NFDI4Ing Conference 2022

How to use re3data – Creating, editing, and analyzing information on research data repositories

Wednesday, October 26th, 10:30 - 12:00

Rouven Schabinger, Dorothea Strecker & Nina Leonie Weisweiler

Schedule

10:30 - 10:45	Introduction to re3data
10:45 - 11:15	The editorial process in re3data, new entries and change requests
11:15 - 11:45	The re3data API: Obtaining and using metadata
11:45 - 12:00	Summary and questions

Registration with CoCalc

If you want to participate in querying the re3data API later in the session, please create an (anonymous) account with CoCalc:

https://cocalc.com/auth/sign-up

A Project Invite Token will be shared later.

DataCite

KIT Karlsruher Institut für Technologie

PURDUE UNIVERSITY

Global Registry of Research Data Repositories

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

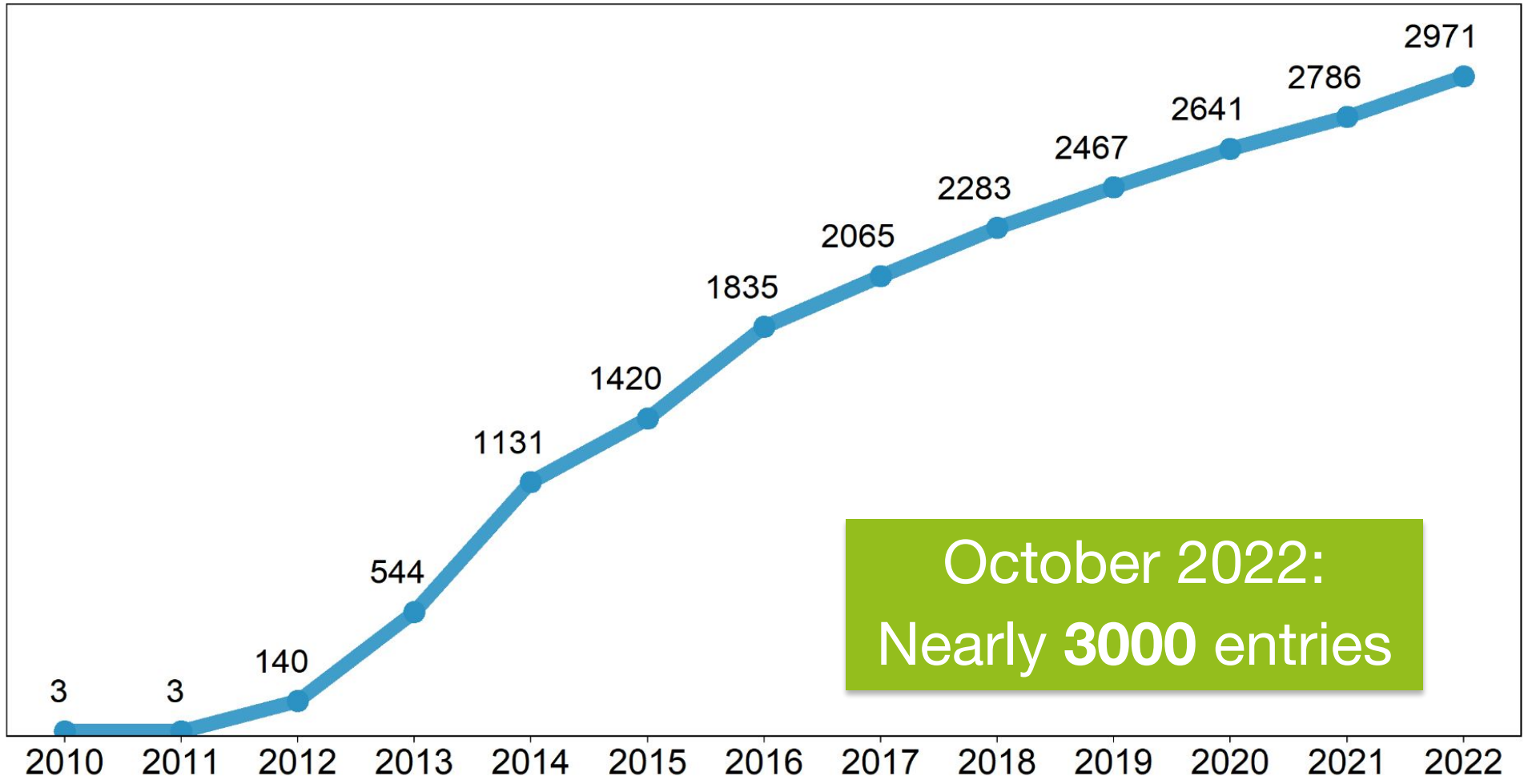
Search... Search

across **all research disciplines**

… presents **repositories** for the **permanent storage** and **access** of research data sets to researchers, funding bodies, publishers and scholarly institutions.

…promotes a culture of **sharing, increased access and better visibility** of research data

Number of repositories indexed in re3data per year:



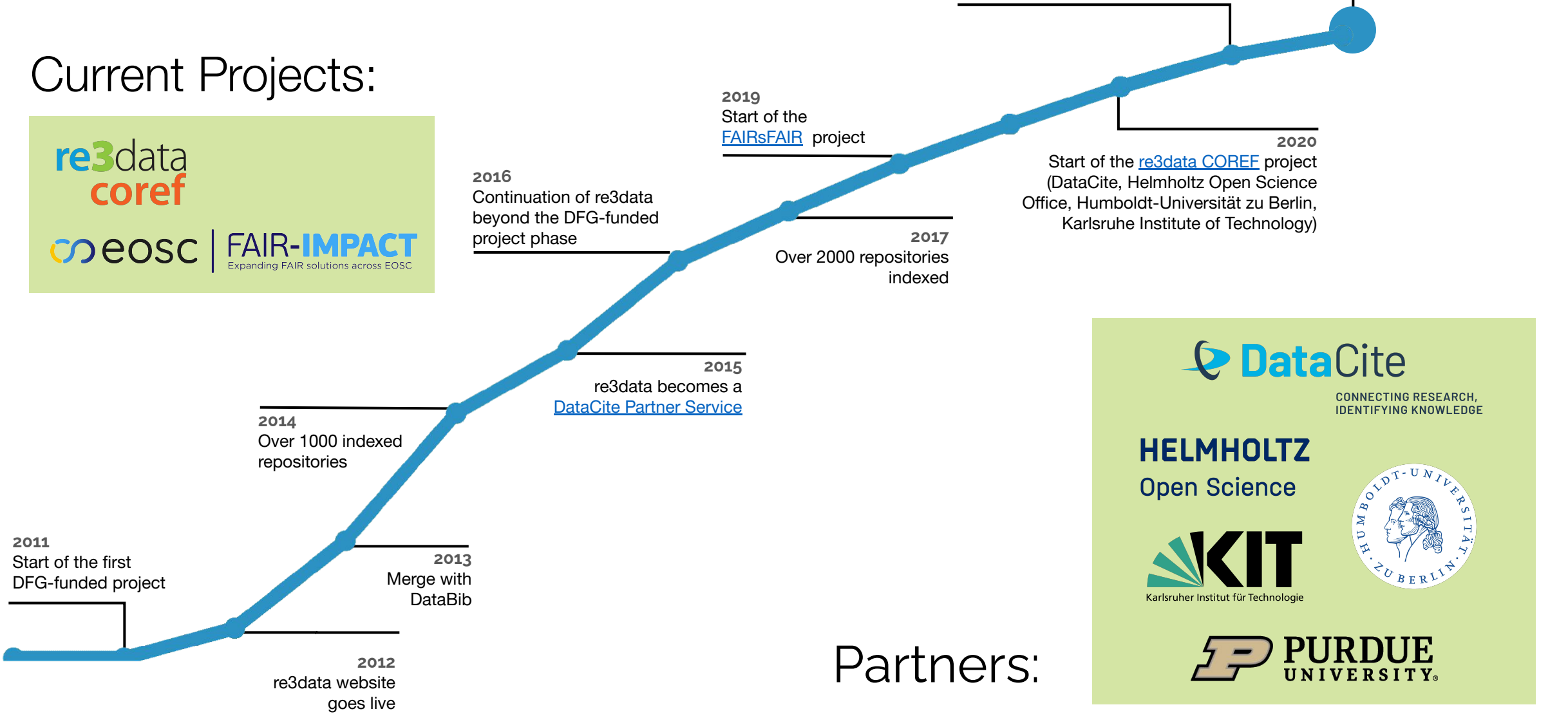
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Celebrating 10 Years of re3data

Current Projects:

re3data coref

eosc | FAIR-IMPACT
Expanding FAIR solutions across EOSC



Partners:

DataCite
CONNECTING RESEARCH, IDENTIFYING KNOWLEDGE

HELMHOLTZ Open Science

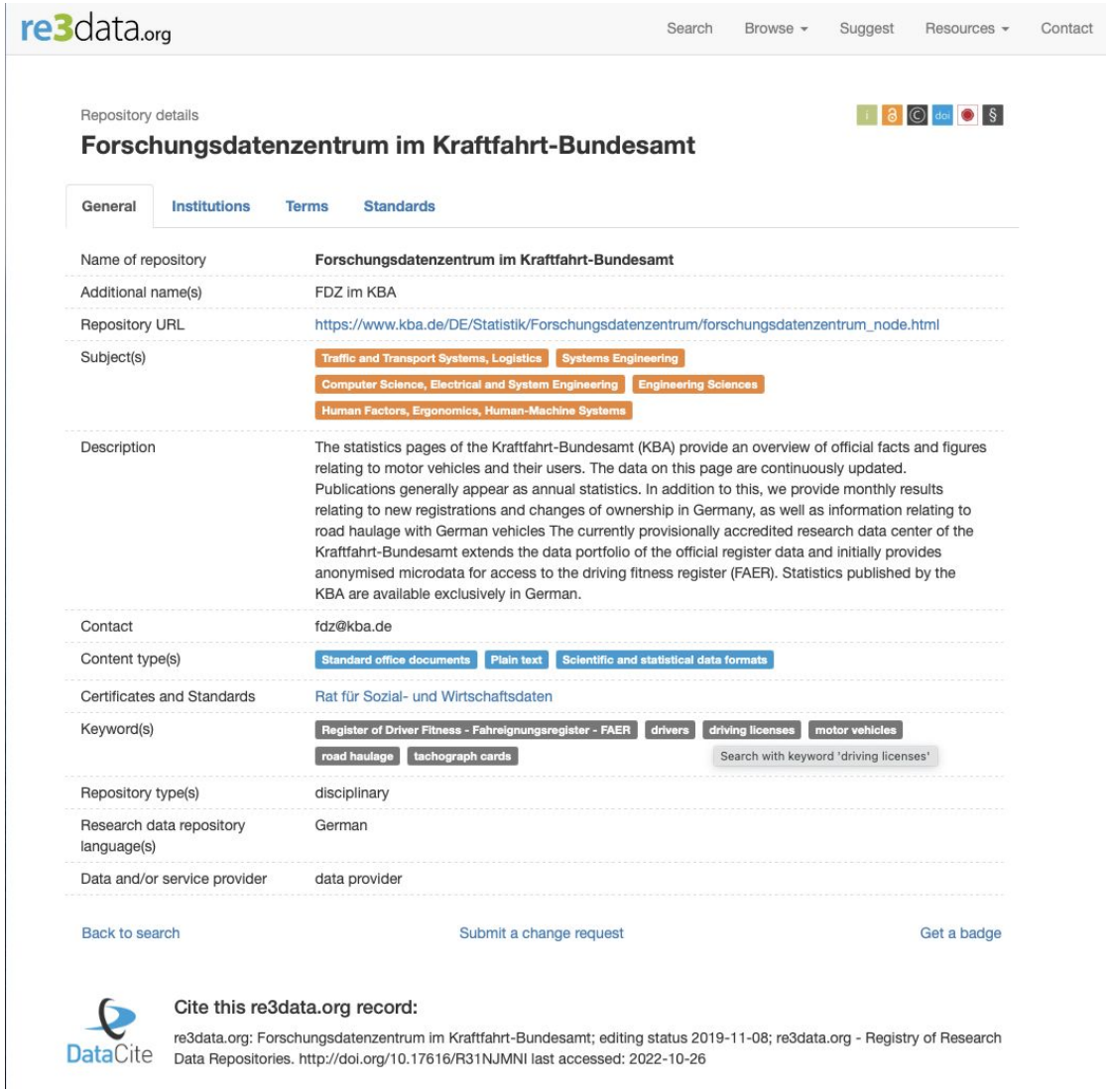
KIT Karlsruher Institut für Technologie

HUMBOLDT-UNIVERSITÄT ZU BERLIN

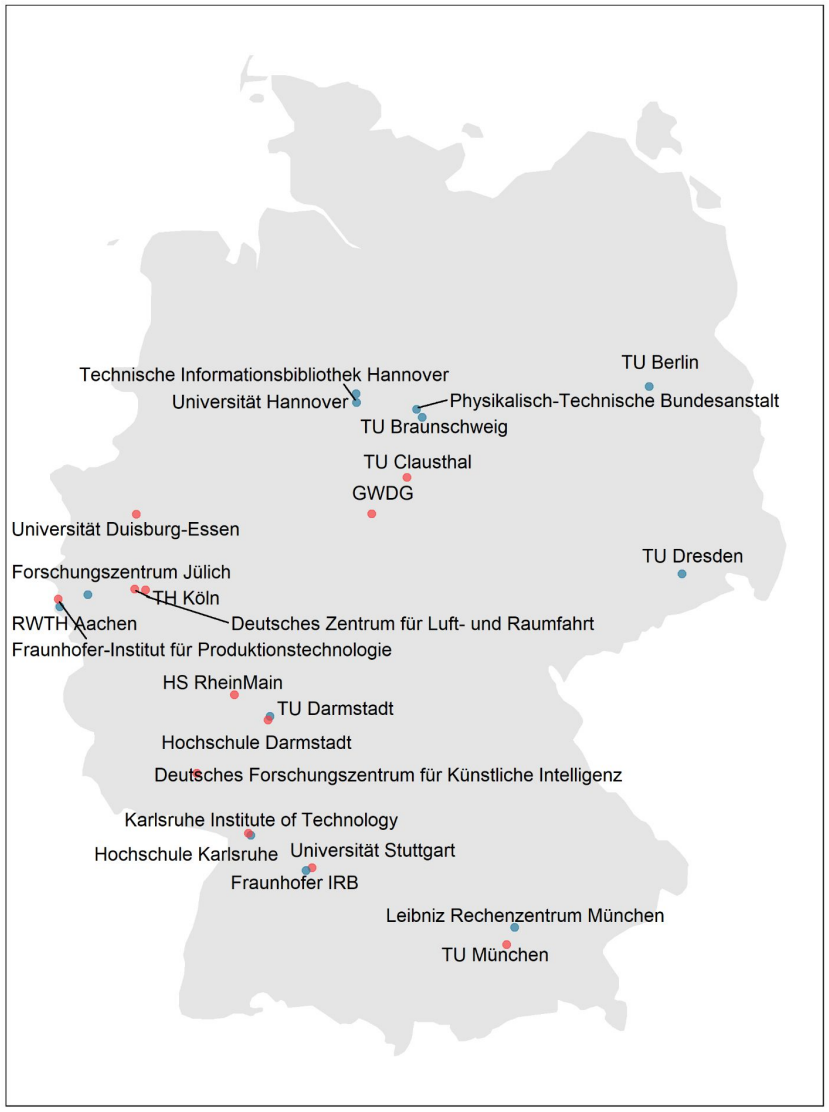
PURDUE UNIVERSITY

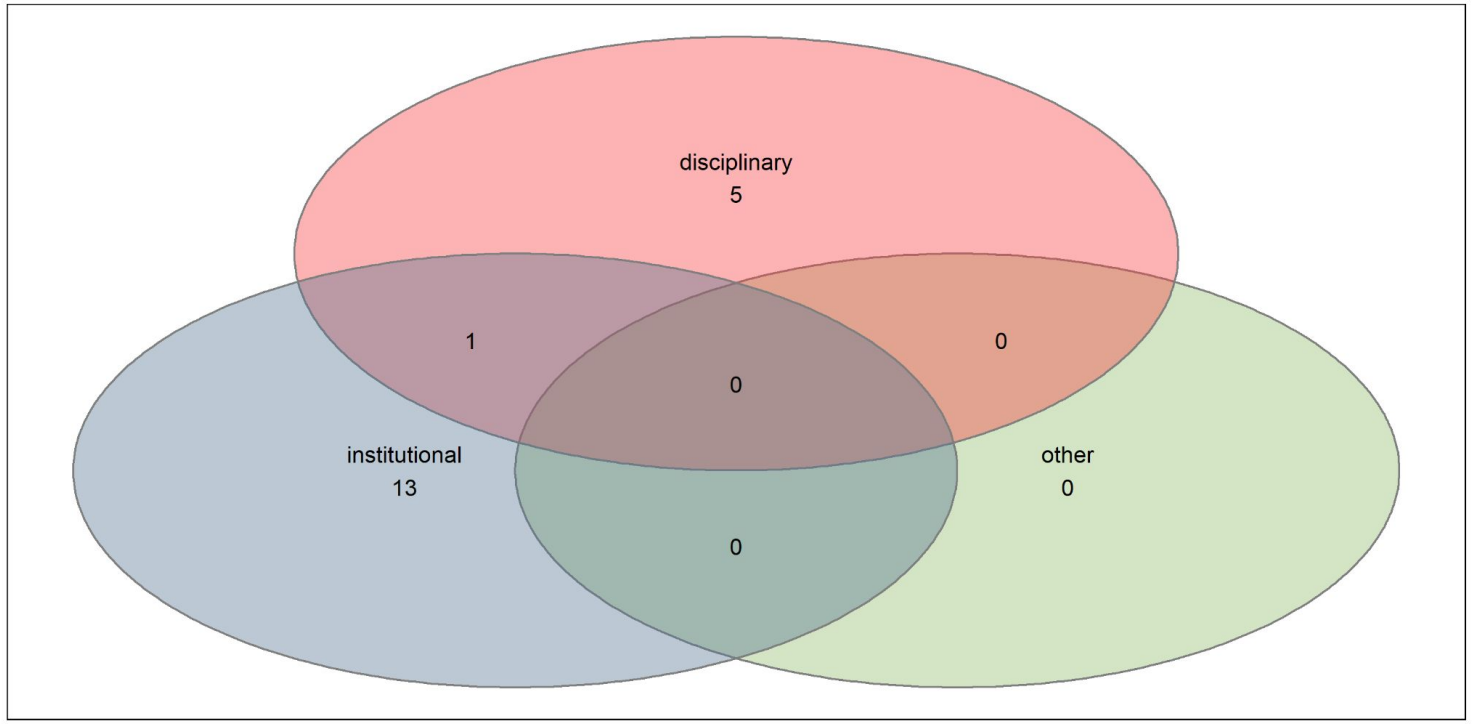
What is a “repository” in re3data?

