# Building FAIR research repositories in practice

Lars Holm Nielsen

Head of Open Science Infrastructure CERN, IT Department

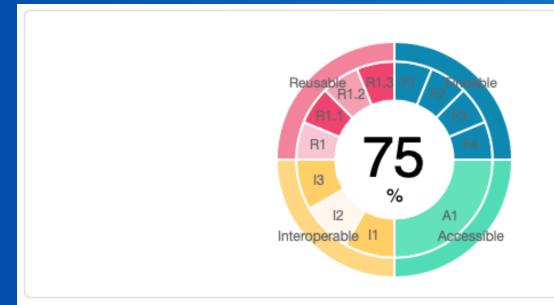




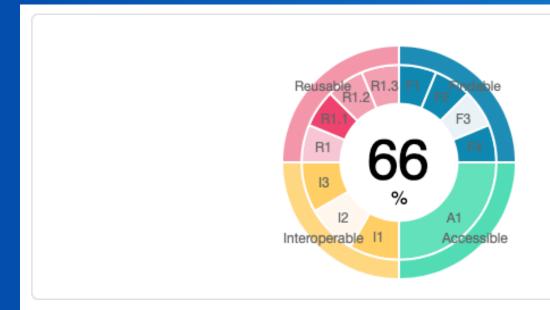
## Make it FAIR



## Zenodo: State of FAIRness



	Score earned:	Fair level:
Findable:	7 of 7	advanced
Accessible:	2 of 3	moderate
Interoperable:	3 of 4	moderate
Reusable:	6 of 10	moderate



	Score earned:	Fair level:
Findable:	6 of 7	moderate
Accessible:	2 of 3	moderate
Interoperable:	3 of 4	moderate
Reusable:	5 of 10	moderate



## Zenodo: State of FAIRness



Search records...

Q

Communities My dashboard

→ Log in

🕜 Sign up



Industrial Ecology and Sustainability Research

Published October 21, 2021 | Version 3.8.2

Dataset



#### JZ I | VEISION 3.0.

#### **EXIOBASE 3**

Stadler, Konstantin 1 (1); Wood, Richard 1 (1); Bulavskaya, Tatyana 2; Södersten, Carl-Johan 1;

Simas, Moana 1 (b): Schmidt, Sarah 1; Usubiaga, Arkaitz 3 (b): Acosta-Fernández, José 3;

Kuenen, Jeroen<sup>2</sup>; Bruckner, Martin<sup>4</sup>; Giljum, Stefan<sup>4</sup>; Lutter, Stephan<sup>4</sup>; Merciai, Stefano<sup>5</sup>;

Schmidt, Jannick H<sup>5</sup>; Theurl, Michaela C<sup>6</sup>; Plutzar, Christoph<sup>6</sup>; Kastner, Thomas<sup>7</sup> (D);

Eisenmenger, Nina<sup>6</sup>; Erb, Karl-Heinz<sup>6</sup>; Koning, Arjan<sup>8</sup>; Tukker, Arnold<sup>8</sup> (D)

Show affiliations

**EXIOBASE 3** provides a time series of environmentally extended multi-regional input-output (EE MRIO) tables ranging from 1995 to a recent year for 44 countries (28 EU member plus 16 major economies) and five rest of the world regions. EXIOBASE 3 builds upon the previous versions of EXIOBASE by using rectangular supply-use tables (SUT) in a 163 industry by 200 products classification as the main building blocks. The tables are provided in current, basic prices (Million EUR).

For any questions regarding access, support or licence clarification please email: exiobase-support@googlegroups.com . The database is provided free of charge to users under a CC-BY-SA license. There is a discussion about different licence options, please reach out for information. For help in use of EXIOBASE data for spend-based emission factors, email exiobase-support@googlegroups.com

EXIOBASE 3 is the culmination of work in the FP7 DESIRE project and builds upon earlier work on EXIOBASE 2 in the FP7 CREEA project and EXIOBASE 1 of the FP6 EXIOPOL project. These databases are available at the official EXIOBASE website.

A special issue of Journal of Industrial Ecology (Volume 22, Issue 3) describes the build process and some use cases of EXIOBASE 3. This includes the article by Stadler et. al 2018 describing the compilation of EXIOBASE 3. Further informations (data quality, updates, ...) can be found in the blog post describing a previous release at the Environmental Footprints webpage. Various concordance tables for the database are available here.

For more (background) information see the Readme file. For any questions regarding access, support or licence options please email: exiobase-support@googlegroups.com

 132K ★ DOWNLOADS

Mar 8, 2021

Nov 19, 2020

Dec 18, 2019

Show more details

#### Versions

Version 3.8.2	Oct 21, 2021
10.5281/zenodo.5589597	

10.5281/zenodo.4588235

Version 3.8

Version 3.8.1

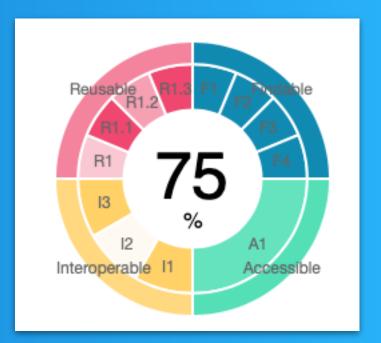
10.5281/zenodo.4277368

Version 3.7

10.5281/zenodo.3583071

#### View all 4 versions

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.3583070. This DOI represents all versions, and will always resolve to the latest one. Read more.





## Zenodo: State of FAIRness

zenodo

Search records...

Q

Communities

My dashboard

Published June 1, 2021 | Version v1



#### iPhone XS Case

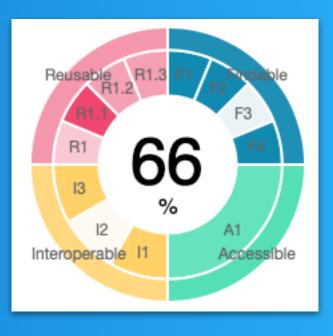
iPhone XS Case

iPhone XS Case- The Best Type of Case For Your iPhone

There are many iPhone XS Cases to choose from, so it is essential to find the best one for your phone. However, many of the cases are similar and may not offer the type of protection you need. Before choosing an iPhone XS Case, make sure you know what is most important to you as a user. For example, if you frequently use your phone outdoors or in a harsh environment, you will need something more durable than a cheap plastic case. Fortunately, there are many top-rated cases from top companies like iPhone and case sealer cases. These three iPhone XS cases come highly recommended, but in case you also vouch for their performance, be sure to check out the other great options too.

This iPhone XS case offers an effortless yet professional look. It has smooth lines and is made from high-quality silicone that fits your phone perfectly. The cutout for the camera and dock area is positioned so that it is difficult to knock against without damaging your device. The overall size of this iPhone XS case is slightly larger than that of the iPhone Plus case, and for this reason, it can take a little more effort to fit into your pocket. If you will be carrying your iPhone xs in your pocket constantly, then this cutout is a necessity.

