EU OPEN DATA DAYS

EU Open Research Repository

Lars Holm Nielsen

CERN, Head of Open Science Infrastructure

19 – 20 MARCH 2025

This work is licensed under Creative Commons Attribution 4.0 International, except where noted on each slide.

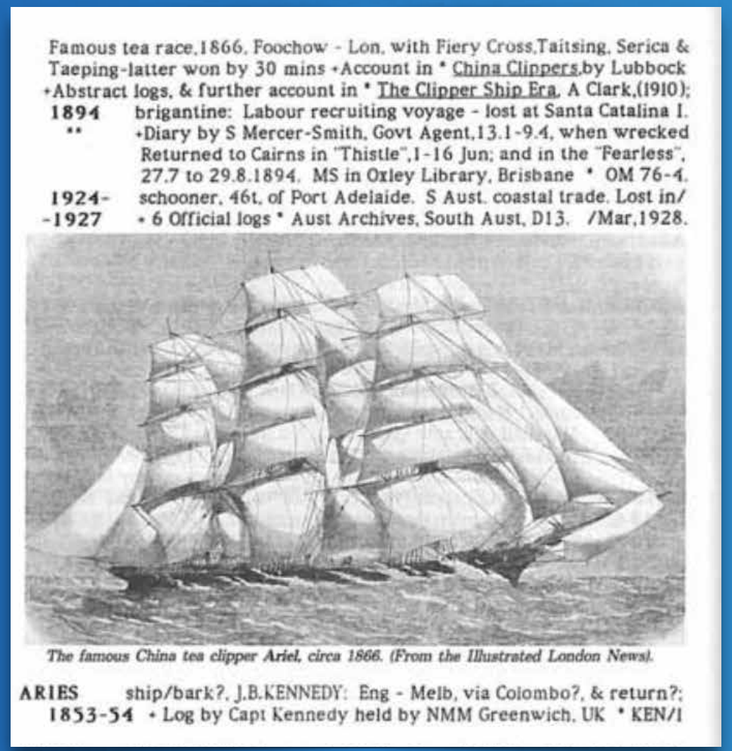
Funded by the European Union under Horizon Europe grant agreement number 101122956.

€100 billion

EU-27 yearly spending on academic R&D

Source: OECD, Main Science and Technology Indicators (May 2023)

Reuse?



Famous tea race,1866, Foochow - Lon, with Fiery Cross,Taitsing, Serica & Taeping-latter won by 30 mins +Account in * China Clippers,by Lubbock +Abstract logs, & further account in * The Clipper Ship Era, A Clark,(1910);

1894 brigantine: Labour recruiting voyage - lost at Santa Catalina I.
** +Diary by S Mercer-Smith, Govt Agent,13.1-9.4, when wrecked Returned to Cairns in "Thistle",1-16 Jun; and in the "Fearless", 27.7 to 29.8.1894. MS in Oxley Library, Brisbane * OM 76-4.

1924--1927 schooner, 46t, of Port Adelaide. S Aust. coastal trade. Lost in/ + 6 Official logs * Aust Archives, South Aust, D13. /Mar,1928.

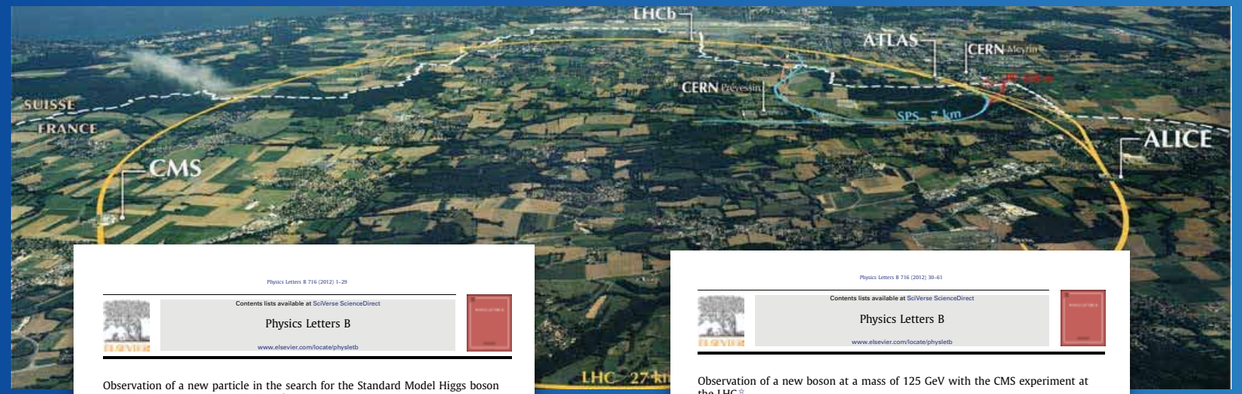
*The famous China tea clipper **Ariel**, circa 1866. (From the Illustrated London News).*

ARIES ship/bark?, J.B.KENNEDY: Eng - Melb, via Colombo?, & return?;
1853-54 + Log by Capt Kennedy held by NMM Greenwich, UK * KEN/I

Ships logs from 18th century used for climate research

Source: University of Sunderland. "18th Century Ships' Logs Predict Future Weather Forecast." ScienceDaily. ScienceDaily, 6 October 2009. <www.sciencedaily.com/releases/2009/10/091006104627.htm>

Reproducibility?



Physics Letters B 716 (2012) 1–29

Contents lists available at SciVerse ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC ☆

ATLAS Collaboration *

This paper is dedicated to the memory of our ATLAS colleagues who did not live to see the full impact and significance of their contributions to the experiment.

ARTICLE INFO

Article history:
Received 31 July 2012
Received in revised form 8 August 2012
Accepted 11 August 2012
Available online 14 August 2012
Editor: W.-D. Schlatter

ABSTRACT

A search for the Standard Model Higgs boson in proton–proton collisions with the ATLAS detector at the LHC is presented. The datasets used correspond to integrated luminosities of approximately 4.8 fb^{-1} collected at $\sqrt{s} = 7$ TeV in 2011 and 5.8 fb^{-1} at $\sqrt{s} = 8$ TeV in 2012. Individual searches in the channels $H \to ZZ^{(*)} \to 4\ell$, $H \to \gamma\gamma$ and $H \to WW^{(*)} \to e\nu\mu\nu$ in the 8 TeV data are combined with previously published results of searches for $H \to ZZ^{(*)}$, $WW^{(*)}$, $b\bar{b}$ and $\tau^+\tau^-$ in the 7 TeV data and results from improved analyses of the $H \to ZZ^{(*)} \to 4\ell$ and $H \to \gamma\gamma$ channels in the 7 TeV data. Clear evidence for the production of a neutral boson with a measured mass of 126.0 ± 0.4 (stat) ± 0.4 (sys) GeV is presented. This observation, which has a significance of 5.9 standard deviations, corresponding to a background fluctuation probability of 1.7×10^{-9}, is compatible with the production and decay of the Standard Model Higgs boson.

© 2012 CERN. Published by Elsevier B.V. All rights reserved.

1. Introduction

The Standard Model (SM) of particle physics [1–4] has been tested by many experiments over the last four decades and has been shown to successfully describe high energy particle interactions. However, the mechanism that breaks electroweak symmetry in the SM has not been verified experimentally. This mechanism [5–10], which gives mass to massive elementary particles, implies the existence of a scalar particle, the SM Higgs boson. The search for the Higgs boson, the only elementary particle in the SM that has not yet been observed, is one of the highlights of the Large Hadron Collider [11] (LHC) physics programme.

Indirect limits on the SM Higgs boson mass of $m_H < 158$ GeV at 95% confidence level (CL) have been set using global fits to precision electroweak results [12]. Direct searches at LEP [13], the Tevatron [14–16] and the LHC [17,18] have previously excluded, at 95% CL, a SM Higgs boson with mass below 600 GeV, apart from some mass regions between 116 GeV and 127 GeV.

