SHARING
ONTOLOGY MAPPINGS



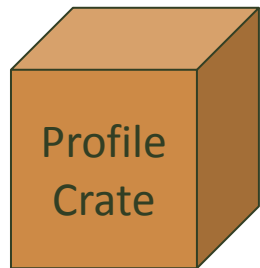
FAIRCORE4EOSC



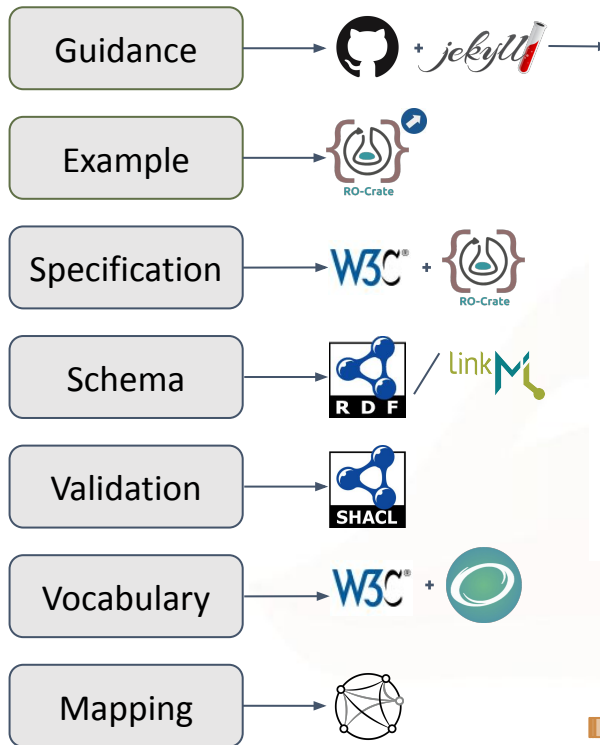
D5.2 Pilot integration of workflows and data products organised through FDO with common community workflows and data products used within the RIs

Working space →

<https://github.com/BioDT/biodt-fair>



PROF
vocabulary



Dataset Attributes

Format: the name of each metadata attribute includes a link to [Schema.org](https://schema.org) (or another vocabulary to which the property belongs), and is followed by a definition in the "Description" line. "Type" indicates the expected values for each property, while "Cardinality" specifies the amount and whether they are optional or mandatory. Lastly, additional remarks might be added as "Comments", and an example is given in the final line.

url

- **Description:** The location of a page describing the dataset.
- **Type:** [URL](#)
- **Cardinality:** 1/1
- **Comments:**
- **Example:** "url": "http://www.ebi.ac.uk/metabolights/MTBL5234"

variableMeasured

- **Description:** The variables that are measured in the dataset, either described as text or as pairs of identifier and description using `PropertyValue`, or more explicitly as a `StatisticalVariable`.
- **Type:**
 - `Property`
 - `PropertyValue`
 - `StatisticalVariable`
 - `Text`
- **Cardinality:** 1/many
- **Comments:** Often but not necessarily each `variableMeasured` will have an explicit representation as (or mapping to) an property such as those defined in [Schema.org](https://schema.org), or other RDF vocabularies and "knowledge graphs". In that case the subproperty of `variableMeasured` called `measuredProperty` is applicable.
- **Example:** "variablesMeasured": "species cover"

measurementTechnique

- **Description:** A technique, method or technology used for measuring the corresponding variable(s) in the dataset. The `measurementTechnique` property helps when extra clarification is needed about how a `measuredProperty` was measured.
- **Type:**
 - `DefinedTerm`
 - `MeasurementMethodEnum`
 - `Text`

➞ MS27 Semantic standards used for data, quality and workflows

→ By Christoph Wohner

- ❑ Deposit data on [B2Share](https://b2share.eudat.eu/)
 - ❑ Under LTER Domain
 - ❑ Add site references:
<https://deims.org/11696de6-0ab9-4c94-a06b-7ce40f56c964>
 - ❑ BioDT keywords (e.g. [BioDT project DOI](#) and “Grassmind pDT”)
- ❑ Generate metadata records on eLTER DAR:
<https://catalogue.lter-europe.net/elter/documents>
- ❑ Write workflow to generate FDOs based on DAR metadata




[HELP](#)
[COMMUNITIES](#)
[UPLOAD](#)
[CONTACT](#)

RECORDS » 6D43BD5AC35945A6AE749DC16CAFB9EF

Grassland Dynamics - Schrankogel (Austria) - 1994, 2004, 2014

Apr 30, 2024

Abstract: species cover according to the GLORIA Field Manual: https://gloria.ac.at/downloads/Manual_5thEd_ENG.pdf (p 38); 1x1m plots arranged in transects at southern slopes of Mt. Schrankogel: <https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.15290>; visual estimate used for the Horizon Europe project “Biodiversity Digital Twin for Advanced Modelling, Simulation and Prediction Capabilities” (<https://biodyn.eu/>).