- We generally refer to all re3data entries as “repositories”.
  - Repositories in re3data are defined as “[...] a subtype of a sustainable information infrastructure providing long-term storage of and access to research data.”(Metadata Schema for the Description of Research Data Repositories, Version 3.1)
- The re3data registration policy is intentionally inclusive to cover a broad range of infrastructures for research data.



Repositories of NFDI4Ing institutions





As of February 2022, 18 repositories indexed in re3data are affiliated with NFDI4Ing (co-)applicant and participant institutions and also cater to the engineering sciences.

Most of them are institutional repositories.

Registration Policy

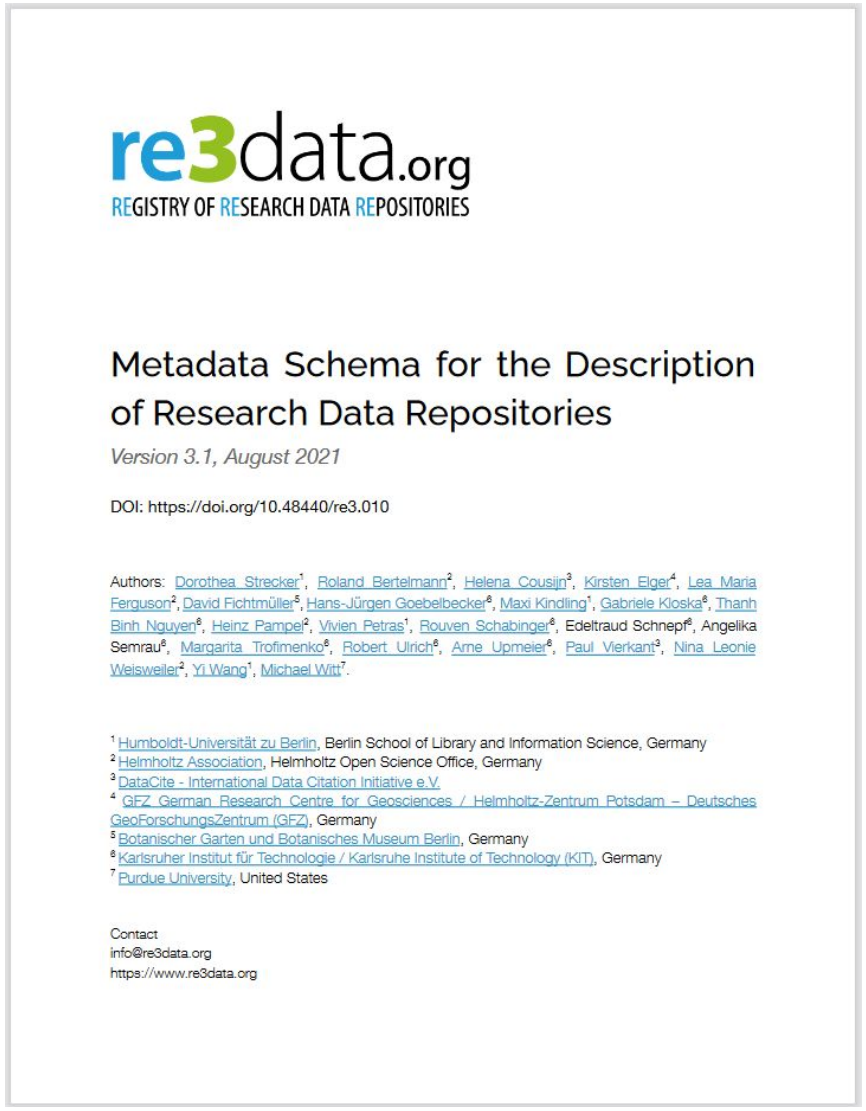
re3data.org Registration Policy

To be registered in re3data.org a research data repository must

- be run by a legal entity, such as a sustainable institution (e.g. library, university)
- clarify access conditions to the data and repository as well as the terms of use
- **have focus on research data**

A research data repository is a subtype of a sustainable information infrastructure which provides long-term storage and access to research data that is the basis for a scholarly publication. Research data means information objects generated by scholarly projects for example through experiments, measurements, surveys or interviews.

Metadata Schema Version 3.1



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Metadata Schema for the Description of Research Data Repositories

Version 3.1, August 2021

DOI: https://doi.org/10.48440/re3.010

Authors: Dorothea Strecker[1], Roland Bertelmann[2], Helena Cousijn[3], Kirsten Elger[4], Lea Maria Ferguson[2], David Fichtmüller[5], Hans-Jürgen Goebelbecker[6], Maxi Kindling[1], Gabriele Kloska[6], Thanh Binh Nguyen[6], Heinz Pampel[2], Vivien Petras[1], Rouven Schabinger[6], Edeltraud Schnepf[6], Angelika Semrau[6], Margarita Trofimenko[6], Robert Ulrich[6], Arne Upmeier[6], Paul Vierkant[3], Nina Leonie Weisweiler[2], Yi Wang[1], Michael Witt[7].

[1] Humboldt-Universität zu Berlin, Berlin School of Library and Information Science, Germany
[2] Helmholtz Association, Helmholtz Open Science Office, Germany
[3] DataCite - International Data Citation Initiative e.V.
[4] GFZ German Research Centre for Geosciences / Helmholtz-Zentrum Potsdam – Deutsches GeoForschungsZentrum (GFZ), Germany
[5] Botanischer Garten und Botanisches Museum Berlin, Germany
[6] Karlsruher Institut für Technologie / Karlsruhe Institute of Technology (KIT), Germany
[7] Purdue University, United States

Contact
info@re3data.org
https://www.re3data.org

Strecker S., et al. (2021): Metadata Schema for the Description of Research Data Repositories: version 3.1, https://doi.org/10.48440/re3.010

42 Properties on…

- General information
- Responsibilities
- Policies
- Legal aspects
- Technical standards
- Quality standards
- DOI for repository descriptions
- Since 2020: ROR IDs for institutions

Version 3.1 of the Metadata Schema:
https://www.re3data.org/schema

Icons – facilitating the selection process of appropriate research data repositories

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

re3data as a metadata provider

re3data is a **central reference point and metadata source** for other service providers, see for example:

- Open Science Observatory
- European Open Science Monitor
- DataCite Commons
- F-UJI Automated FAIR Data Assessment Tool
- DARIAH-EU Data Deposit Recommendation Service

re3data coref

COMMUNITY DRIVEN OPEN REFERENCE FOR RESEARCH DATA REPOSITORIES

Project blog: coref.project.re3data.org

Conceptual Service Model

- Embed the registry within the research community and infrastructure landscape to meet the emerging needs for a trusted repository reference.