Both the ATLAS and CMS Collaborations reported excesses of events in their 2011 datasets of proton–proton (pp) collisions at centre-of-mass energy $\sqrt{s} = 7$ TeV at the LHC, which were compatible with SM Higgs boson production and decay in the mass region 124–126 GeV, with significances of 2.9 and 3.1 standard deviations (σ), respectively [17,18]. The CDF and DØ experiments at the Tevatron have also recently reported a broad excess in the mass region 120–135 GeV, using the existing LHC constraints, the observed local significances for $m_H = 125$ GeV are 2.7σ for CDF [14], 1.1σ for DØ [15] and 2.2σ for their combination [16].

The previous ATLAS searches in 4.6–4.8 fb^{-1} of data at $\sqrt{s} = 7$ TeV are combined here with new searches for $H \to ZZ^{(*)} \to 4\ell$, $H \to \gamma\gamma$ and $H \to WW^{(*)} \to e\nu\mu\nu$ in the 5.8–5.9 fb^{-1} of pp collision data taken at $\sqrt{s} = 8$ TeV between April and June 2012.

The data were recorded with instantaneous luminosities up to 6.8×10^{33} cm^{-2} s^{-1}; they are therefore affected by multiple pp collisions occurring in the same or neighbouring bunch crossings (pile-up). In the 7 TeV data, the average number of interactions per bunch crossing was approximately 10; the average increased to approximately 20 in the 8 TeV data. The reconstruction, identification and isolation criteria used for electrons and photons in the 8 TeV data are improved, making the $H \to ZZ^{(*)} \to 4\ell$ and $H \to \gamma\gamma$ searches more robust against the increased pile-up. These analyses were re-optimised with simulation and frozen before looking at the 8 TeV data.

In the $H \to WW^{(*)} \to \ell\nu\ell\nu$ channel, the increased pile-up deteriorates the event missing transverse momentum, E_T^{miss}, resolution, which results in significantly larger Drell–Yan background in the same-flavour final states. Since the $e\mu$ channel provides most of the sensitivity of the search, only this final state is used in the analysis of the 8 TeV data. The kinematic region in which a SM Higgs boson with a mass between 110 GeV and 140 GeV is

☆ © CERN for the benefit of the ATLAS Collaboration.

* E-mail address: atlas.publications@cern.ch.

0370-2693/ © 2012 CERN. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physletb.2012.08.020

[1] The symbol ℓ stands for electron or muon.

Physics Letters B 716 (2012) 30–61

Contents lists available at SciVerse ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC ☆

CMS Collaboration *

CERN, Switzerland

This paper is dedicated to the memory of our colleagues who worked on CMS but have since passed away. In recognition of their many contributions to the achievement of this observation.

ARTICLE INFO

Article history:
Received 31 July 2012
Received in revised form 9 August 2012
Accepted 11 August 2012
Available online 18 August 2012
Editor: W.-D. Schlatter

Keywords:
CMS
Physics
Higgs

ABSTRACT

Results are presented from searches for the standard model Higgs boson in proton–proton collisions at $\sqrt{s} = 7$ and 8 TeV in the Compact Muon Solenoid experiment at the LHC, using data samples corresponding to integrated luminosities of up to 5.1 fb^{-1} at 7 TeV and 5.3 fb^{-1} at 8 TeV. The search is performed in five decay modes: $\gamma\gamma$, ZZ, WW, $\tau^+\tau^-$, and $b\bar{b}$. An excess of events is observed above the expected background, with a local significance of 5.0 standard deviations, at a mass near 125 GeV, signalling the production of a new particle. The expected significance for a standard model Higgs boson of that mass is 5.8 standard deviations. The excess is most significant in the two decay modes with the best mass resolution, $\gamma\gamma$ and ZZ; a fit to these signals gives a mass of 125.3 ± 0.4(stat.) ± 0.5(syst.) GeV. The decay to two photons indicates that the new particle is a boson with spin different from one.

© 2012 CERN. Published by Elsevier B.V. All rights reserved.

1. Introduction

The standard model (SM) of elementary particles provides a remarkably accurate description of results from many accelerator and non-accelerator based experiments. The SM comprises quarks and leptons as the building blocks of matter, and describes their interactions through the exchange of force carriers: the photon for electromagnetic interactions, the W and Z bosons for weak interactions, and the gluons for strong interactions. The electromagnetic and weak interactions are unified in the electroweak theory. Although the predictions of the SM have been extensively confirmed, the question of how the W and Z gauge bosons acquire mass whilst the photon remains massless is still open.

Nearly fifty years ago it was proposed [1–6] that spontaneous symmetry breaking in gauge theories could be achieved through the introduction of a scalar field. Applying this mechanism to the electroweak theory [7–9] through a complex scalar doublet field leads to the generation of the W and Z masses, and to the prediction of the existence of the SM Higgs boson (H). The scalar field also gives mass to the fundamental fermions through the Yukawa interaction. The mass m_H of the SM Higgs boson is not predicted by theory. However, general considerations [10–13] suggest that m_H should be smaller than ~1 TeV, while precision electroweak measurements imply that $m_H < 152$ GeV at 95% confidence level (CL) [14]. Over the past twenty years, direct searches for the Higgs boson have been carried out at the LEP collider, leading to a lower bound of $m_H > 114.4$ GeV at 95% CL [15], and at the Tevatron proton–antiproton collider, excluding the mass range 162–166 GeV at 95% CL [16] and detecting an excess of events, recently reported in [17–19], in the range 120–135 GeV.

The discovery or exclusion of the SM Higgs boson is one of the primary scientific goals of the Large Hadron Collider (LHC) [20]. Previous direct searches at the LHC were based on data from proton–proton collisions corresponding to an integrated luminosity of 5 fb^{-1} collected at a centre-of-mass energy $\sqrt{s} = 7$ TeV. The CMS experiment excluded at 95% CL a range of masses from 127 to 600 GeV [21], The ATLAS experiment excluded at 95% CL the ranges 111.4–116.8, 119.4–122.1 and 129.2–541 GeV [22]. Within the remaining allowed mass region, an excess of events near 125 GeV was reported by both experiments. In 2012 the proton–proton centre-of-mass energy was increased to 8 TeV and by the end of June an additional integrated luminosity of more than 5 fb^{-1} had been recorded by each of these experiments, thereby enhancing significantly the sensitivity of the search for the Higgs boson.

This Letter reports the results of a search for the SM Higgs boson using samples collected by the CMS experiment, comprising data recorded at $\sqrt{s} = 7$ and 8 TeV. The search is performed in

☆ © CERN for the benefit of the CMS Collaboration.

* E-mail address: cms-publication-committee-chair@cern.ch.

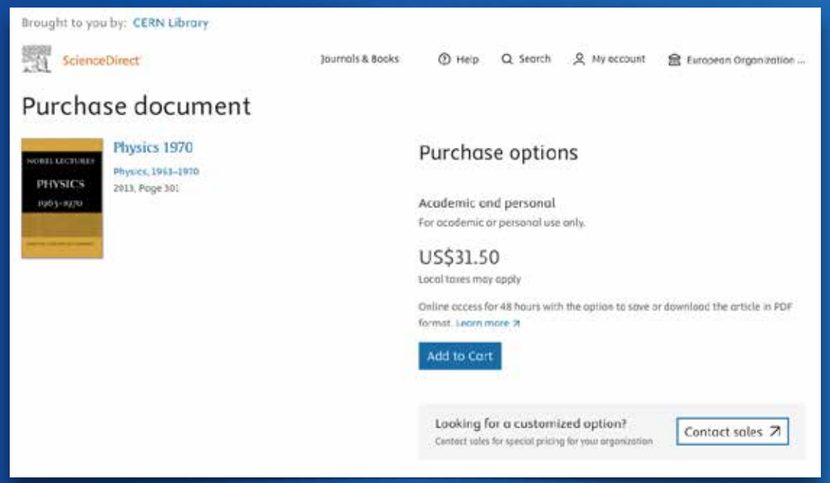
0370-2693/ © 2012 CERN. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physletb.2012.08.021

Only 6 of 53 landmark papers in cancer research were able to be confirmed

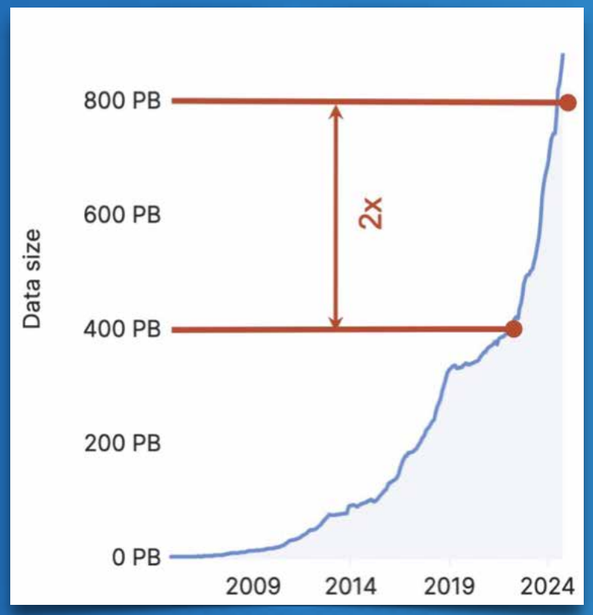
Begley, C., Ellis, L. Raise standards for preclinical cancer research. Nature 483, 531–533 (2012). https://doi.org/10.1038/483531a

Access?