Methods: <https://doi.org/10.2777/867331>

Disciplines: 3.1.13 → Biology → Ecology

Keywords: BioDT; Grassland pDT;

DOI: [10.23728/b2share.6d43bd5ac35945a6ae749dc16cafb9ef](https://doi.org/10.23728/b2share.6d43bd5ac35945a6ae749dc16cafb9ef)

PID: [11304/9a4d8ab0-e34d-45b1-a77a-e330ace65066](https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.15290)

Views 4

File Downloads 10

Files	5
Total Size	3.8 MB

Share

[Twitter](#) [Facebook](#)

Cite as

[Unknown]. (2024). Grassland Dynamics - Schrankogel (Austria) - 1994, 2004, 2014 [Data set]. <https://b2share.eudat.eu/>
<https://doi.org/10.23728/b2share.6d43bd5ac35945a6ae749dc16cafb9ef>
[Copy BibTeX](#)
 More citation choices

Files

<https://doi.org/10.23728/b2share.6d43bd5ac35945a6ae749dc16cafb9ef>

Basic metadata

Open access True ✓

LTER Metadata

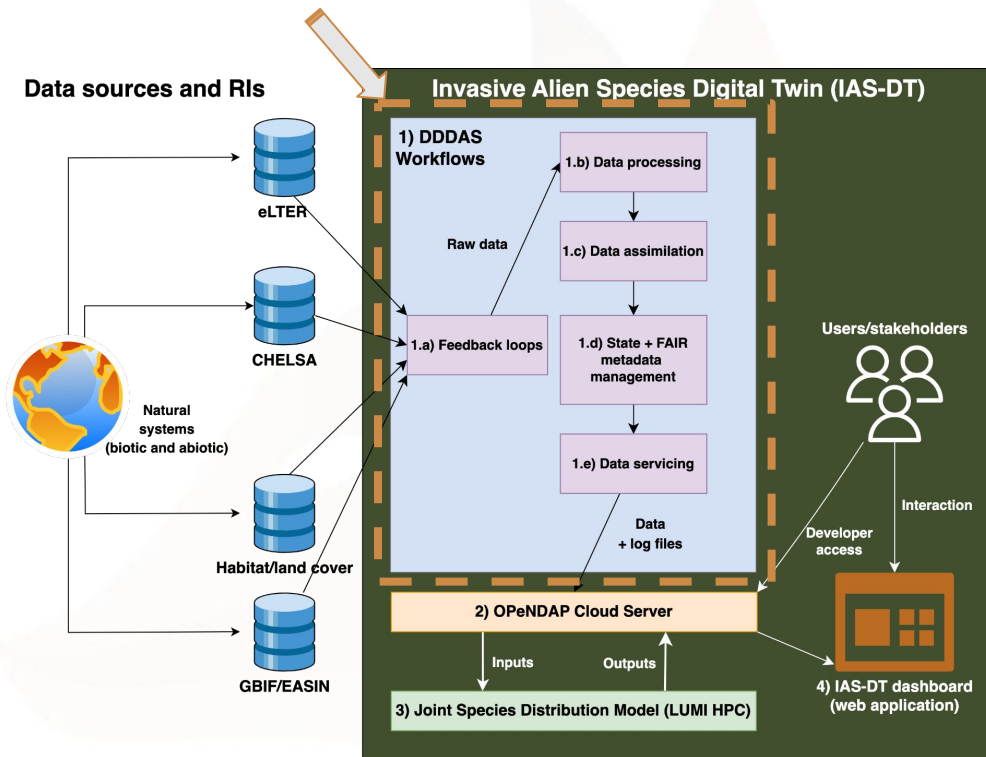
Metadata URL <https://deims.org/11696de6-0ab9-4c94-a06b-7ce40f56c964>

(Adapted from D6.1 - Pipeline Release and Validation)

“DDDAS is a conceptual framework that synergistically combines models and data to facilitate the analysis and prediction of physical phenomena. It is a two-layer system design where data assimilation and feedback loops create a bidirectional information flow.”

There are four types of metadata in the IASDT:

1. Feedback-loop metadata: JSON files that track changes in the data sources.
2. **State metadata:** JSON files that show the difference between consecutive input data from workflow run to workflow run (i.e., what has changed over time).
3. Runtime metadata: Log files for the workflow runs.
4. Model metadata: Log files of the model runs + JSON files.



DoIt Automation Tool

doit comes from the idea of bringing the power of **build-tools** to execute any kind of **task**



```
pip install doit
```

People often compare **doit** to tools like *make*, *grunt*, *rake*, *scons*, *snakemake*.

They appreciate **doit** strong features, flexibility, simplicity of authoring and ease of use.