Metadata Schema

- Revision and enhancement of the re3data metadata schema for the description of research data repositories.

Technical Infrastructure

- Implement new features and the updated metadata schema.
- Provide the necessary technical infrastructure and services

Quality & Transparency

- Conduct a comprehensive study on quality assurances at research data repositories to improve re3data metadata information
- Improve re3data metadata quality, trust and transparency.

Community Building

- Extensive outreach activities, networking and policy development.

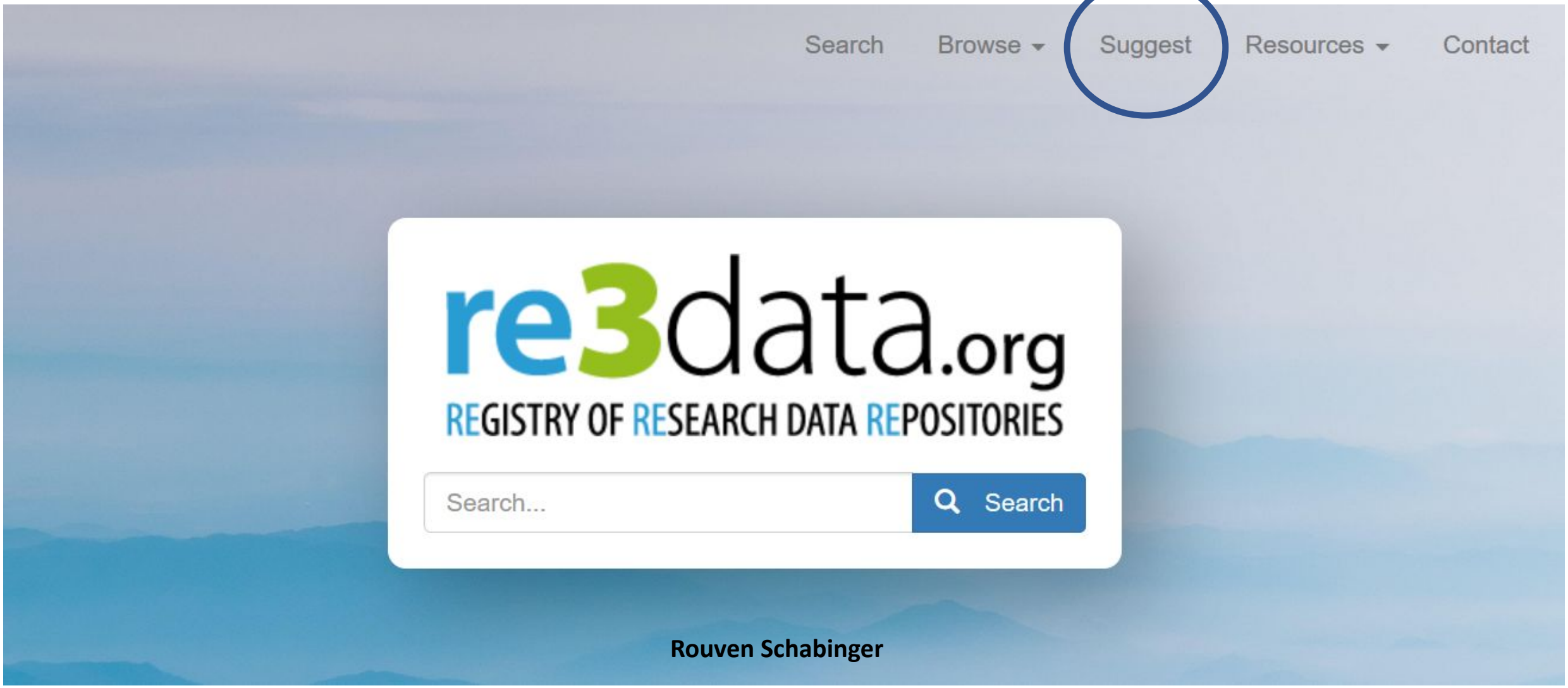
DataCite

HELMHOLTZ Open Science

HUMBOLDT-UNIVERSITÄT ZU BERLIN

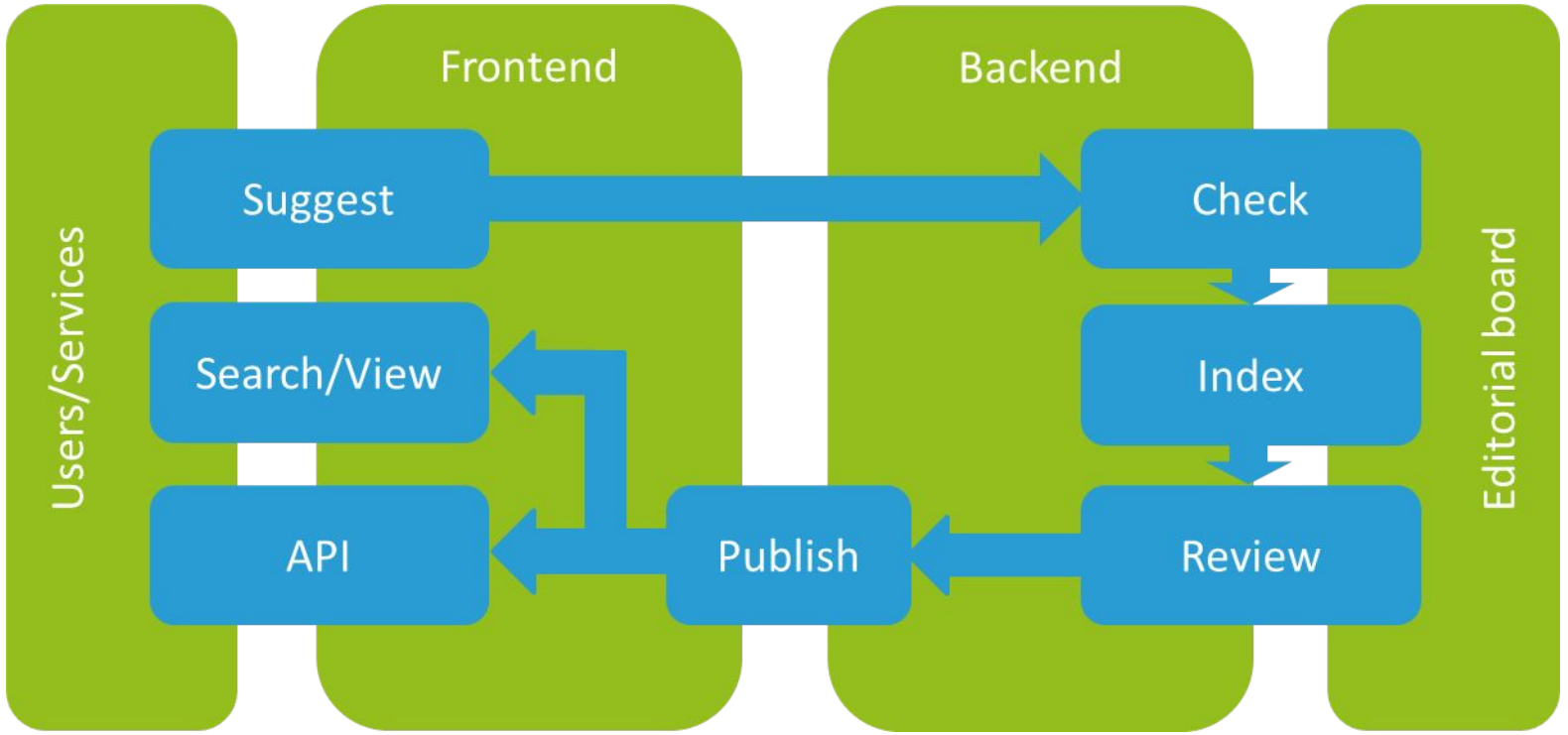
KIT Karlsruher Institut für Technologie

What makes a good registry record?



Rouven Schabinger

Editorial Workflow



Workflow for new repository suggestions and change requests for re3data

Editorial Board

Catherine Jones

Science and Technology Facilities Council (STFC)

Jiban K. Pal

Indian Statistical Institute (ISI)

Iris Lindenmann

University of Basel

Rouven Schabinger

Karlsruhe Institute of Technology (KIT)

Angelika Semrau

Karlsruhe Institute of Technology (KIT)

Edeltraud Schnepf

Karlsruhe Institute of Technology (KIT)

Sarah J. Wright

Cornell University

Hui Wang

Chinese Academy of Sciences (CAS) / National Science Library (NSL)

Gabriele Weickert

Karlsruhe Institute of Technology (KIT)