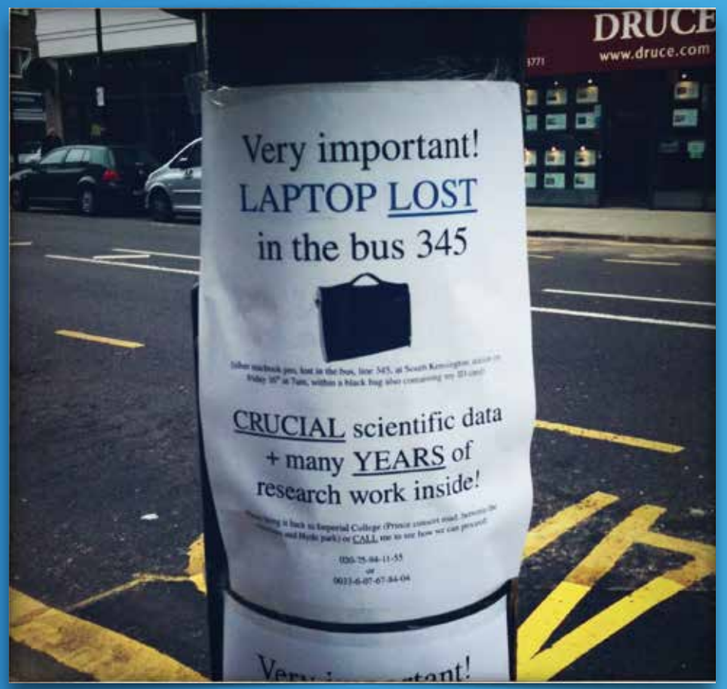
Paywalls



Data storage/growth



Disasters



EU Open Science Policy

Accessible

Transparent

Reusable

I.e. squeeze more value out of 100 billions EUR/year

Horizon Europe Grants

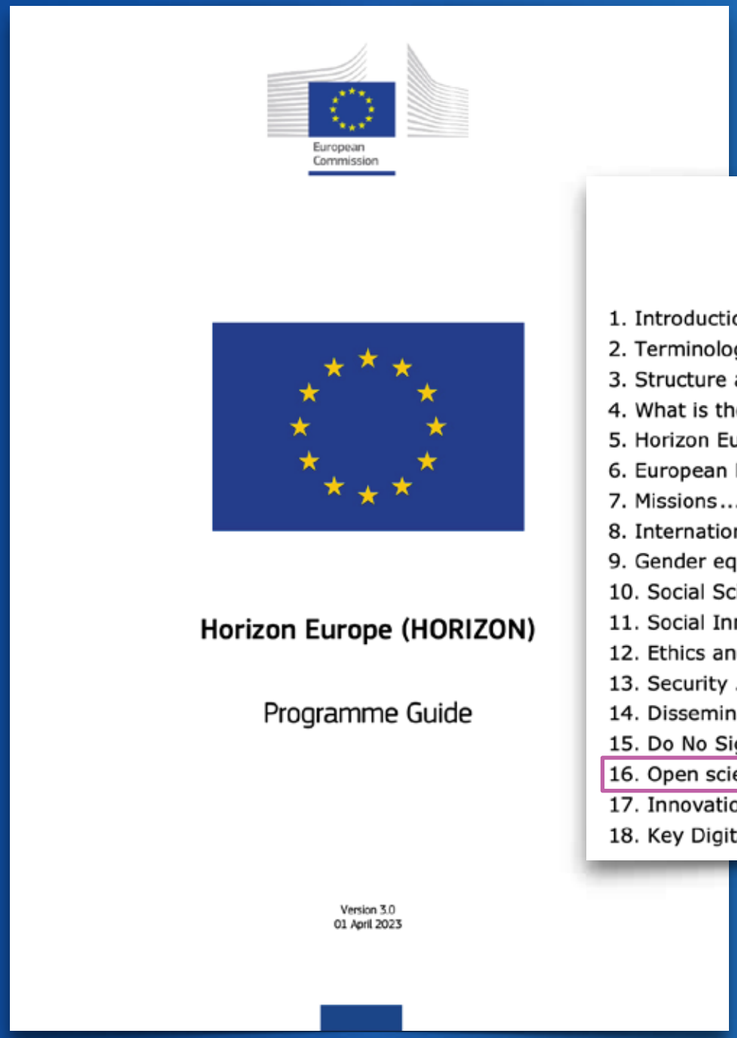


Table of contents

Requirements

- Open access
- Open & FAIR data
- + more

Horizon Europe Grants

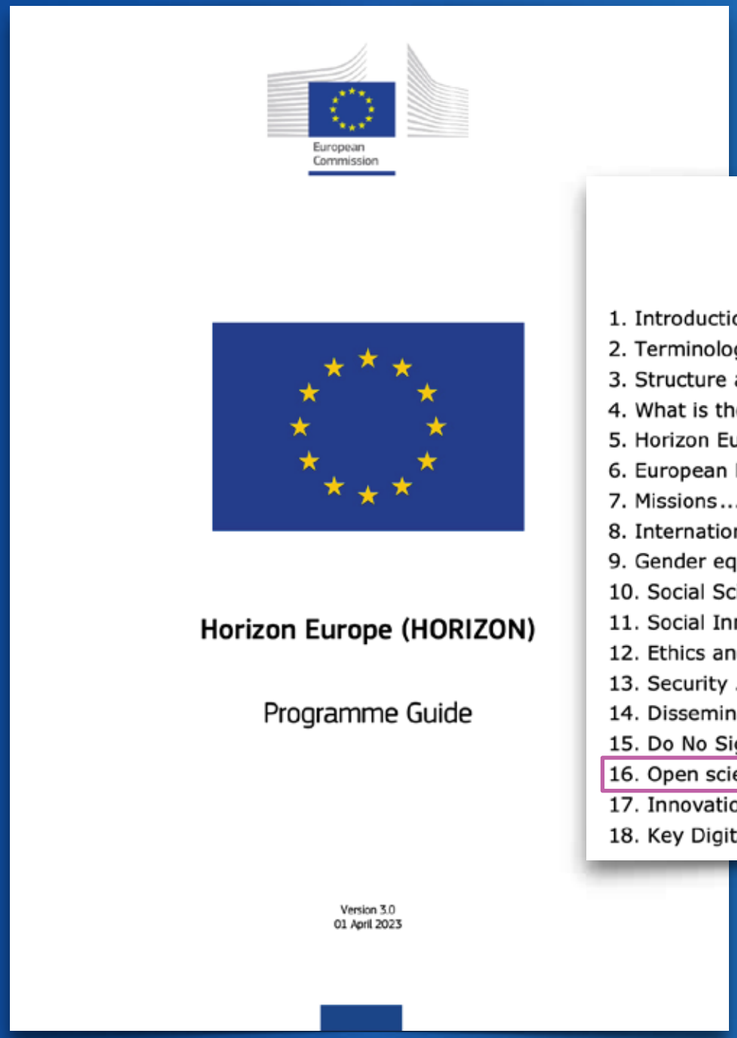


Table of contents

Requirements

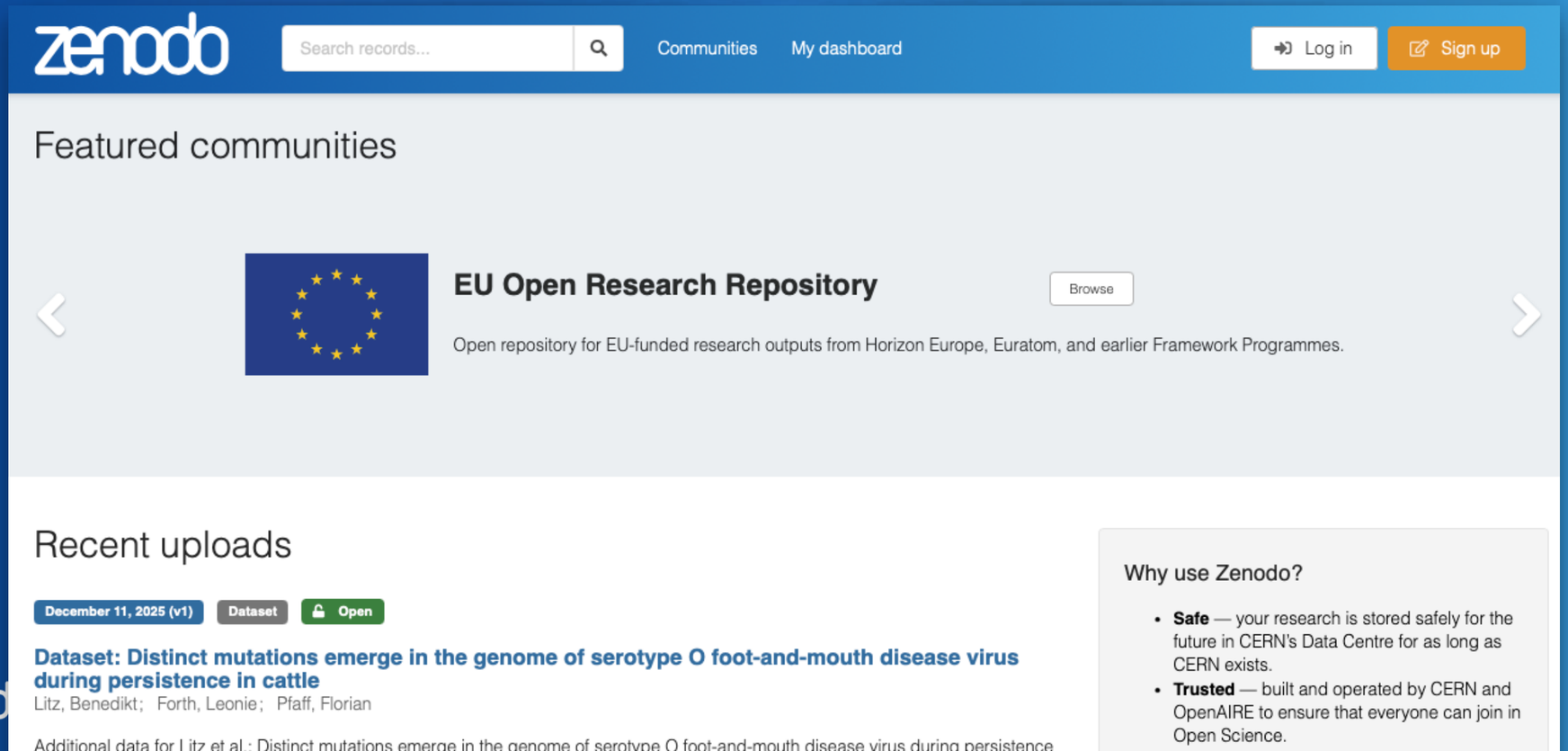
- Open access
- Open & FAIR data
- + more

How?

zenodo

CERN

A general-purpose repository for research products



NIH National Institutes of Health
Turning Discovery Into Health

ARCADIA

Alfred P. Sloan FOUNDATION

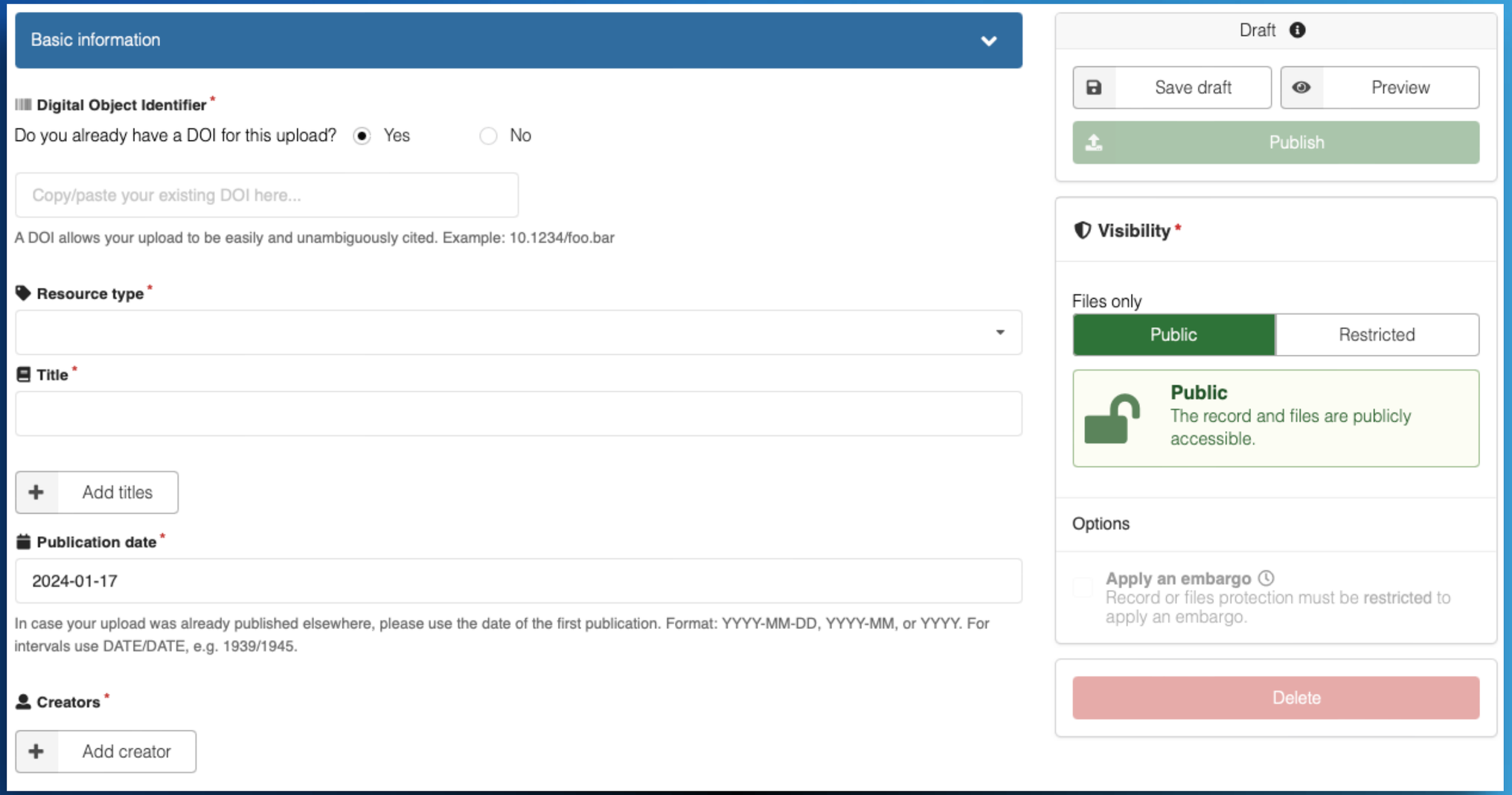
Upload

Publications Data Software Presentations

Posters Videos Image

zenodo CERN

Describe



zenodo CERN

Publish

zenodo | Search records... | Communities | My dashboard | Log in | Sign up

Industrial Ecology and Sustainability Research

Published January 28, 2025 | Version 3.9.4

Dataset | Open

EXIOBASE 3

Stadler, Konstantin[1, 2]; Wood, Richard[1, 2]; Bulavskaya, Tatyana[3]; Södersten, Carl-Johan[1]; Simas, Moana[1]; Schmidt, Sarah[1]; Usubiaga, Arkaitz[4]; Acosta-Fernández, José[4]; Kuenen, Jeroen[3]; Bruckner, Martin[5]; Giljum, Stefan[5]; Lutter, Stephan[5]; Merciai, Stefano[6]; Schmidt, Jannick H[6]; Theurl, Michaela C[7]; Plutzar, Christoph[7]; Kastner, Thomas[8]; Eisenmenger, Nina[7]; Erb, Karl-Heinz[7]; Koning, Arjan[9]; Deck, Candy Eugenie Charlotte Anquetil Ep[1]; Rasul, Kajwan[2]; Hertwich, Edgar[10, 2]; Tukker, Arnold[9]

Show affiliations

EXIOBASE 3: For best in class environmental-economic accounting data. Get insight into global supply-chains and the environmental impacts of consumption.