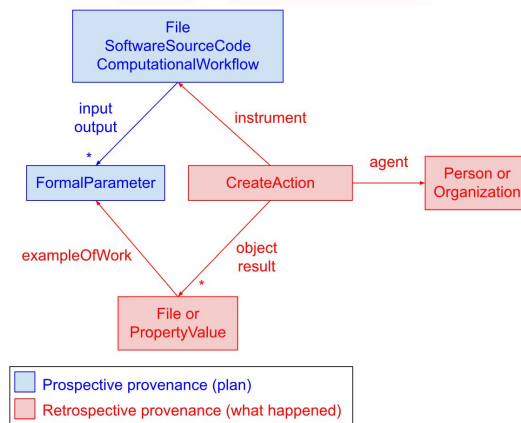
A **bioinformatics developer** uses **doit** to

- CREATE A REPRODUCEABLE COMPUTATIONAL PIPELINE
- MANAGE A COMPLEX WORKFLOW (SET OF DEPENDING TASKS) <https://pydoit.org/index.html>
- OPTIMIZE PROCESSING TIME BY SKIPPING TASKS ALREADY DONE

[View all use-cases](#)

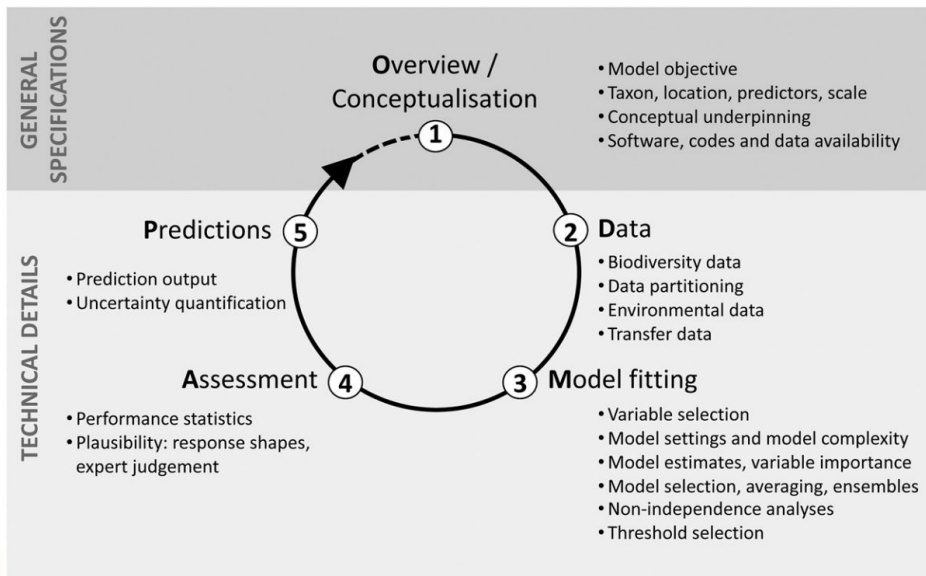


Workflow Run RO-Crate



<https://www.researchobject.org/workflow-run-crate/>

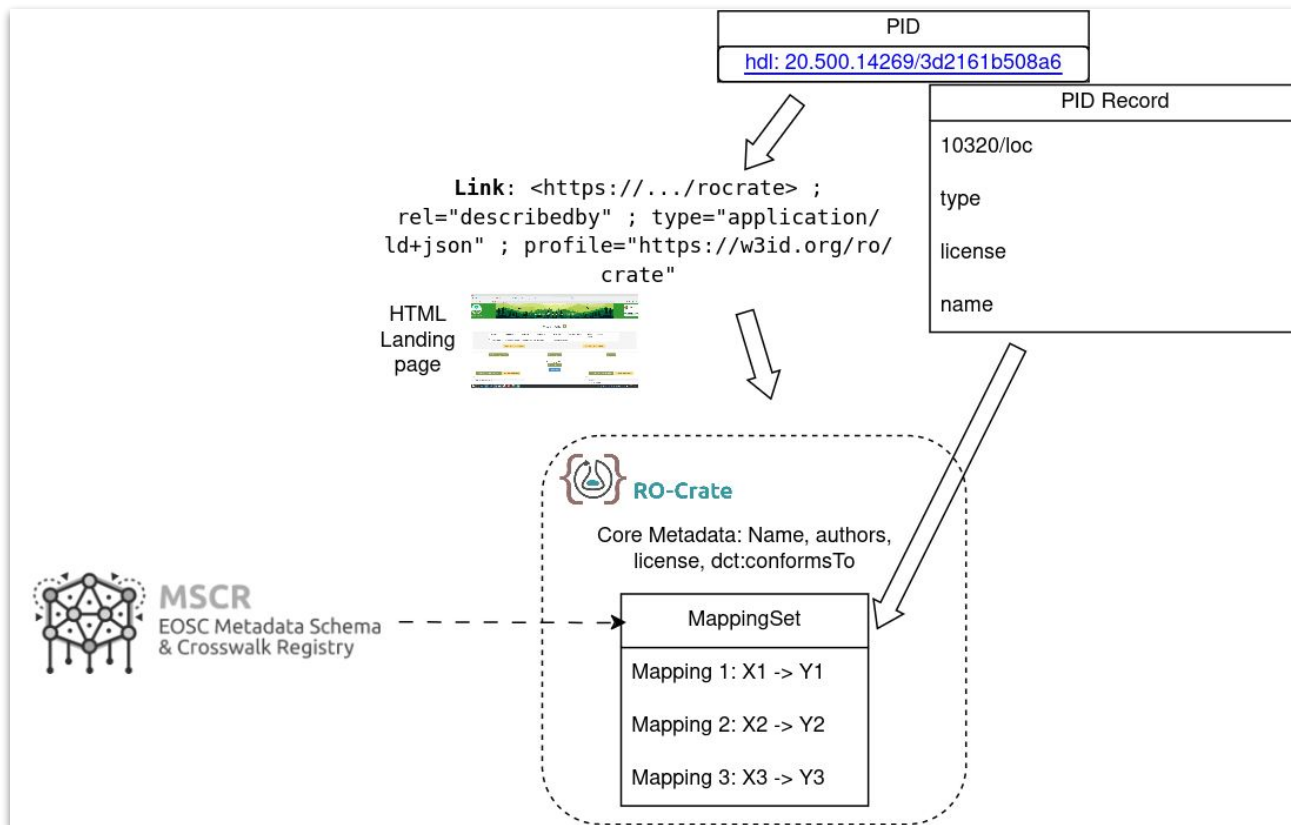
- ❑ [ODMAP protocol](#) for reporting species distribution models
- ❑ Progress made on generating the human-readable report from code
- ❑ Could have an equivalent RO-Crate metadata file