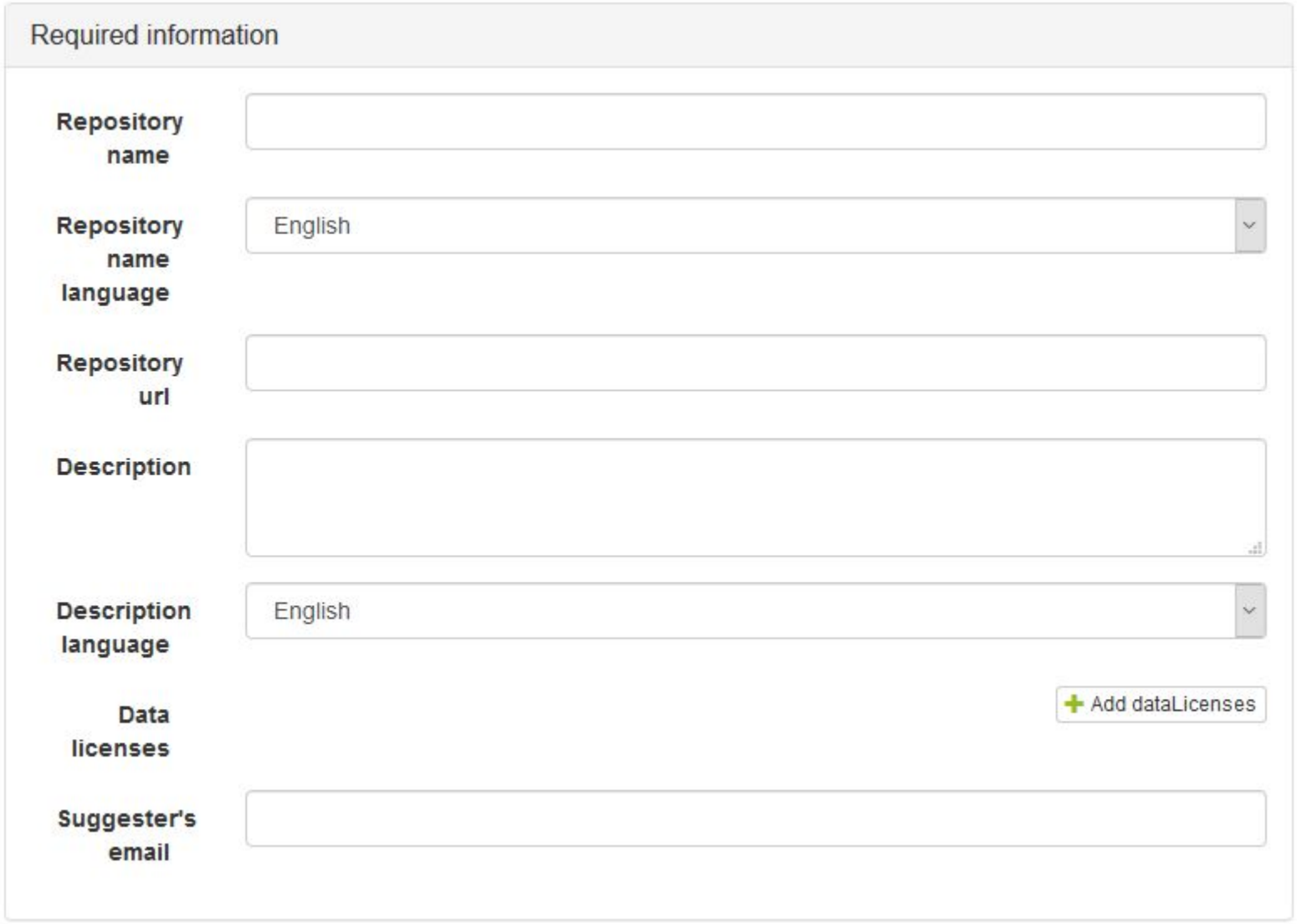
Sarah Williams

University of Illinois

Michael Witt

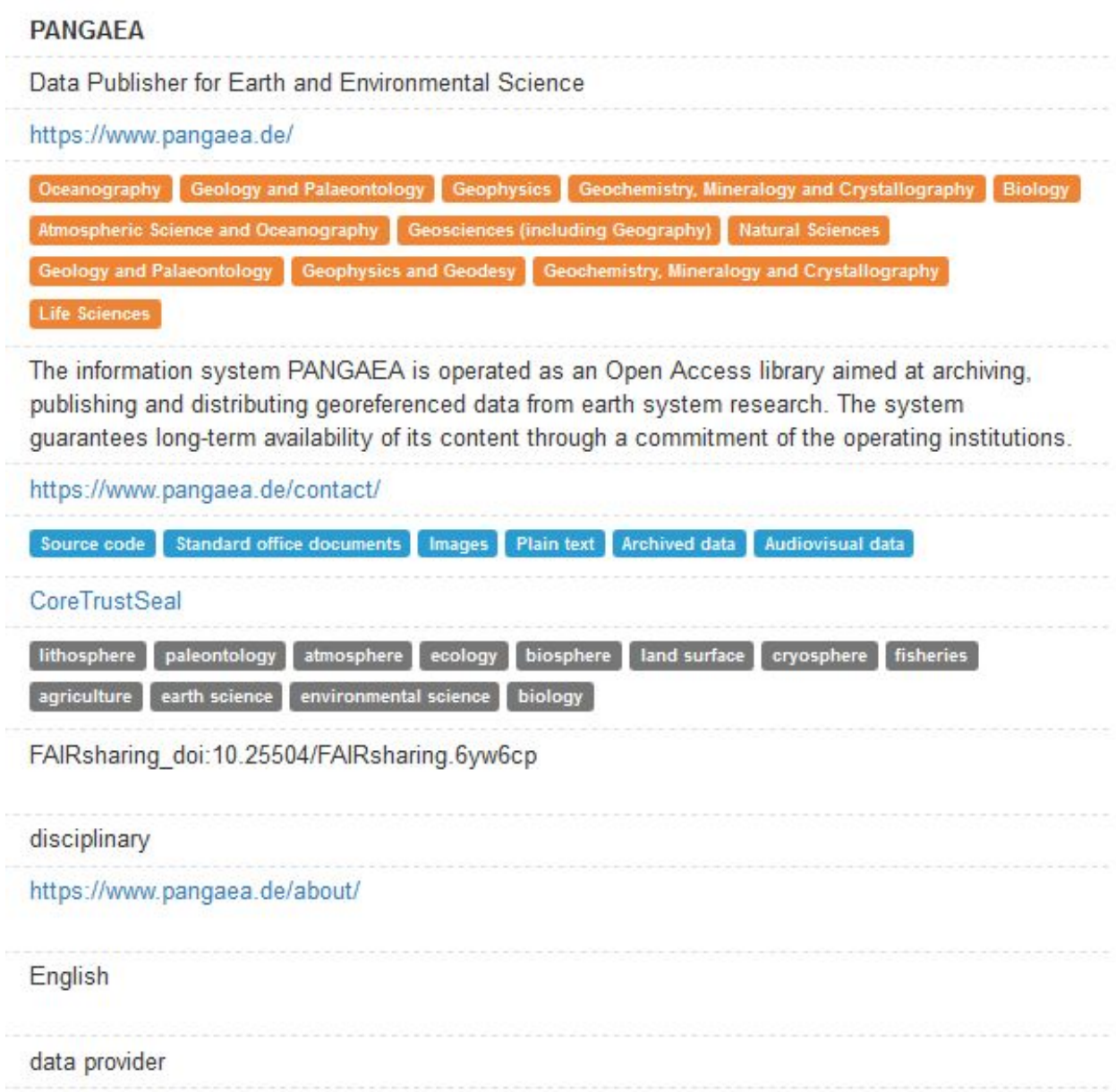
Purdue University

Required information



General

- Additional names
- **Subjects**
- **Repository contacts**
- Content types
- **Certificates**
- Keywords
- **Repository identifiers**
- Size
- **Types**
- Mission statement URL
- Start date
- Repository languages
- Provider types



Institutions

- Name
- Name language
- **Country**
- Type
- URL
- Responsibility start date
- Responsibility end date
- Additional names
- Contacts
- Identifiers
- **Responsibility types**

Terms

- Policies
- **Database access**
- **Data accesses**
- **Data uploads**
- Data upload licenses

Policies (2)

Policy Name	CoreTrustSeal assessment
Policy Name	Data policy of the information system PANGAEA

Database access

Type of access to research data repository	open

Database licenses (1)

Database License	CC0

Data access (1)

Type of access to data	open

Data licenses (4)

DataLicense	CC
DataLicense	other

Access (property)	*Open Access*		*Restricted Access*		*Closed Access*
Access to Repository (20.1 databaseAccessType)	open		open or restricted		closed
Access to Data (22.1 dataAccessType)	open (embargoed, restricted, closed)		restricted (embargoed, closed)		closed
Data Upload (24.1 dataUploadType)	open or restricted	closed	open or restricted	closed	-

Standards

- Software
- Versioning
- **PID systems**
- Citation guideline URL
- AID systems
- Enhanced publication