EXIOBASE 3 provides a time series of environmentally extended multi-regional input-output (EE MRIO) tables ranging from 1995 to 2020 (plus now-casted tables for 2021 and 2022) for 44 countries (27 EU member plus 17 major economies) and five rest of the world regions.

EXIOBASE is maintained by the EXIOBASE consortium, with XIO Sustainability Analytics now working on providing annual updates to

177K VIEWS | 160K DOWNLOADS

Show more details

Versions

Version 3.9.4 — Jan 28, 2025
10.5281/zenodo.14614930

Version 3.8.2 — Oct 21, 2021

DOI 10.5281/zenodo.14869924

10.5281/zenodo.4588235

Version 3.8 — Nov 19, 2020
10.5281/zenodo.4277368

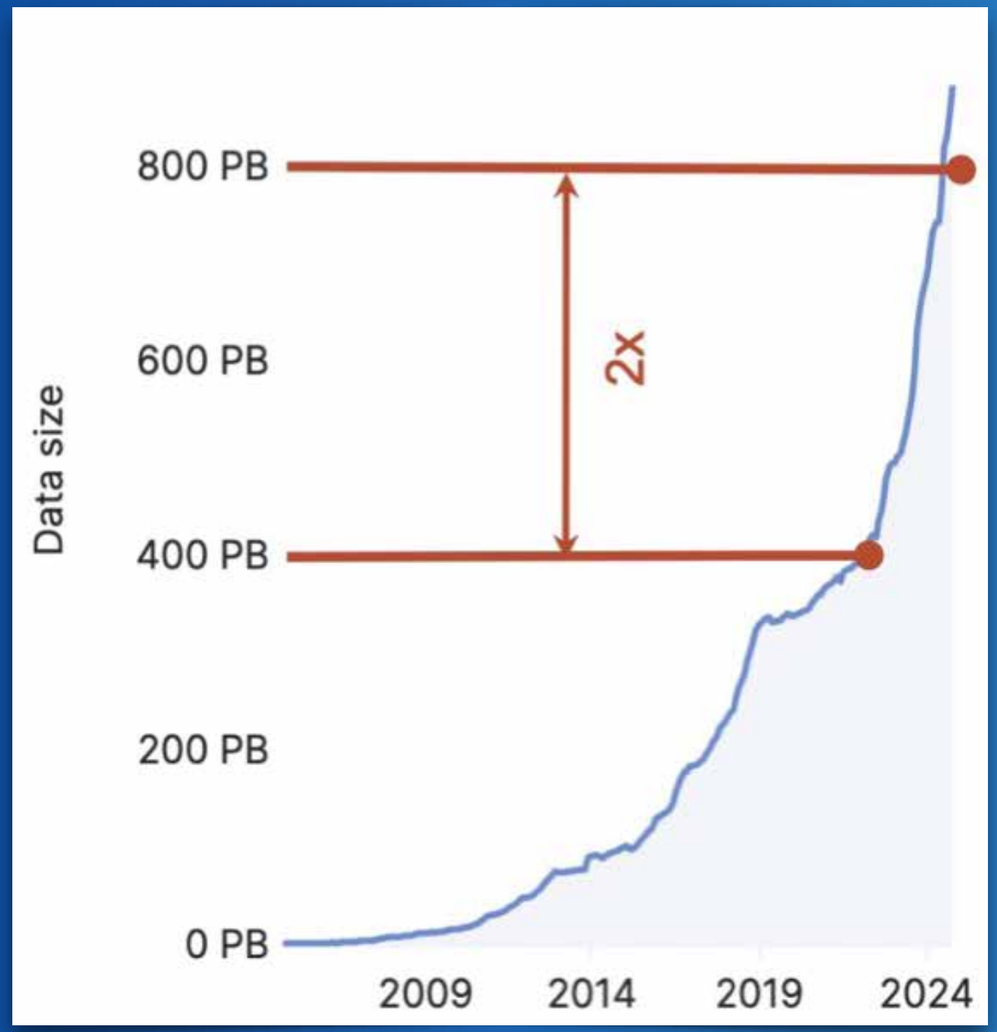
Since 2013 …

400.000 users

9000 organisations

161 countries

Zenodo: A small layer on top of CERN existing infrastructure



CERN

1 EXABYTE

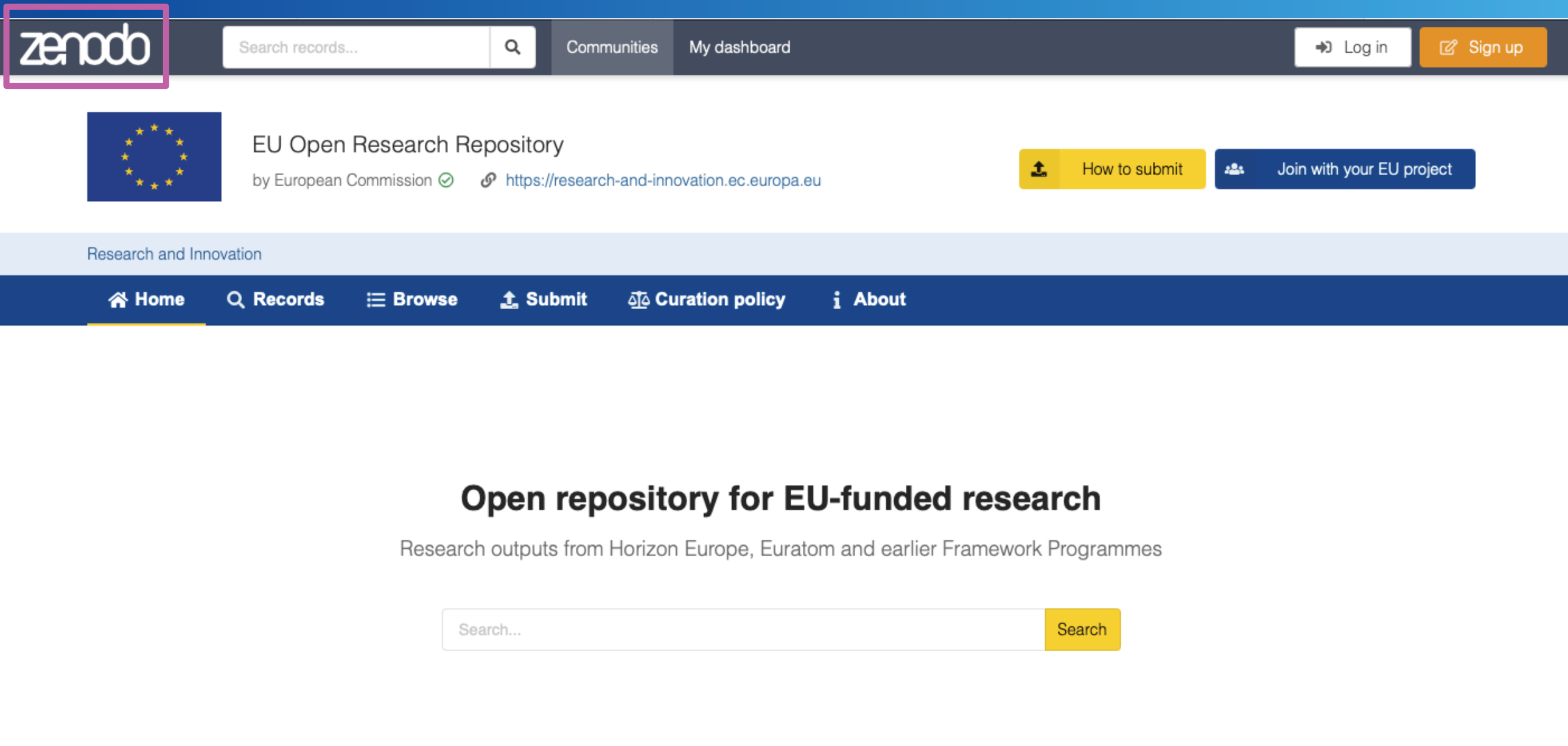