<https://odmap.wsl.ch/>

mapping.bio - mapping editor





Other aspects not covered...

- ☐ Publishing
- ☐ Workflows
- ☐ “Open access”
- ☐ FOSS (Free and Open source Software)
- ☐ Linked Open Data
- ☐ Open Hardware
- ☐ Reproducible papers
- ☐ Version Control
- ☐ Best software engineering practices
- ☐ ...



BIODT

biodiversitydigitaltwin

School & Hackaton





BIODT

biodiversitydigitaltwin

School & Hackaton

Break: Open Science “Escape Room”



Games of Horror for Open Science Training



Games



GitHub

The GHOST collective is a group of Open Science enthusiasts from universities across the Netherlands. We make Open Science games that can be used as training materials or just for fun.

<https://ghostcollective.github.io/>



Data
Horror
Week



Data Horror Escape Room

The Data Horror Escape Room is an online game to explore the tricks and treats of research data management.



Software Horror Escape Room

The Software Horror Escape Room is an online game to explore the tricks and treats of research software management.



Open Science Escape Room

The Open Science Escape Room is an online game to explore the tricks and treats of Open Science.



Open Science Against Humanity

Open Science Against Humanity is a card game based on Cards Against Humanity, but with an Open Science twist!



BIODT

biodiversitydigitaltwin

School & Hackaton

Tools and Resources

(see the last slides or Training Environment for the links!)

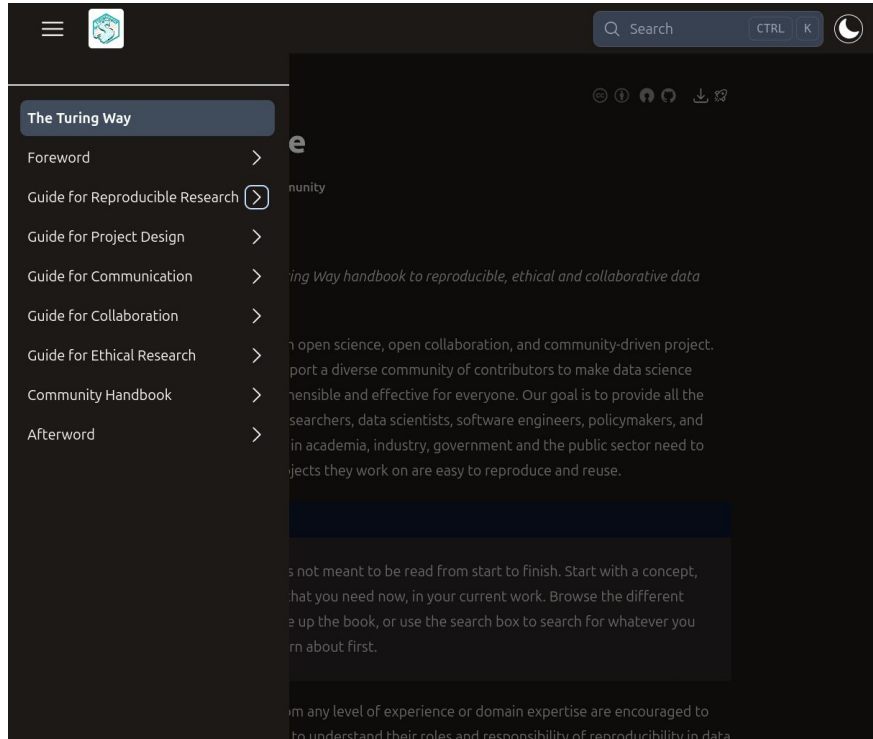


FIVE RECOMMENDATIONS FOR FAIR SOFTWARE

LET'S GO! →



ENDORSE



The Research Data Management toolkit for Life Sciences

Best practices and guidelines to help you make your data FAIR (Findable, Accessible, Interoperable and Reusable)

What can we help you find?

 Search RDMkit

Browse all topics by



Data life cycle

Start here to get an overview of research data management based on stages in the data life cycle.



Your role

Identify your role in research data management, find data management resources relevant for you, and information to help you progress in your career path.



Your domain

Learn about data management tasks that affect your domain or research community, and the solutions adopted to address them.



Your tasks

Find guidelines and solutions for tackling common data management tasks.



Tool assembly

Find concrete combinations of tools and resources assembled into an ecosystem for research data management.