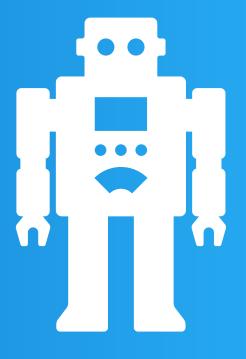
Best-Selling iPhone XS Cases







Goal: Reuse



US\$ 1.7 trillion
Global spending on R&D

Source: UNESCO Institute for Statistics

## Experiences from our FAIR journey







# 400.000 users 9000 organisations 161 countries

Total records

4,558,565

Records

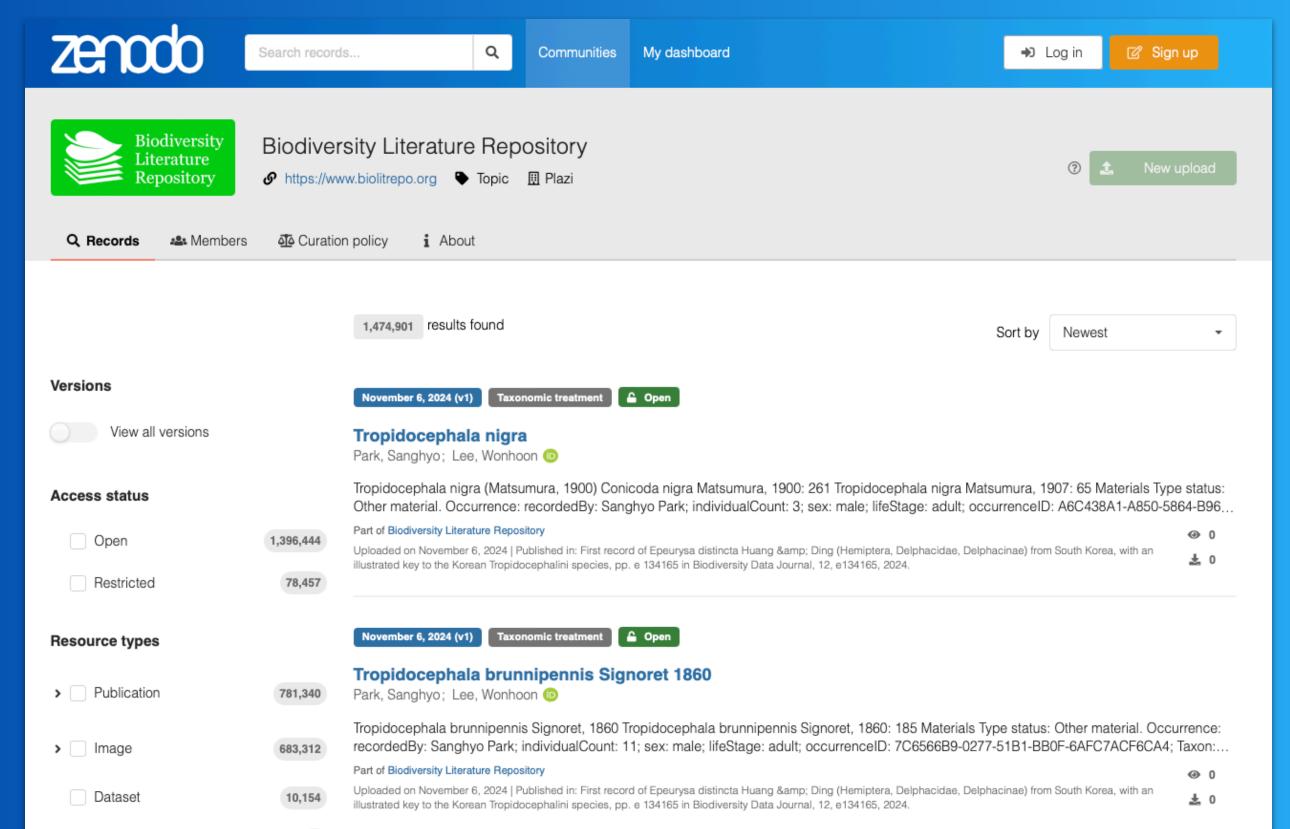
Total data volume

1,003.4TB

Data volume



# Biodiversity Literature Repository



# Taxonomic treatments

Describe the discovery of new biological species

#### **Example:**

Journal article describing 22 new millipedes, published in European Journal of Taxonomy





ISSN 2118-9773 www.europeanjournaloftaxonomy.eu 2018 · Enghoff H

This work is licensed under a Creative Commons Attribution 3.0 License.

#### Managranh

urn:lsid:zoobank.org:pub:852A3F68-B728-413A-B12E-56F306D56C35

A mountain of millipedes VII: The genus *Eviulisoma* Silvestri, 1910, in the Udzungwa Mountains, Tanzania, and related species from other Eastern Arc Mountains. With notes on *Eoseviulisoma* Brolemann, 1920, and *Suohelisoma* Hoffman, 1963 (Diplopoda, Polydesmida, Paradoxosomatidae)

#### Henrik ENGHOFF

Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 København Ø, Denmark.

Email: henghoff@snm.ku.dk

urn: lsid: zoobank. org: author: FB09A817-000D-43C3-BCC4-2BC1E5373635

Abstract. Twenty-two new species of the genus Eviulisoma Silvestri, 1910, from the Eastern Arc Mountains, Tanzania, are described: E. acaciae sp. nov., E. aequilobatum sp. nov., E. akkariae sp. nov., E. angulatum sp. nov., E. articulatum sp. nov., E. biquintum sp. nov., E. breviscutum sp. nov., E. cetafi sp. nov., E. chitense sp. nov., E. commelina sp. nov., E. coxale sp. nov., E. ejti sp. nov., E. grumslingslak sp. nov., E. kalimbasiense sp. nov., E. navuncus sp. nov., E. nessiteras sp. nov., E. ottokrausi sp. nov., E. paradisiacum sp. nov., E. sternale sp. nov. and E. zebra sp. nov. from the Udzungwa Mts, E. culter sp. nov. from the Rubeho Mts and E. kangense sp. nov. from the Kanga Mts. Eviulisoma kwabuniense Kraus, 1958, and E. dabagaense Kraus, 1958, both from the Udzungwa Mts, are redesribed based on new material. Notes are provided on E. iuloideum (Verhoeff, 1941) based on type material. Eoseviulisoma Brolemann, 1920, is synonymized under Eviulisoma, based on newly collected material of E. julinum (Attems, 1909), type species of Eoseviulisoma. New material of Suohelisoma ulugurense Hoffman, 1964, type species of Suohelisoma Hoffman, 1964, has revealed that the gonopod structure is more similar to that of Eviulisoma than originally thought, but Suohelisoma is retained as a valid genus. Four species groups are recognized among Eviulisoma species from the Udzungwa Mts, but the need for a revision of the entire genus is emphasized. Two types of epizootic fungi are recorded from Eviulisoma spp., and an enigmatic amorphous mass, which may be a kind of plugging substance, is recorded from the gonopod tips and excavated sixth sternum of several species.

Keywords. Taxonomy, new species, epizootic fungi, copulatory plug.

Enghoff H. 2018. A mountain of millipedes VII: The genus Eviulisoma Silvestri, 1910, in the Udzungwa Mountains Tanzania, and related species from other Eastern Arc Mountains. With notes on Eoseviulisoma Brolemann, 1920 and Suohelisoma Hoffman, 1963 (Diplopoda, Polydesmida, Paradoxosomatidae). European Journal of Taxonomy 445: 1–90. https://doi.org/10.5852/ejt.2018.445

zenodo

Englight H. 2018. A mountain of millipedes VII: The genus Eviulisoma Silvestri, 1910, in the Udzungwa Mountains, Tanzania, and related species from other Eastern Arc Mountains. With notes on Eastern Arc Mountains. With notes on Eastern Arc Mountains and Paradoxosomatidae. European Journal of Taxonomy 445: 1–90. https://doi.org/10.5852/ejt.2018.445

# Treatments: Data in disguise

Eviulisoma breviscutum sp. nov.

urn:lsid:zoobank.org:act:D7C4195B-37DF-4B02-BD3B-4447DBCBB23C Fig. 36

#### **Diagnosis**

Differs from other Udzungwan species of *Eviulisoma* by the combination of unmodified sterna 5 and 6 and a very short *map* (ca half as long as solenophore).

#### **Etymology**

The name is a noun in apposition meaning 'short shield' and refers to the short, shield-like mesal acropodital process.

**Material** (total: 3 ♂♂)

#### Holotype

TANZANIA:  $\circlearrowleft$ , Mwanihana Forest, above Sanje, 1650 m a.s.l., pitfall trap, 18 Aug. 1982, M. Stoltze and N. Scharff leg. (ZMUC).

#### **Paratypes**

TANZANIA: 1 Å, Morogoro Region, Kilombero District, Udzungwa Mts National Park, forest below Mwanihana Peak, 7°49′ S, 36°50′ E, 1800 m a.s.l., sifted from leaf litter, 20 Aug. 2017, T. Pape leg. (ZMUC); 1 Å, Morogoro Region, Udzungwa Mts National Park, Mito Mitatu, above Mang'ula,



# Treatments: Data in disguise

**Geographic coordinates** 

**Date of collection** 

**Collector** 

**Material** (total: 3 ♂♂)

#### Holotype

TANZANIA: 3, Mwanihana Forest, above Sanje, 1650 m a.s.l., pitfall trap, 18 Aug. 1982, M. Stoltze and N. Scharff leg. (ZMUC).