zenodo

1 PETABYTE

1 TERABYTE

zenodo CERN

Since 2024 …


zenodo
Search records...
Communities
My dashboard
Log in
Sign up
EU Open Research Repository
by European Commission
https://research-and-innovation.ec.europa.eu
How to submit
Join with your EU project
Research and Innovation
Home
Records
Browse
Submit
Curation policy
About
Open repository for EU-funded research
Research outputs from Horizon Europe, Euratom and earlier Framework Programmes
Search...
Search

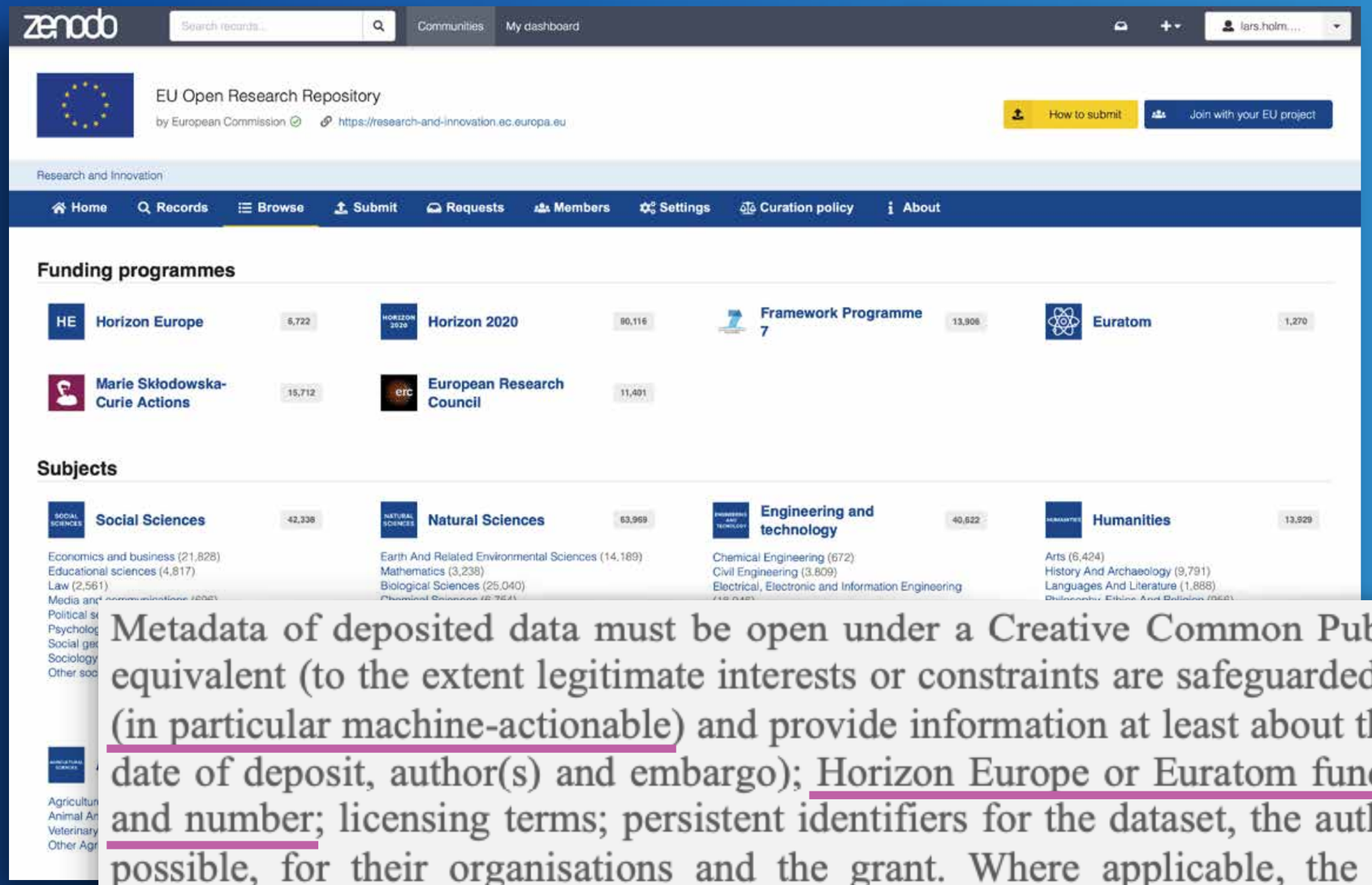
Key challenge 1

Data curation

Estimated cost: 2 FTE for *minimal* curation

Source: Nielsen, L. H. (2024). Sustainability plan for EU Open Research Repository (D5.5) (v1.0). Zenodo. https://doi.org/10.5281/zenodo.11393461

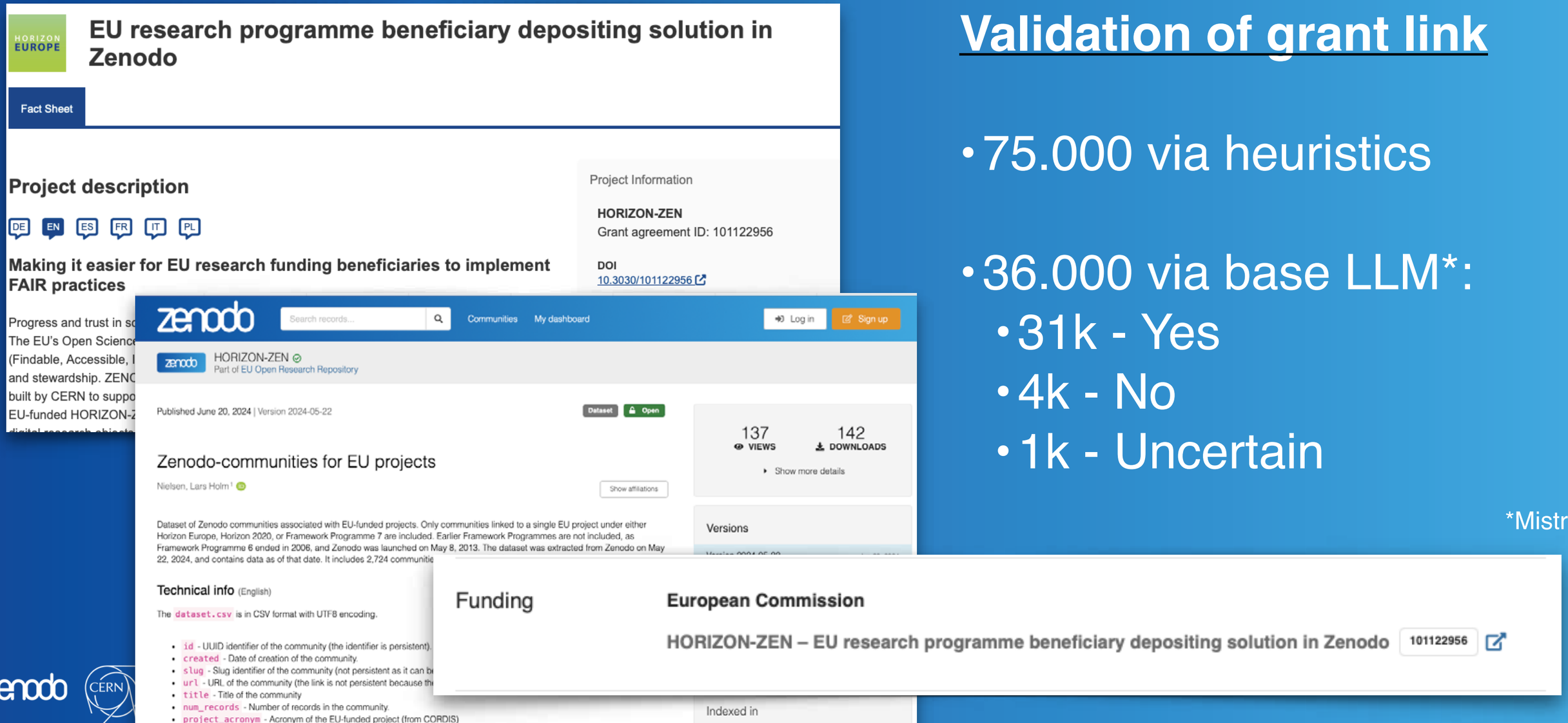
Curating existing content



- 4.5M records - how much is EC funded?
- Trustworthy grant link?