National resources

Find pointers to country specific information resources and national research data management practices.

The **FAIR** Cookbook

for FAIR doers

An online, **open** and **live** resource for the Life Sciences with recipes that help you to make and keep data Findable, Accessible, Interoperable and Reusable; in one word FAIR.

[LEARN MORE](#)[SEE RECIPES](#)



Galaxy Training!

FAIR Data, Workflows, and Research

Learning Pathways



Galaxy Training!

FAIR Data, Workflows, and Research

Learning Pathways



Galaxy Training!

FAIR Data, Workflows, and Research

RO-Crate - Introduction

Author(s)   Stian Soiland-Reyes   Peter Sefton

Reviewers  

Overview

Questions:

- How do I package data in a FAIR way?
- How can I list the authors of individual files?
- Can I use multiple licenses in the same data package?
- How can I visualize JSON-LD metadata?

Objectives:

- Construct an RO-Crate by hand using JSON
- Describe each part of the Research Object
- Learn basic JSON-LD to create FAIR metadata
- Connect different parts of the Research Object using identifiers

RO-Crate in Python

Author(s)   Simone Leo   Bruno P. Kinoshita

Editor(s)   Helena Rasche   Eli Chadwick

Tester(s)   Eli Chadwick

Reviewers     

Overview

Questions:

- What data is contained within an RO-Crate
- How can I create an RO-Crate myself?

Objectives:

- Create a custom, annotated RO-Crate
- Use ORCID IDs and other linked data to annotate datasets contained within the crate

Workflow Run RO-Crate

Author(s)   Simone Leo

Reviewers 

Overview

Questions:

Objectives:

Time estimation: 30 minutes

Supporting Materials:

[FAQs](#) [Available on these Galaxies](#)

Published: May 11, 2023

Last modification: May 11, 2023

License: Tutorial Content is licensed under [Apache-2.0](#) T

PURL: <https://gxy.io/GTN:T00343>

Rating: 4.0 (1 recent ratings, 1 all time)

Revision: 1



GitHub Skills

Learn how to use GitHub with interactive courses designed for beginners and experts.

[Start with Introduction to GitHub](#)

Our courses

First day on GitHub

Introduction to GitHub

Get started using GitHub in less than an hour.

Communicate using Markdown

Organize ideas and collaborate using Markdown, a lightweight language for text formatting.

GitHub Pages

Create a site or blog from your GitHub repositories with GitHub Pages.

First week on GitHub

Review pull requests

Collaborate and work together on GitHub.

Resolve merge conflicts

Learn why conflicts happen and how to resolve them.

Release-based workflow

Practice a release-based workflow and explore branching strategies.

Connect the dots

Find relevant conversations, commits, and projects in a repository.

Code with Codespaces

Develop code using GitHub Codespaces and Visual Studio Code.



git --local-branching-on-the-cheap

About

Documentation

- Reference
- Book
- Videos
- External Links

Downloads

Community

Documentation

Reference



Reference Manual

The official and comprehensive **man pages** that are included in the Git package itself.

Quick reference guides: [GitHub Cheat Sheet](#) | [Visual Git Cheat Sheet](#)

Book



Pro Git

The entire **Pro Git book** written by Scott Chacon and Ben Straub is available to [read online for free](#). Dead tree versions are available on [Amazon.com](#).

Videos



Git Basics: What is Git?

What is Git?
Length: 08:15



Git Basics: What is Version Control?

What is Version Control?
Length: 05:59

**GitLab**

Docs

Getting started

Tutorials

[Find your way around GitLab](#)

Learn GitLab with tutorials

S&H

Choose an open source license

An open source license protects contributors and users. Businesses and savvy developers won't touch a project without this protection.

Which of the following best describes your situation?



I need to work in a community.

Use the [license preferred by the community](#) you're contributing to or depending on. Your project will fit right in.

If you have a dependency that doesn't have a license, ask its maintainers to [add a license](#).



I want it simple and permissive.

The [MIT License](#) is short and to the point. It lets people do almost anything they want with your project, like making and distributing closed source versions.

[Babel](#), [.NET](#), and [Rails](#) use the MIT License.



I care about sharing improvements.

The [GNU GPLv3](#) also lets people do almost anything they want with your project, *except* distributing closed source versions.

[Ansible](#), [Bash](#), and [GIMP](#) use the GNU GPLv3.

What if none of these work for me?

My project isn't software.

There are licenses for that.

I want more choices.

More licenses are available.

I don't want to choose a license.

Here's what happens if you don't.



Licensing Assistant

📁 EUPL 88 64 members

Topics: Legal

Overview

Compatibility Checker

Find and compare software licenses



Products ▾

Developers ▾

Pricing

Support

Blog

Company ▾

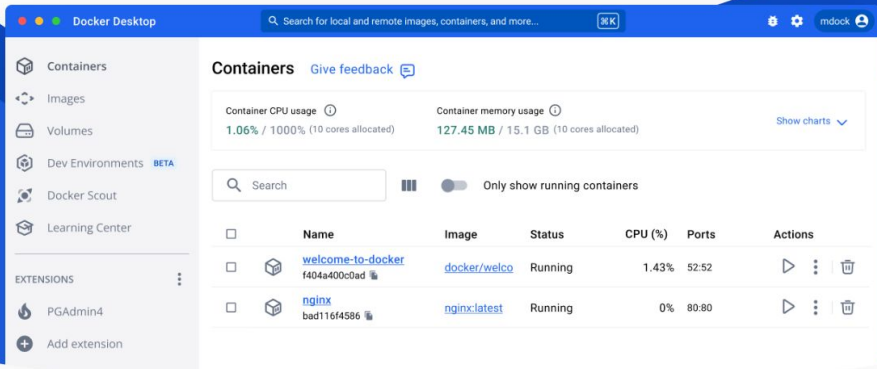


Sign In

Get started

Develop faster. Run anywhere.