- Quality management
- **APIs**
- **Metadata standards**
- Syndications
- Remarks

DOI

other

yes

https://wiki.pangaea.de/wiki/Citation

ORCID

unknown

yes

interfaces (1)

OAI-PMH

Darwin Core

DCC

ISO 19115

DCC

Dublin Core

DCC

DIF - Directory Interchange Format

DCC

Keeping your record up-to-date

- Automated updates (e.g. CTS)
- Reviews by the editorial team
- Your change request?

Contact	https://nanohub.org/about/contact
Content type(s)	Databases · Audiovisual data · Software applications
Keyword(s)	Nanomaterial Registry · nanoBIO · nanoscience · nan
Persistent identifier(s) of the repository	RRID:SCR_013963 OMICS_27120
Repository size	6.139 resources
Repository type(s)	disciplinary
Mission statement for designated community	https://nanohub.org/about
Research data repository language(s)	English
Data and/or service provider	data provider service provider

Back to search

Submit a change request

Submit a change request

Make changes to the properties that need an update. The editorial board will review the submitted record and put it online

General

Repository name nanoHUB

Repository name language English

Additional names — **Text** nanoHUB.org

Language English

Repository url https://nanohub.org/

Subjects — **Name** 31 Chemistry

Scheme DFG

pre-filled metadata fields

Get a badge

Demo Instance

- URL: https://demo.project.re3data.org
- Login: re3data.org
- Password: NFDI4Ing2022

Querying the re3data API with R

We will use the platform CoCalc.