Metadata of deposited data must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: datasets (description, date of deposit, author(s) and embargo); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the dataset, the authors involved in the action, and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for related publications and other research outputs.

Open data, heuristics and AI



Validation of grant link

- 75.000 via heuristics
- 36.000 via base LLM*:
  - 31k - Yes
  - 4k - No
  - 1k - Uncertain

*Mistral

zenodo CERN

Zenodo

Funding programmes

- Horizon Europe — 9,353
- Horizon 2020 — 96,129
- Framework Programme 7 — 14,070
- Euratom — 1,291
- Marie Skłodowska-Curie Actions — 16,671
- European Research Council — 11,968

Subjects

Social Sciences — 45,128

Economics and business (23,354)
Educational sciences (4,836)
Law (2,792)
Media and communications (742)
Political sciences (13,172)
Psychology (1,203)
Social geography (2,308)
Sociology (22,239)
Other social sciences (1)

Natural Sciences — 68,203

Earth And Related Environmental Sciences (15,168)
Mathematics (3,393)
Biological Sciences (26,453)
Chemical Sciences (7,261)
Physical Sciences (16,768)
Computer And Information Sciences (38,268)
Other Natural Sciences (0)

Engineering and technology — 43,100

Chemical Engineering (774)
Civil Engineering (4,044)
Electrical, Electronic and Information Engineering (20,029)
Environmental Biotechnology (761)
Environmental Engineering (16,365)
Industrial Biotechnology (1,731)
Materials Engineering (6,088)
Mechanical Engineering (8,779)
Medical Engineering (758)
Nanotechnology (3,390)
Other Engineering And Technologies (3,874)

Agricultural sciences — 15,452

Agriculture, Forestry, And Fisheries (12,925)
Animal And Dairy Science (2,031)
Veterinary Sciences (70)
Other Agricultural Sciences (0)

Medical and Health sciences — 15,426

Basic Medicine (6,313)
Clinical Medicine (4,958)
Health Sciences (8,509)
Medical Biotechnology (1,942)
Other Medical Sciences (157)

CORDIS

Fields of science

natural sciences > biological sciences > ecology > **ecosystems**

Keywords

biodiversity | genomics | literature | collections | taxon names | linked open data | biodiversity knowledge graph | open science | data interoperability | reproducibility | data extraction | publishing

Programme(s)

H2020-EU.1.4. - EXCELLENT SCIENCE - Research Infrastructures (MAIN PROGRAMME)

H2020-EU.1.4.1.2. - Integrating and opening existing national and regional research infrastructures of European interest

Topic(s)