Build with the **#1** most-used developer tool

[Download Docker Desktop](#)[Learn more about Docker](#)

pod...

[Features](#)[Get Started](#)[Community](#)

The best free & open source container tools

Manage containers, pods, and images with Podman. Seamlessly work with containers and Kubernetes from your local environment.

[Get Started](#)[Download ▾](#)

Latest stable
Podman 5.4.1

- Latest stable Podman
Desktop 1.17.1

- Apache
License 2.0

European alternatives for digital products

We help you find European alternatives for digital service and products, like cloud services and SaaS products.



Support local businesses

When you buy from local businesses, you are supporting yourself down the road. Taxes paid by the company come back to you indirectly and the company creates jobs in your region.



VAT / Billing

As a business that operates in Europe, it is possible to get a VAT refund for products/services of other European companies. European companies also tend to offer payment methods that are commonly used in Europe.



Data protection / GDPR

Some companies outside Europe tend to ignore data protection and related laws such as the GDPR or do not implement them correctly.



Similar legal requirements

Within the EU, many laws and framework conditions are set by the EU, which helps to cover a large market without having to consider large country-specific differences. It is also easier to enforce your rights against another company located in the EU.

- [Five recommendations for FAIR software](#)
- [The Turing Way](#)
- [RDMkit](#)
- [The FAIR Cookbook](#)
- RO-Crate tutorials:
 - [RO-Crate Introduction](#)
 - [RO-Crate in Python](#)
 - [Workflow Run RO-Crate Introduction](#)
- Version Control Systems
 - [Git Documentation](#)
 - [GitLab Tutorials](#)
 - [GitHub Skills](#)
- Licenses
 - [Choose an open source license](#)
 - Interoperable Europe - [Licensing Assistant](#)
- Containers
 - [Podman](#)
 - [Docker](#)
- [European alternatives for digital products](#)
- [Guidelines for recommended metadata standard for research software within EOSC](#)
 - Particularly, *Appendix B: RSMD Checklist*, on pages 38–40



BIODT

biodiversitydigitaltwin

School & Hackaton



@BiodiversityD



T
BioD
T



Funded by
the European Union



BIODT

biodiversitydigitaltwin

School & Hackaton

References



B2SHARE. (n.d.). Retrieved 18 March 2025, from <https://b2share.eudat.eu/>

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), 452–454. <https://doi.org/10.1038/533452a>

Bausell, R. B. (2021). *The Problem with Science: The Reproducibility Crisis and what to Do about it*. Oxford University Press.

B-Cubed. (n.d.). Retrieved 18 March 2025, from <https://b-cubed.eu/>

Biodiversity and Ecosystem Research—LifeWatch ERIC Home. (n.d.). LifeWatch ERIC. Retrieved 18 March 2025, from <https://www.lifewatch.eu/>

Biodiversity Digital Twin for Advanced Modelling, Simulation and Prediction Capabilities | BioDT Project | Fact Sheet | HORIZON. (n.d.). CORDIS | European Commission. Retrieved 18 March 2025, from <https://cordis.europa.eu/project/id/101057437>

Bioschemas—Bioschemas. (n.d.). Retrieved 18 March 2025, from <https://bioschemas.org/>

Connected Open Identifiers for Discovery, Access and Use of Research Resources | FREYA Project | Fact Sheet | H2020. (n.d.). CORDIS | European Commission. Retrieved 18 March 2025, from <https://cordis.europa.eu/project/id/777523>

Crusoe, M. R., Abeln, S., Iosup, A., Amstutz, P., Chilton, J., Tijanić, N., Ménager, H., Soiland-Reyes, S., Gavrilović, B., Goble, C., & Community, T. C. (2022). Methods included: Standardizing computational reuse and portability with the Common Workflow Language. *Communications of the ACM*, 65(6), 54–63. <https://doi.org/10.1145/3486897>

De Smedt, K., Koureas, D., & Wittenburg, P. (2020). FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. *Publications*, 8(2), 21. <https://doi.org/10.3390/publications8020021>

Destination Earth | Shaping Europe's digital future. (2023, February 15). <https://digital-strategy.ec.europa.eu/en/policies/destination-earth>

Distributed System of Scientific Collections. (n.d.). *Open Digital Specimen specification*. <https://github.com/DiSSCo/openDS>

doit: Doit - Automation Tool (Version 0.36.0). (n.d.). [Python; OS Independent, POSIX]. Retrieved 18 March 2025, from <http://pydoit.org>

European Organization For Nuclear Research & OpenAIRE. (2013). *Zenodo*. CERN. <https://doi.org/10.25495/7GXK-RD71>

FAIRCORE4EOSC. (n.d.). Retrieved 18 March 2025, from <https://faircore4eosc.eu/>

Games of Horror for Open Science Training. (n.d.). Retrieved 18 March 2025, from <https://ghostcollective.github.io/>

GBIF. (n.d.). Retrieved 18 March 2025, from <https://www.gbif.org/>

Gray, A., Goble, C., & Jiménez, R. (2017). *Bioschemas: From Potato Salad to Protein Annotation*. International Workshop on the Semantic Web.