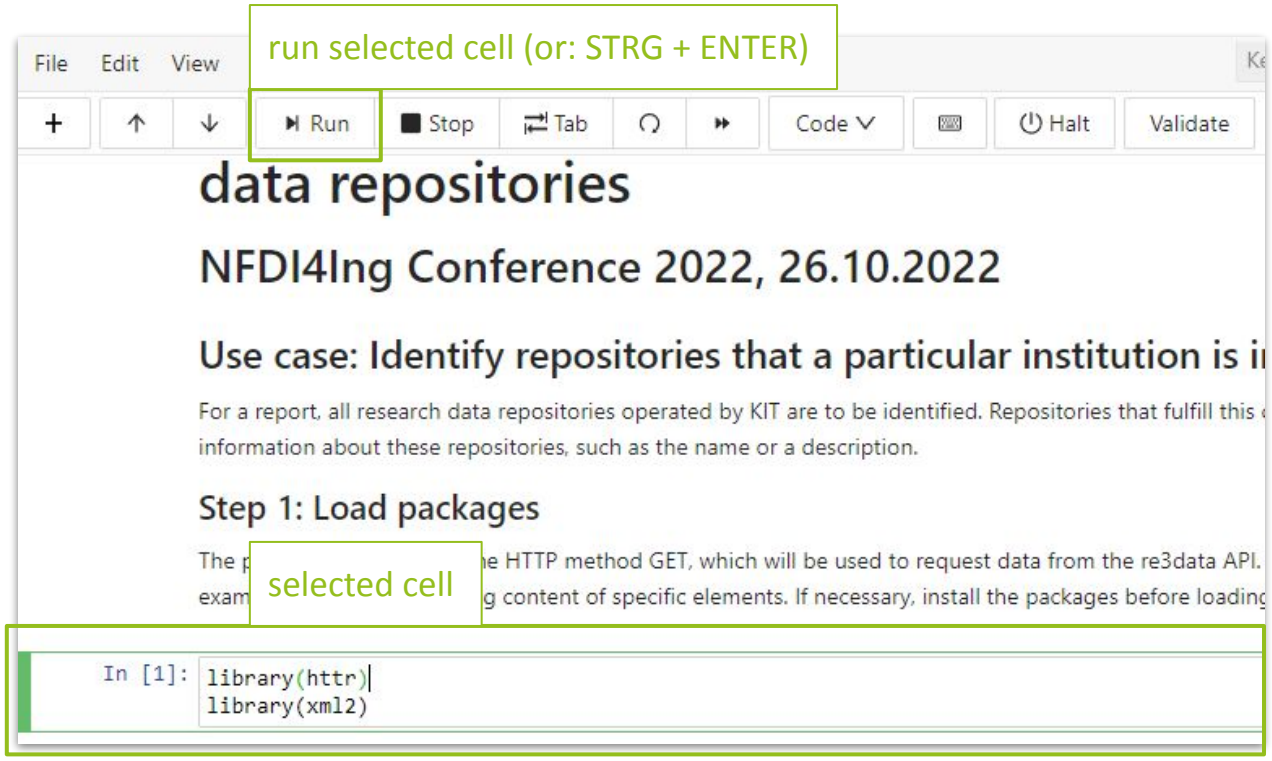
1. log in to CoCalc using the account you created earlier
2. click on “projects” in the top left corner
3. enter the Project Invite Token in the top right corner:

   CLS5VYnBuB7WJTZE

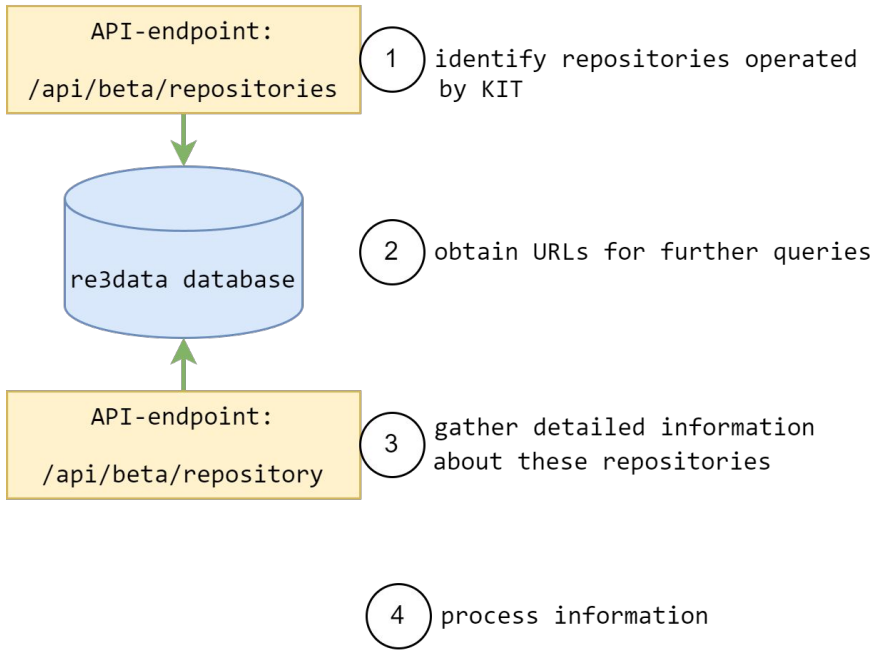
4. start the project NFDI4Ing Conference 2022

Querying the re3data API with R

Using Jupyter Notebooks



Schematic overview of the process



re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

Stay in touch!

Web: **www.re3data.org**

Mail: **info@re3data.org**

Twitter: **@re3data**

Project Blog:

coref.project.re3data.org

Rouven Schabinger (Karlsruher Institut für Technologie)

Dorothea Strecker (Humboldt-Universität zu Berlin)

Nina Weisweiler (Helmholtz-Gemeinschaft, Helmholtz Open Science Office)

https://creativecommons.org/licenses/by/4.0/