#### **Paratypes**

TANZANIA: 1 Å, Morogord Region, Kilombero District, Udzungwa Mts National Park, forest below Mwanihana Peak, 7°49′ S, 36°50′ E, 1800 m a.s.l., sifted from leaf litter, 20 Aug. 2017, T. Pape leg. (ZMUC); 1 Å, Morogoro Region, Udzungwa Mts National Park, Mito Mitatu, above Mang'ula,

07°49′3″ S, 36°52′58″ E, 1487 m a.s.l., 16 Dec. 2016,



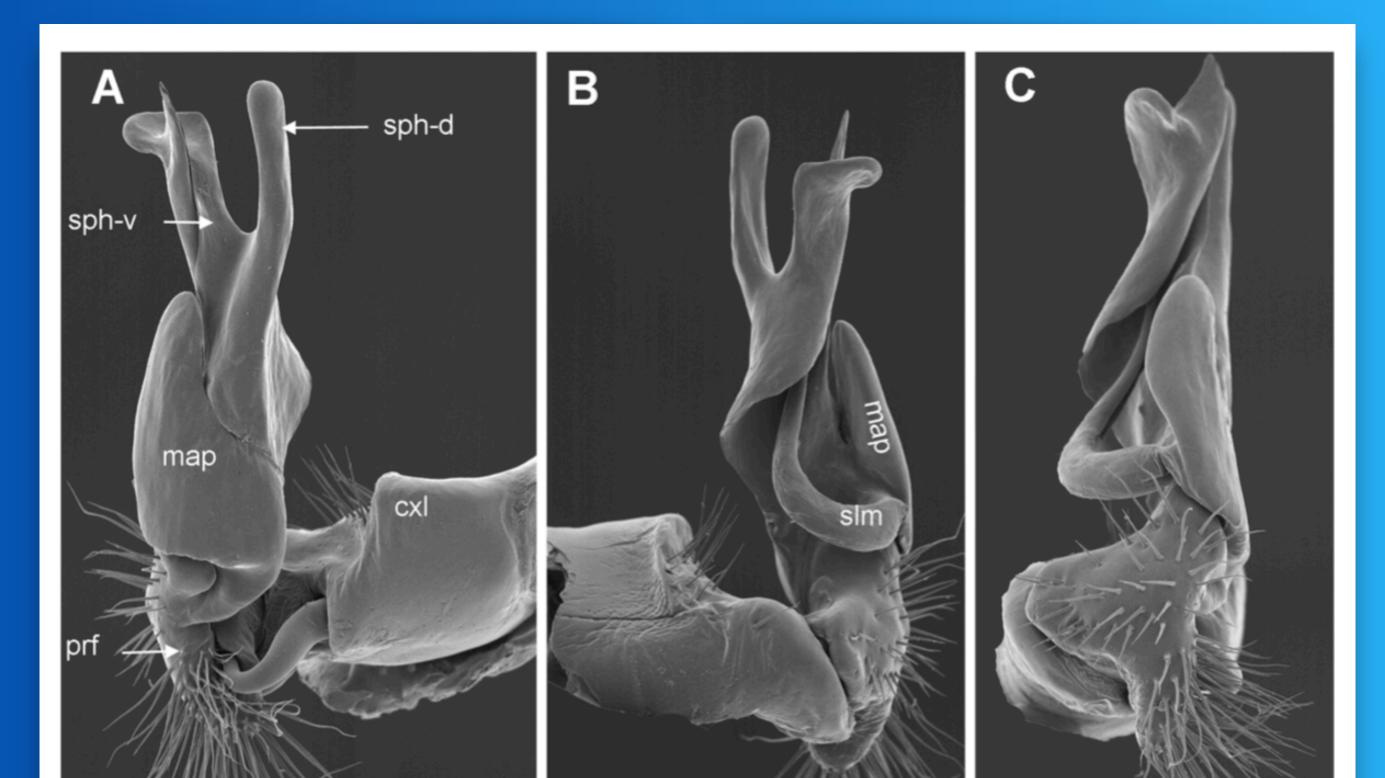
EUROPEAN JOURNAL OF TAXONOMY
MATERIAL CITATIONS FORMATTING GUIDE

**Host collection** 



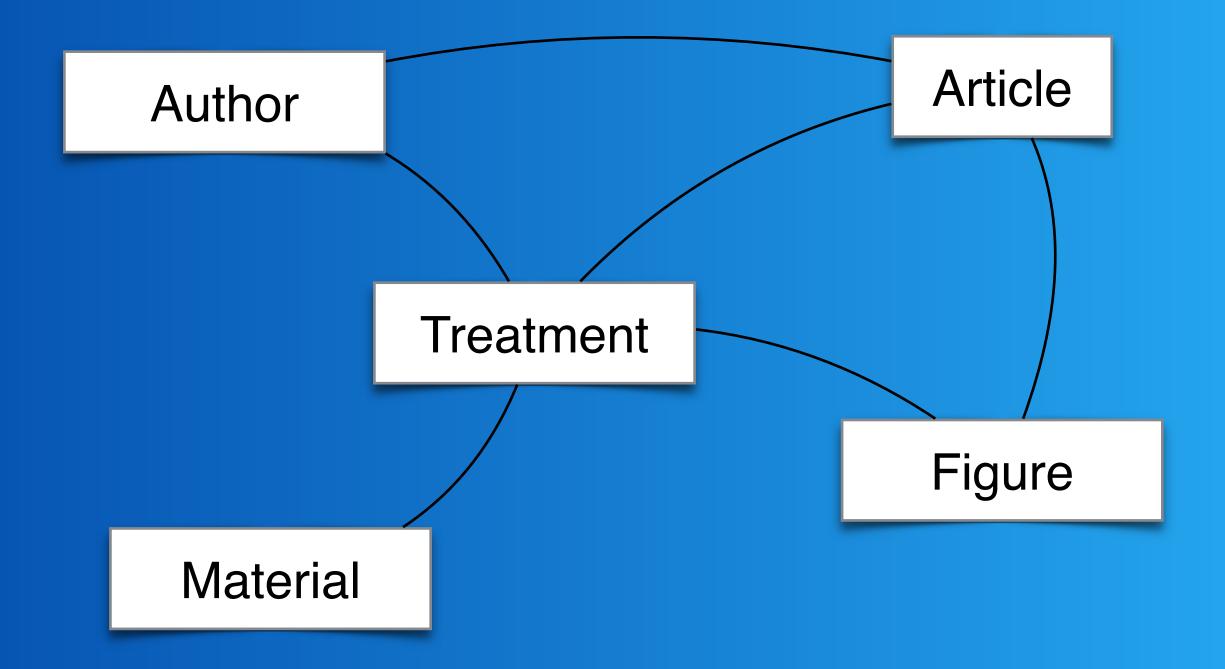
In accordance with the European Journal of Taxonomy's <u>FAIR & Open Science policy</u>, the formatting guide for entomology, zoology and palaeozoology material citations is provided below (guidelines for botany available soon). Authors are encouraged to prepare their manuscripts according to this model prior to submission, but they will also be given the opportunity to comply upon acceptance of the article.