<https://www.semanticscholar.org/paper/Bioschemas%3A-From-Potato-Salad-to-Protein-Annotation-Gray-Goble/74eca9c89622b7f731b21b03acb4f2400a0f00fa>

Gruenpeter, M., Granger, S., Monteil, A., Chue Hong, N., Breitmoser, E., Antonioletti, M., Garijo, D., González Guardia, E., Gonzalez Beltran, A., Goble, C., Soiland-Reyes, S., Juty, N., & Mejias, G. (2023).

D4.4—Guidelines for recommended metadata standard for research software within EOSC. <https://zenodo.org/record/8097537>

Gustafsson, O. J. R., Wilkinson, S. R., Bacall, F., Pireddu, L., Soiland-Reyes, S., Leo, S., Owen, S., Juty, N., Fernández, J. M., Grüning, B., Brown, T., Ménager, H., Capella-Gutierrez, S., Coppens, F., & Goble, C.

(2024). *WorkflowHub: A registry for computational workflows* (No. arXiv:2410.06941). arXiv. <https://doi.org/10.48550/arXiv.2410.06941>

Hernández, J. A., & Colom, M. (2023). *Repeatability, Reproducibility, Replicability, Reusability (4R) in Journals' Policies and Software/Data Management in Scientific Publications: A Survey, Discussion, and Perspectives*. <https://hal.science/hal-04322522>

Home Page—Software Heritage. (n.d.). Retrieved 18 March 2025, from <https://www.softwareheritage.org/>

Jekyll • Simple, blog-aware, static sites. (n.d.). Jekyll • Simple, Blog-Aware, Static Sites. Retrieved 18 March 2025, from <https://jekyllrb.com/>

Jonquet, C., Graybeal, J., Bouazzouni, S., Dorf, M., Fiore, N., Kechagioglou, X., Redmond, T., Rosati, I., Skrenchuk, A., Vendetti, J. L., & Musen, M. (2023). *Ontology Repositories and Semantic Artefact Catalogues with the OntoPortal Technology*. *Semantic Web, ISWC 2023. ISSN: 0302-9743, 14266*, 38–58. https://doi.org/10.1007/978-3-031-47243-5_3

JSON-LD - JSON for Linking Data. (n.d.). Retrieved 25 May 2023, from <https://json-ld.org/>

Korbmacher, M., Azevedo, F., Pennington, C. R., Hartmann, H., Pownall, M., Schmidt, K., Elsherif, M., Breznau, N., Robertson, O., Kalandadze, T., Yu, S., Baker, B. J., O'Mahony, A., Olsnes, J. Ø.-S., Shaw, J. J., Gjoneska, B., Yamada, Y., Röer, J. P., Murphy, J., ... Evans, T. (2023). The replication crisis has led to positive structural, procedural, and community changes. *Communications Psychology*, 1(1), 1–13. <https://doi.org/10.1038/s44271-023-00003-2>

Leo, S. (n.d.). *Workflow run RO-crate introduction (galaxy training materials)*. [url{https://training.galaxyproject.org/training-material/topics/fair/tutorials/ro-crate-workflow-run-ro-crate/tutorial.html}](https://training.galaxyproject.org/training-material/topics/fair/tutorials/ro-crate-workflow-run-ro-crate/tutorial.html)

Leo, S., Crusoe, M. R., Rodríguez-Navas, L., Sirvent, R., Kanitz, A., Geest, P. D., Wittner, R., Pireddu, L., Garijo, D., Fernández, J. M., Colonnelli, I., Gallo, M., Ohta, T., Suetake, H., Capella-Gutierrez, S., Wit, R. de, Kinoshita, B. P., & Soiland-Reyes, S. (2024). Recording provenance of workflow runs with RO-Crate. *PLOS ONE*, 19(9), e0309210. <https://doi.org/10.1371/journal.pone.0309210>

Leo, S., & Kinoshita, B. P. (n.d.). *RO-crate in python (galaxy training materials)*. [url{https://training.galaxyproject.org/training-material/topics/fair/tutorials/ro-crate-in-python/tutorial.html}](https://training.galaxyproject.org/training-material/topics/fair/tutorials/ro-crate-in-python/tutorial.html)

Matentzoglou, N., Balhoff, J. P., Bello, S. M., Bizon, C., Brush, M., Callahan, T. J., Chute, C. G., Duncan, W. D., Evelo, C. T., Gabriel, D., Graybeal, J., Gray, A., Gyori, B. M., Haendel, M., Harmse, H.,

Harris, N. L., Harrow, I., Hegde, H. B., Hoyt, A. L., ... Mungall, C. J. (2022). A Simple Standard for Sharing Ontological Mappings (SSSOM). *Database*, 2022, baac035.