INFRAIA-02-2020 - Integrating Activities for Starting Communities

Call for proposal

H2020-INFRAIA-2018-2020

See other projects for this call

Sub call

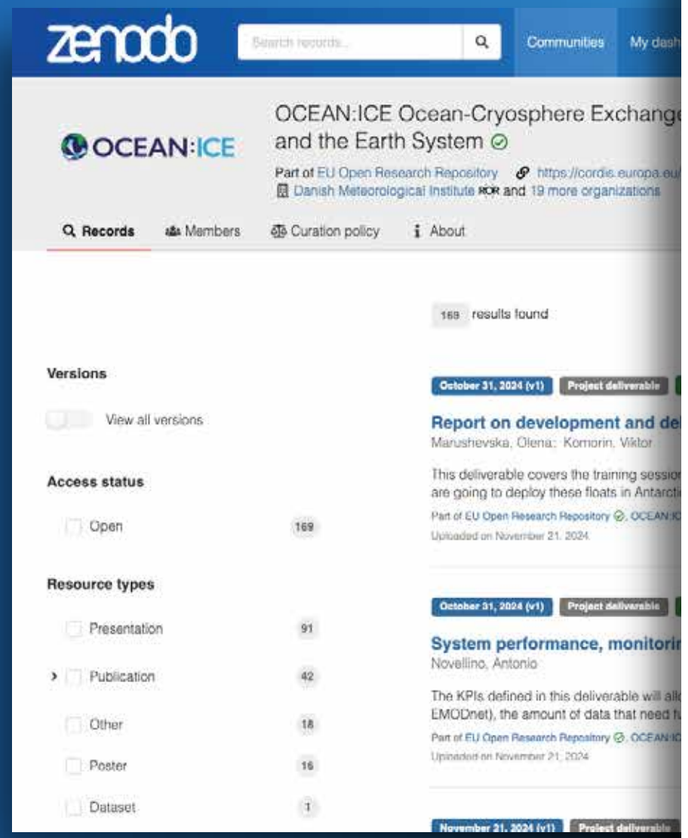
H2020-INFRAIA-2020-1

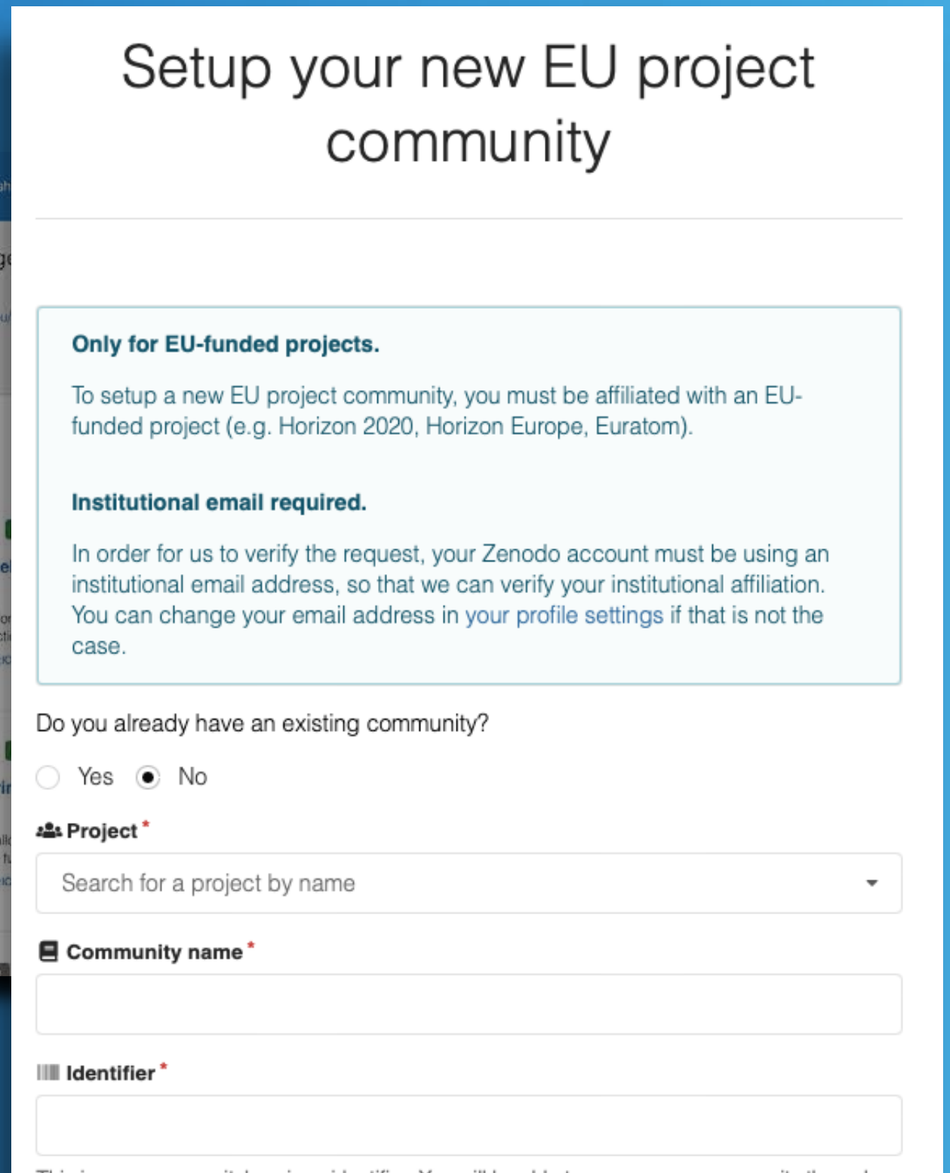
Funding Scheme

zenodo CERN

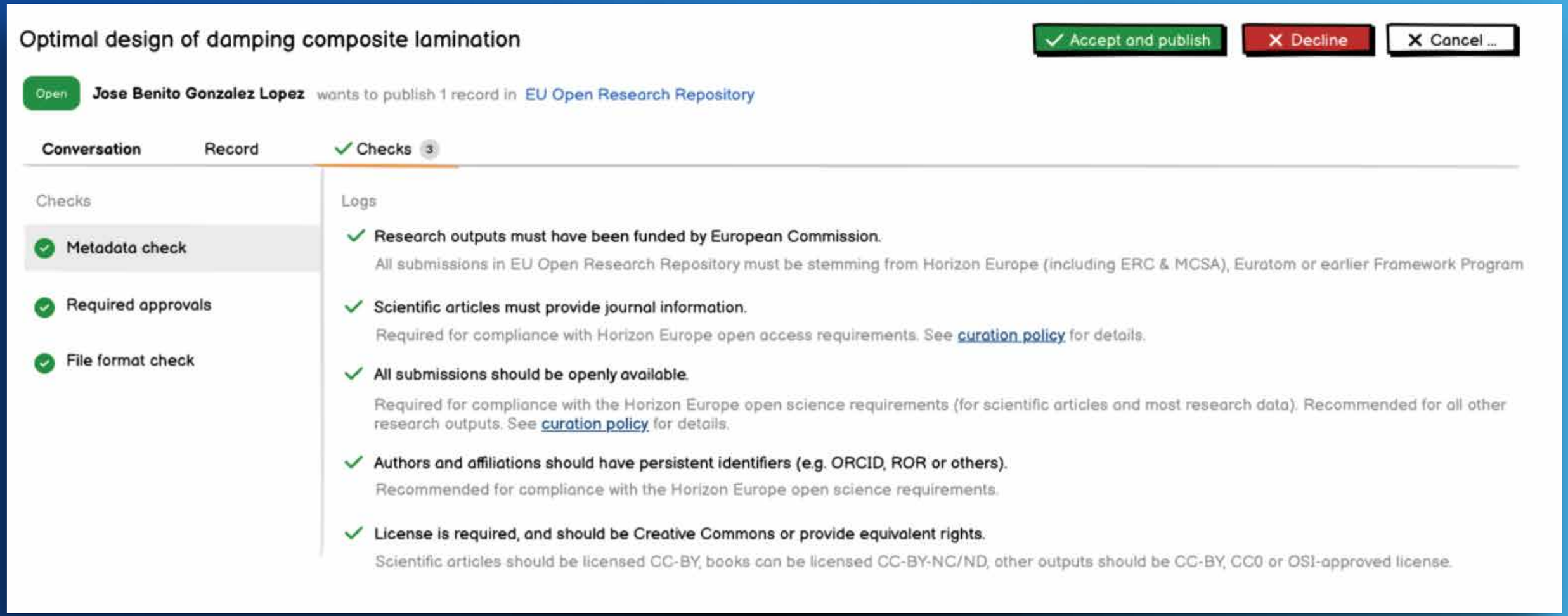
New content: Distributed curation

- Currently 2700+ EU projects
- Easy collaboration across institutions
- Harmonisation?





Harmonisation & compliance



Planned launch in April 2025

zenodo CERN

Data curation framework

Distrbute

Automate

Incentivise

= Lower cost, higher quality

Key challenge 2

FAIR* Data Principles

*Findable, Accessible, Interoperable, Reusable

Example: Description of species

Geographic coordinates

Date of collection

Collector

Material (total: 3 ♂♂)

Holotype

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

Paratypes

TANZANIA: 1 ♂, 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 ♂, 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, T. Pape and N. Scharff leg. (ZMUC).

Host collection

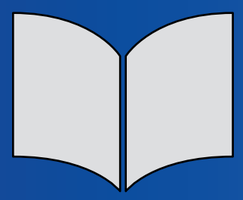
Locked up data

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



Article

PLAZI
TAKING CARE OF FREEDOM

Text and data mining

zenodo

Preserve & disseminate

SIB LS
SWISS INSTITUTE OF BIOINFORMATICS
LITERATURE SERVICES
BiodiversityPMC

GBIF Global Biodiversity Information Facility

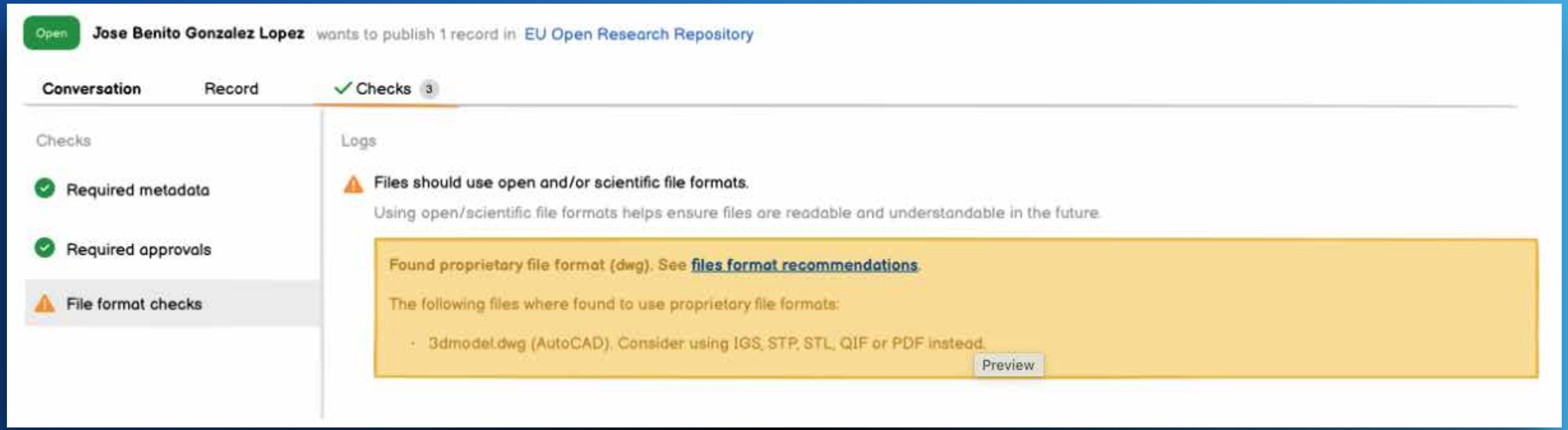
Catalogue of Life

Search, retrieve & linking

FAIR Metadata

What about the actual data?

File format checks



Planned launch in April 2025

zenodo CERN

Next

Ease of use

Using AI + Supporting AI

EU OPEN DATA DAYS

19 – 20 MARCH 2025

EU Open Research Repository

Enabler of open science policy

Ease of use

Lower the cost to increase quality

Funded by the European Union under Horizon Europe grant agreement number 101122956.

EU OPEN DATA DAYS

19 – 20 MARCH 2025

Thank you!

EU Open Research Repository

Enabler of open science policy

Ease of use

Lower the cost to increase quality

Funded by the European Union under Horizon Europe grant agreement number 101122956.