# Treatments: Figures





# Treatments: Relations





# Locked up data

## Unanswerable questions:

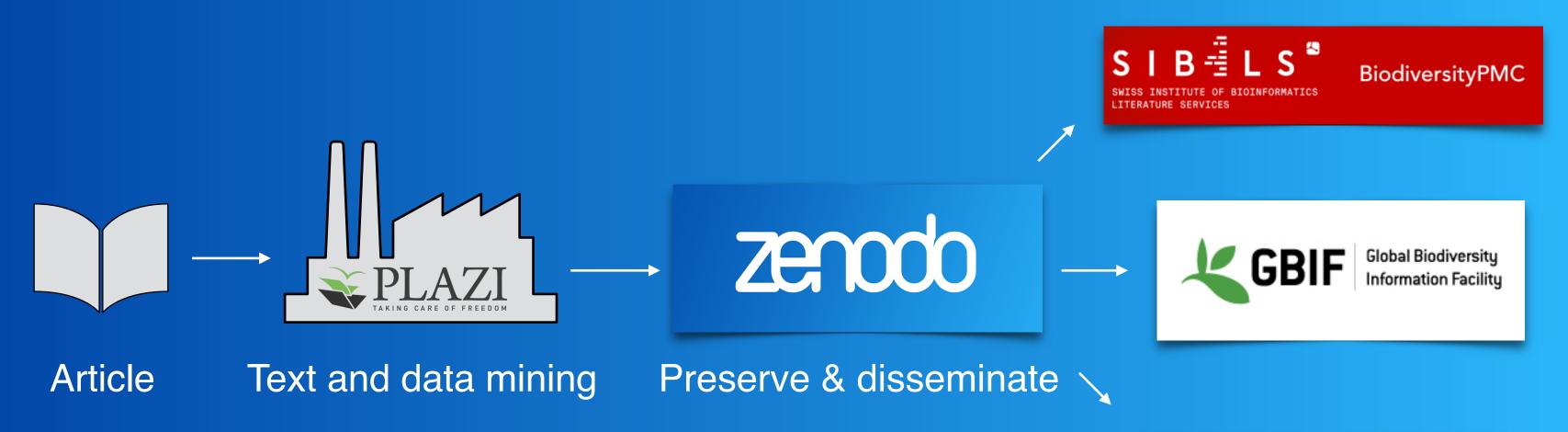
- How many species have been described by my collection?
- Give me a list of all new species?
- Retrieve all images for a given taxon?
- What's known about a geographic region?

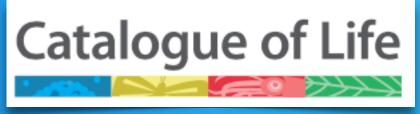
### • Treatments:

- Past 260 years: ~10+ millions published
- •Every year: ~17k new / ~130k augmented