<https://doi.org/10.1093/database/baac035>

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. <https://doi.org/10.1145/3287560.3287596>

Miyakawa, T. (2020). No raw data, no science: Another possible source of the reproducibility crisis. *Molecular Brain*, 13(1), 24. <https://doi.org/10.1186/s13041-020-0552-2>

Moxon, S., Kalita, P., Saunders, J., Hegde, H., Patil, S., Schaper, K., Harris, N., Cox, C., Linke, D., Unni, D., Vaidya, G., Putman, T., Solbrig, H., All GitHub Contributors, <https://github.com/linkml/linkml/graphs/contributors>, Haendel, M., & Mungall, C. (n.d.). *LinkML* [Computer software]. <https://github.com/linkml/linkml>

ORCID. (n.d.). Retrieved 18 March 2025, from <https://orcid.org/>

Pauli, H. (2024). *Grassland Dynamics—Schränkogel (Austria)—1994, 2004, 2014* (p. 5 files, 4.0 MB) [Dataset]. <https://b2share.eudat.eu>.

<https://doi.org/10.23728/B2SHARE.6D43BD5AC35945A6AE749DC16CAFB9EF>

Pensoft. (n.d.). *eLTER Homepage*. Retrieved 18 March 2025, from <https://elter-ri.eu/>

RDF - Semantic Web Standards. (n.d.). Retrieved 18 March 2025, from <https://www.w3.org/RDF/>

RDMkit. (n.d.). Retrieved 18 March 2025, from https://rdmkit.elixir-europe.org/machine_actionability.html

Research Activity Identifier. (n.d.). Research Activity Identifier. Retrieved 18 March 2025, from <https://raid.org>

Research Organization Registry (ROR). (n.d.). Research Organization Registry (ROR). Retrieved 18 March 2025, from <https://ror.org/>

Schema.org—Schema.org. (n.d.). Retrieved 25 May 2023, from <https://schema.org/>

Science has been in a “replication crisis” for a decade. Have we learned anything? | Vox. (n.d.). Retrieved 18 March 2025, from

<https://www.vox.com/future-perfect/21504366/science-replication-crisis-peer-review-statistics>

Shapes Constraint Language (SHACL). (2017, July 20). <https://www.w3.org/TR/shacl/>

Soiland-Reyes, S., & Goble, C. (2023, April 14). *Building diverse FDO Collections using RO-Crate: An overview of the RO-Crate concept and its implementations*. <https://doi.org/10.5281/zenodo.7915603>

Soiland-Reyes, S., & Sefton, P. (n.d.). *RO-crate—Introduction (galaxy training materials)*. <url{https://training.galaxyproject.org/training-material/topics/fair/tutorials/ro-crate-intro/tutorial.html}>

Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L. J., Coppens, F., Fernández, J. M., Garijo, D., Grüning, B., La Rosa, M., Leo, S., Ó Carragáin, E., Portier, M., Trisovic, A., RO-Crate Community, Groth, P., & Goble, C. (2022). Packaging research artefacts with RO-Crate. *Data Science*, 5(2), 97–138. <https://doi.org/10.3233/DS-210053>

Solve the Reproducibility Crisis with Software Reproducibility. (n.d.). Retrieved 18 March 2025, from <https://www.activestate.com/blog/trust-security-and-the-reproducibility-crisis-in-software/>

The Reproducibility Crisis in Science—These Researchers Have a Fix—USC Viterbi | School of Engineering. (n.d.). Retrieved 18 March 2025, from

<https://viterbischool.usc.edu/news/2022/11/the-reproducibility-crisis-in-science-these-researchers-have-a-fix/>

The Turing Way Community, & Scriberia. (2024). *Illustrations from The Turing Way: Shared under CC-BY 4.0 for reuse* [Graphic]. Zenodo. <https://doi.org/10.5281/ZENODO.3332807>

What is a DOI? (n.d.). Retrieved 18 March 2025, from <https://www.doi.org/the-identifier/what-is-a-doi/>

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., & Vieglais, D. (2012). Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE*, 7(1), e29715. <https://doi.org/10.1371/journal.pone.0029715>

Wikipedia contributors. (2025a). *ISBN — Wikipedia, the free encyclopedia*. <https://en.wikipedia.org/w/index.php?title=ISBN&oldid=1279630315>

Wikipedia contributors. (2025b). *Metadata—Wikipedia, the free encyclopedia*. <https://en.wikipedia.org/w/index.php?title=Metadata&oldid=1278695576>

Wikipedia contributors. (2025c). *Replication crisis—Wikipedia, the free encyclopedia*. https://en.wikipedia.org/w/index.php?title=Replication_crisis&oldid=1277101005

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018.

<https://doi.org/10.1038/sdata.2016.18>

Your tasks: *Machine actionability* | *RDMkit*. (n.d.). Retrieved 18 March 2025, from https://rdmkit.elixir-europe.org/machine_actionability.html

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillerá-Aroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., ... Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43(9), 1261–1277.

<https://doi.org/10.1111/ecog.04960>