# It takes an ecosystem

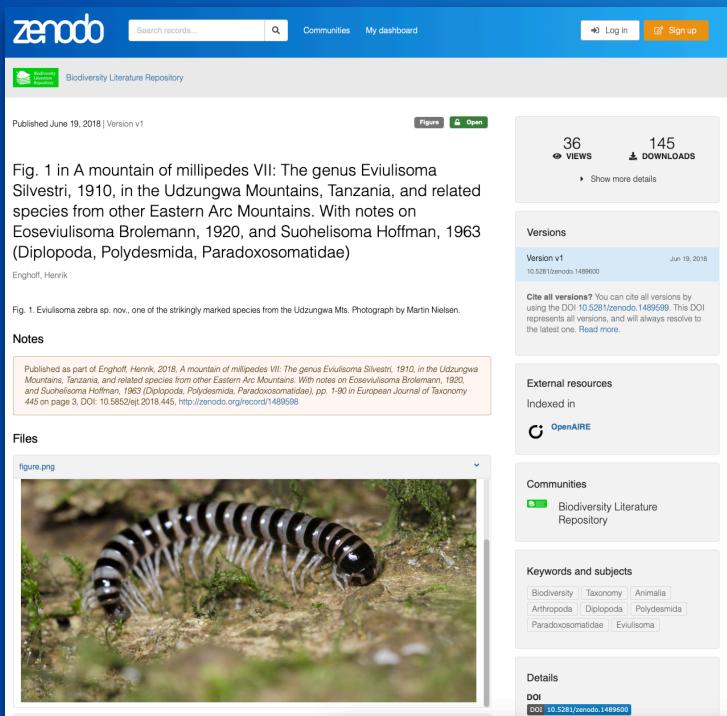


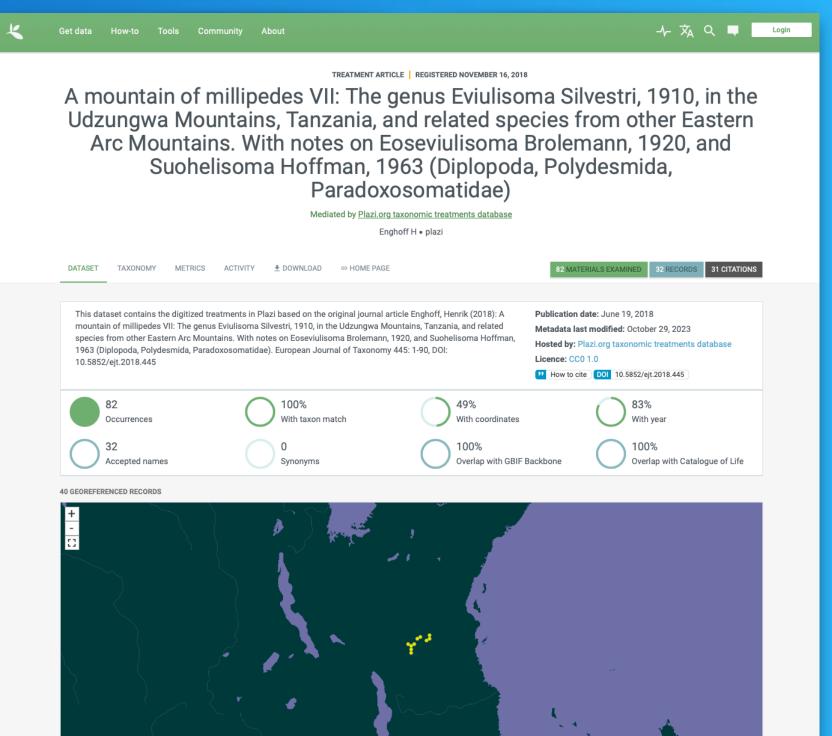


Search, retrieve & linking



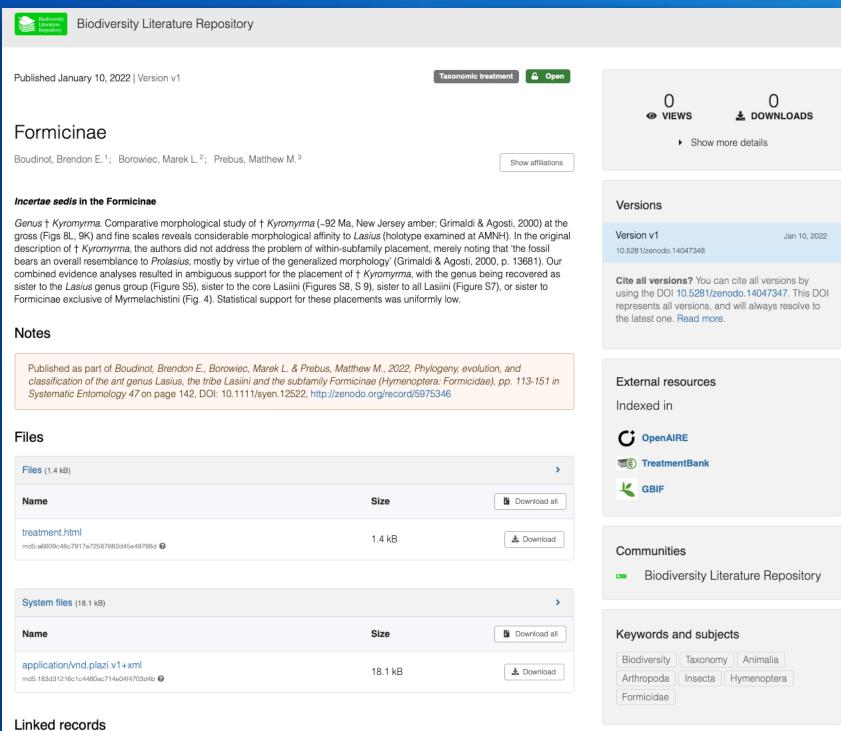
# FAIR Data

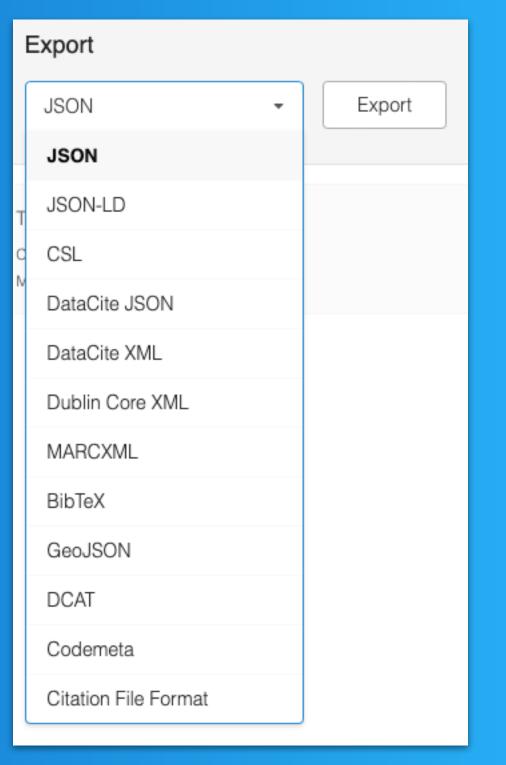






# Interdisciplinary metadata





# Domain-specific metadata

#### Related works

Figure: 10.5281/zenodo.5975370 (DOI) Figure: 10.5281/zenodo.5975374 (DOI)

Figure: 10.5281/zenodo.5975362 (DOI) Figure: 10.5281/zenodo.5975368 (DOI) Figure: 10.5281/zenodo.5975358 (DOI)

#### Is part of

Journal article: 10.1111/syen.12522 (DOI)

Journal article: http://zenodo.org/record/5975346 (URL)

Journal article: http://publication.plazi.org/id/FF80FFD14372FF82FF5FF95FF99D837 (URL)
Journal article: http://zoobank.org/016059BA-33C3-43B2-ADAD-6807DC5CB6D8 (URL)

#### Is source of

https://biodiversitypmc.sibils.org/collections/plazi/03B987A9436FFF9FFCA4FC53FAEDDDA9 (URL)

https://www.gbif.org/species/245722473 (URL)

https://www.checklistbank.org/dataset/20773/taxon/03B987A9436FFF9FFCA4FC53FAEDDDA9.taxon (URL)

#### Biodiversity

#### Family 🗹

Formicidae

#### Kingdom 🗹

Animalia

#### Order 🗹

Hymenoptera

#### Phylum 🗹

Arthropoda

#### Scientific name authorship [2]

Lepeletier de Saint-Fargeau

Taxon rank <a>I</a>

subFamily

#### Research object metadata

Indexed and searchable metadata

Interdisciplinary layer (DataCite terms)

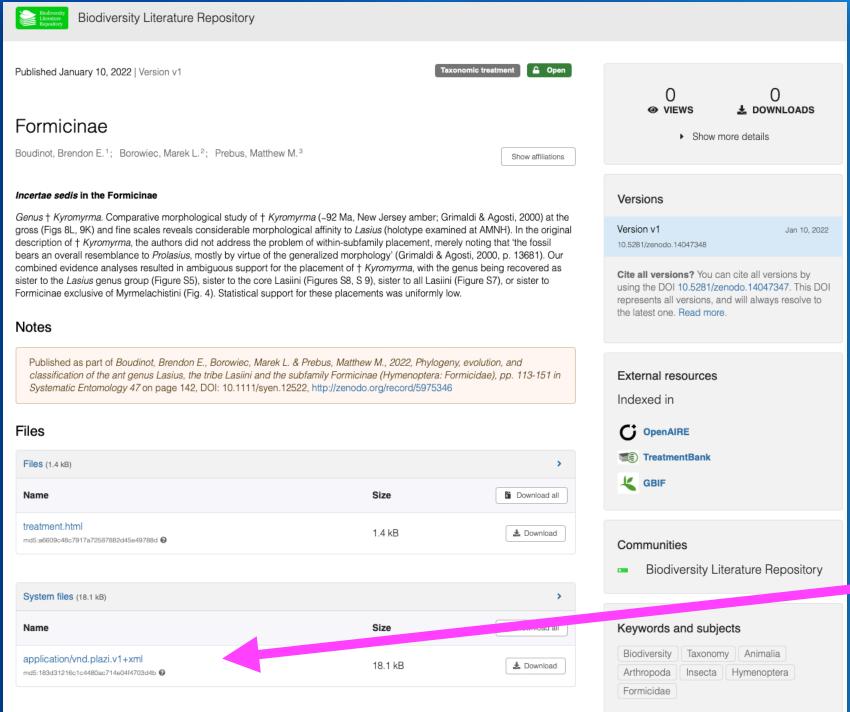
Domain layer MeSH vocabulary Domain layer Darwin Core terms

Discipline metadata file

Discipline metadata file Discipline metadata file



# Domain-specific metadata



#### Research object metadata

Indexed and searchable metadata

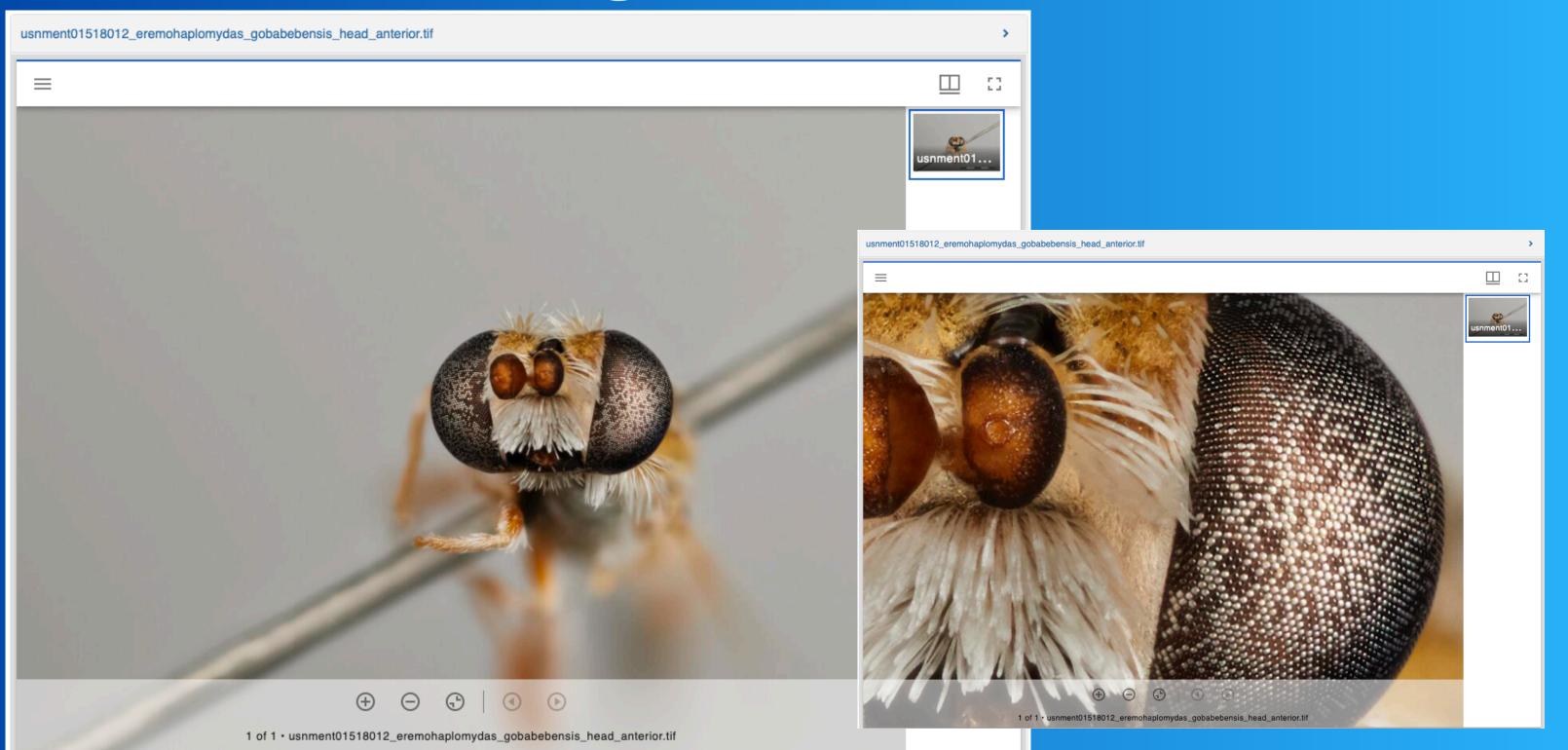
Interdisciplinary layer (DataCite terms)

Domain layer MeSH vocabulary Domain layer Darwin Core terms

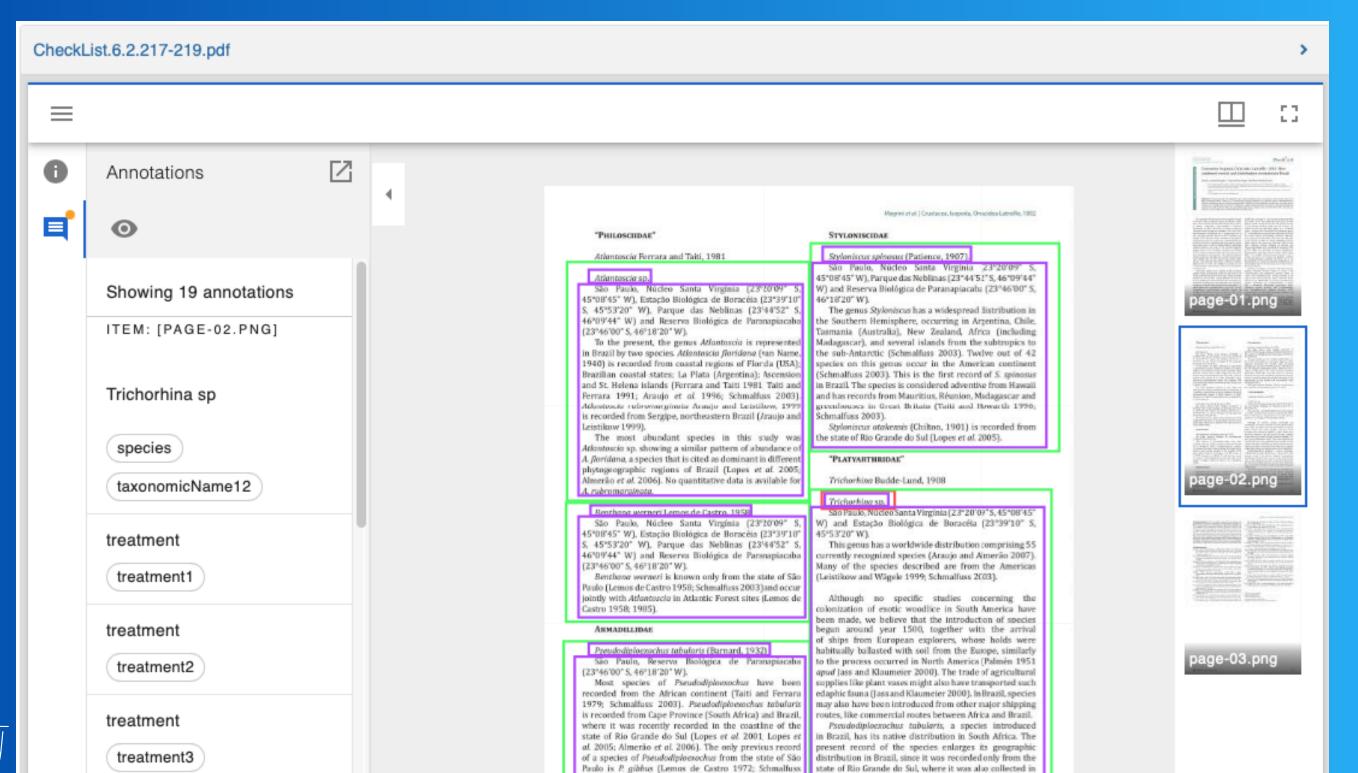
Discipline metadata file

Discipline metadata file Discipline metadata file

# Image Zoom: IIIF



# Annotations: WADM



the litter layer of Atlantic forests. Therefore, we expect that



## How do machines access?

# FAIR Signposting

```
lnielsen@lnielsen-mbp16-10 ~ % curl -I -X HEAD https://zenodo.org/records/13325981
HTTP/1.1 200 0K
server: nginx
date: Wed, 06 Nov 2024 22:12:57 GMT
content-type: text/html; charset=utf-8
content-length: 91348
vary: Accept-Encoding
link: <https://zenodo.org/api/records/13325981> ; rel="linkset"; type="application/linkset+json"
```



# FAIR Signposting

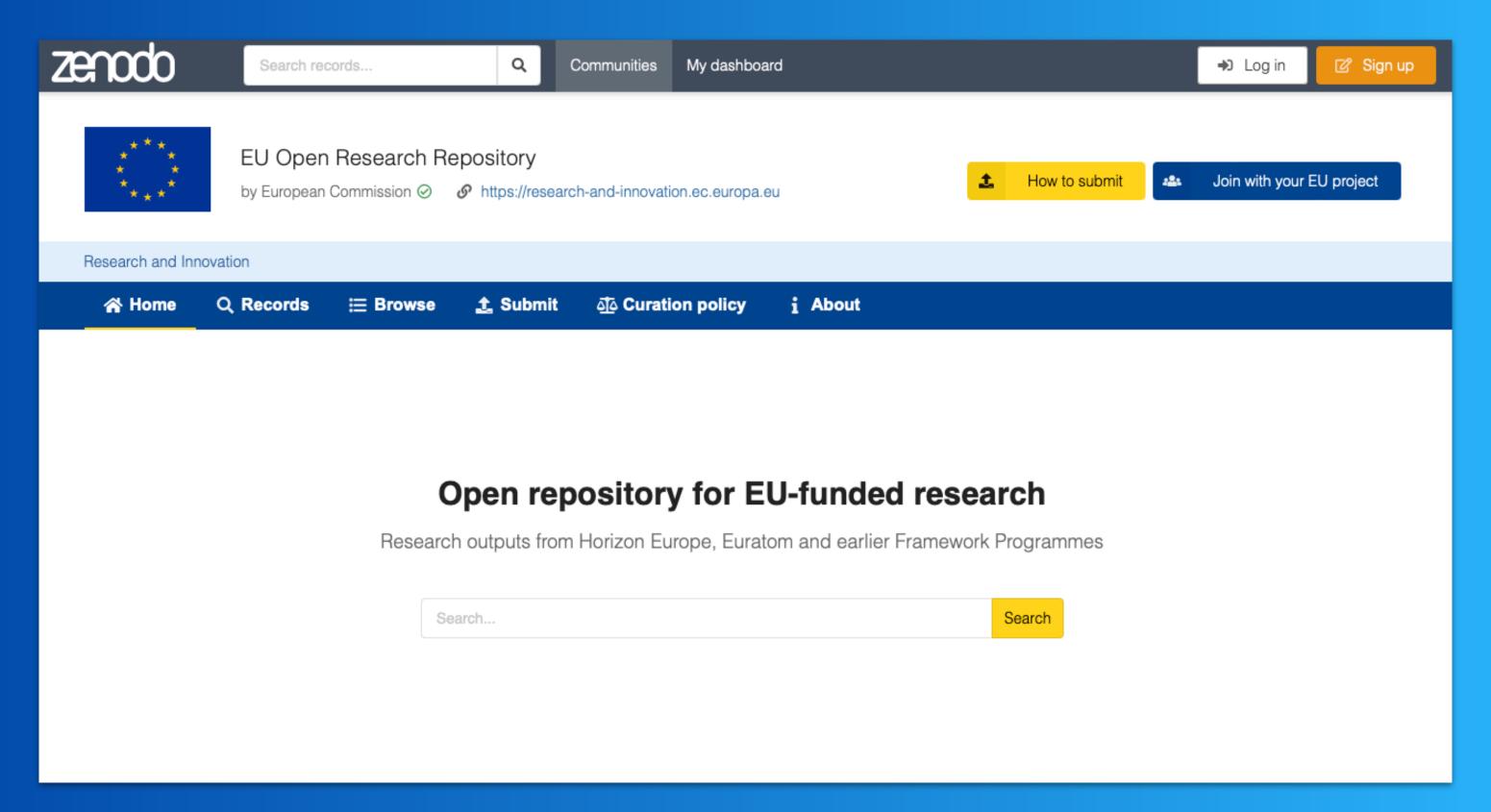
```
lnielsen@lnielsen-mbp16-10 ~ % curl -X GET -H "Accept: application/linkset+json" https://zenodo.org/api/records/13325981
            % Received % Xferd Average Speed
 % Total
                                                        Time
                                                                 Time Current
                                Dload Upload
                                                Total
                                                        Spent
                                                                 Left Speed
    2061 100 2061
                             0 8243
 "linkset": [
      "anchor": "https://zenodo.org/records/13325981",
     "cite-as": [
         "href": "https://doi.org/10.5281/zenodo.13325981"
      "describedby": [
         "href": "https://zenodo.org/api/records/13325981",
         "type": "application/dcat+xml"
         "href": "https://zenodo.org/api/records/13325981",
         "type": "application/json"
```



# Embedded JSON-LD

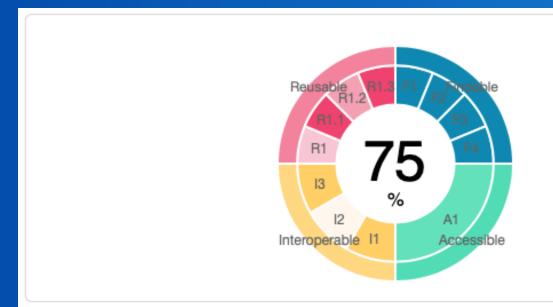


## What about researchers?

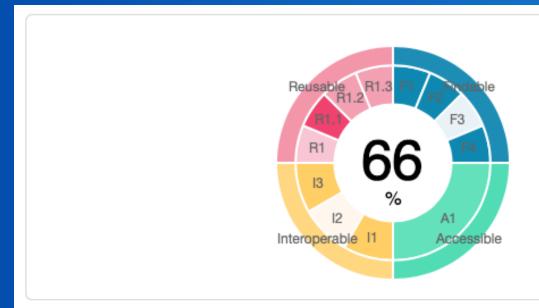




## Can we integrate FAIR evaluation tools?



	Score earned:		Fair level:
Findable:	7 of 7	0	advanced
Accessible:	2 of 3	0	moderate
Interoperable:	3 of 4	0	moderate
Reusable:	6 of 10	0	moderate



	Score earned:		Fair level:
Findable:	6 of 7	Ö	moderate
Accessible:	2 of 3	0	moderate
Interoperable:	3 of 4	0	moderate
Reusable:	5 of 10	0	moderate



## Challenges





GET https://zenodo.org/records/1234

200 OK

Success



## Challenges: Unpublished







**403 FORBIDDEN** 

Fail

GET https://zenodo.org/uploads/1234?token=...

200 OK

Success



## Challenges: DOI not yet registered





GET https://doi.org/10.5281/zenodo.1234

**404 NOT FOUND** 

Fail

=> Lower FAIR score outside of control of user



## Failed tests due to unregistered DOI

- FsF-F1-02D (fail): Data is assigned a persistent identifier.
- FsF-F4-01M (lower score): Metadata is registered in major research data registries (DataCite).
- FsF-I1-01M (fail): Parsable, graph data (RDF, JSON-LD) is accessible through content negotiation, typed links or sparql endpoint.

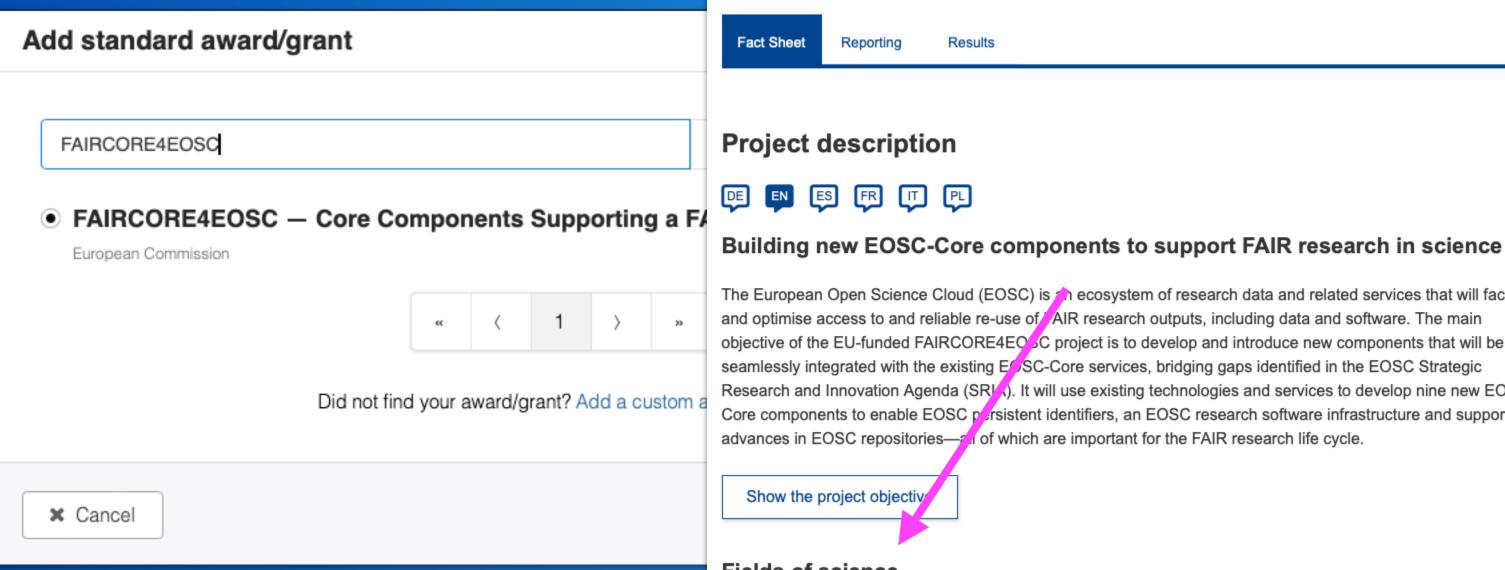


## What then?

Metadata: Automated checks + Subject information

Data curation: File format checks

## Subject information





#### **Core Components Supporting a FAIR EOSC**

**Fact Sheet** 

Reporting

Results

#### **Project description**













The European Open Science Cloud (EOSC) is necosystem of research data and related services that will facilitate and optimise access to and reliable re-use of AIR research outputs, including data and software. The main objective of the EU-funded FAIRCORE4EQC project is to develop and introduce new components that will be seamlessly integrated with the existing EOSC-Core services, bridging gaps identified in the EOSC Strategic Research and Innovation Agenda (SRK). It will use existing technologies and services to develop nine new EOSC-Core components to enable EOSC persistent identifiers, an EOSC research software infrastructure and support for advances in EOSC repositories—a of which are important for the FAIR research life cycle.

Show the project objective

#### Fields of science

natural sciences > computer and information sciences > software



## Subject information

Subject

Grant

Record

#### **Subjects**



#### Social **Sciences**

42,392

Natural

64,051

#### Engineering and

Chemical Engineering (672)

Civil Engineering (3,820)

Engineering (18,958)

(3,762)

40,670

#### technology

Electrical, Electronic and Information

Environmental Biotechnology (685)

Environmental Engineering (15,335)

Other Engineering And Technologies

Industrial Biotechnology (1,633)

Materials Engineering (5,730) Mechanical Engineering (8,483) Medical Engineering (729) Nanotechnology (3,176)

Arts (6,424)

History And Archaeology (9,792) Languages And Literature (1,897)

Humanities

13,946

Philosophy, Ethics And Religion (956)

Other Humanities (4,666)

Economics and business (21,861) Educational sciences (4,819) Law (2,562) Media and communications (696) Political sciences (12,602) Psychology (1,136) Social geography (2,125) Sociology (21,395)

Other social sciences (1)

#### **Sciences**

Earth And Related Environmental Sciences (14,220)

Mathematics (3,243)

Biological Sciences (25,085)

Chemical Sciences (6,763)

Physical Sciences (16,030)

Computer And Information Sciences

(35.886)

Other Natural Sciences (0)



**Agricultural** 

sciences

14,807

#### Medical and Health sciences

14,506

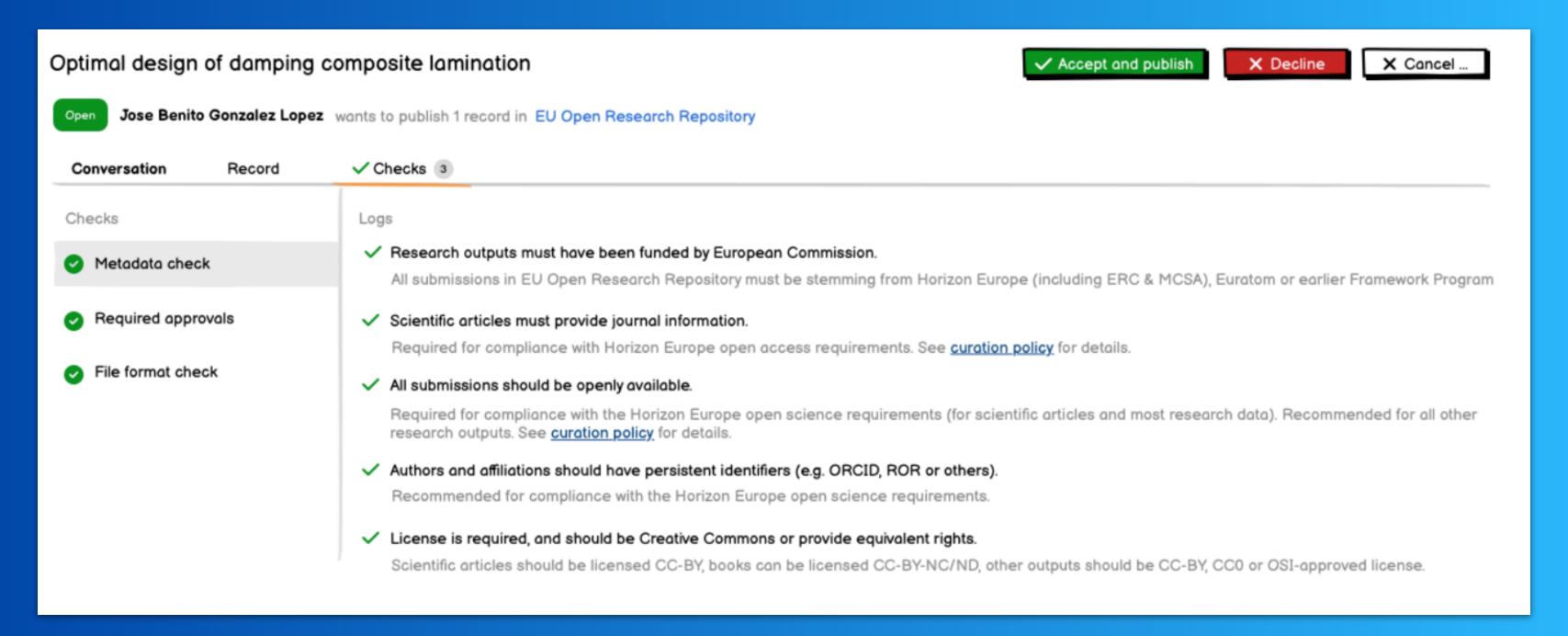
Agriculture, Forestry, And Fisheries (12,351)Animal And Dairy Science (1,974)

Veterinary Sciences (70) Other Agricultural Sciences (0) Basic Medicine (5,990) Clinical Medicine (4,621) Health Sciences (8,030) Medical Biotechnology (1,763) Other Medical Sciences (156)



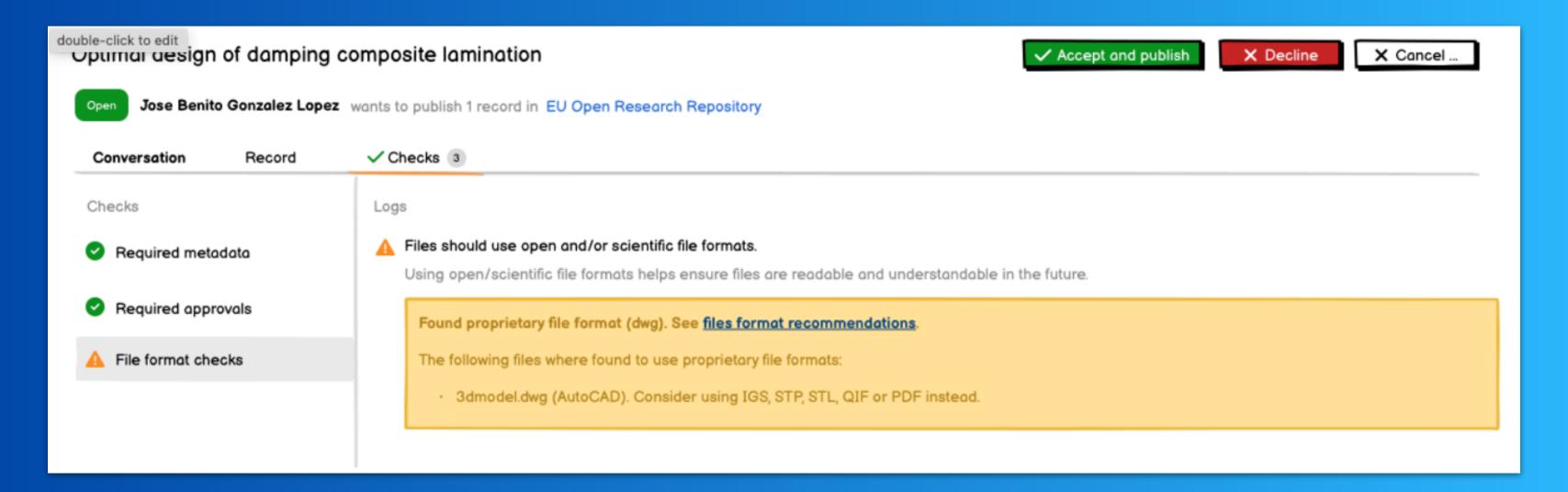


## Metadata and data curation





## Metadata and data curation





The current incompatibilities of the platforms and tools make it impossible to access existing information through a common interface, leading to waste of time, frustration and obsolete answers to simple data lookup.

#### WorldWideWeb:

### Proposal for a HyperText Project

T. Berners-Lee / CN, R. Cailliau / ECP

Abstract: HyperText is a way to link and access information of various kinds as a web of nodes in which the user can browse at will. Potentially, HyperText provides a single user-interface to many large classes of stored information such as reports, notes, data-bases, computer documentation and on-line systems help. We propose the implementation of a simple scheme to incorporate several different servers of machine-stored information already available at CERN, including an analysis of the requirements for information access needs by experiments.

#### Introduction

The current incompatibilities of the platforms and tools make it impossible to access existing information through a common interface, leading to waste of time, frustration and obsolete answers to simple data lookup. There is a potential large benefit from the integration of a variety of systems in a way which allows a user to follow links pointing from one piece of information to another one. This forming of a web of information nodes rather than a hierarchical tree or an ordered list is the basic concept behind HyperText.

