Research Data Management

Data Archiving & Repositories

Silvia Sofianos / silvia.sofianos@cvtisr.sk

Gabriela Fišová / gabriela.fisova@cvtisr.sk

Slovak Centre of Scientific and Technical Information

3rd March 2026, STU in Bratislava

Financované Európskou úniou

PROGRAM SLOVENSKO

MINISTERSTVO ŠKOLSTVA, VÝSKUMU, VÝVOJA A MLÁDEŽE SLOVENSKEJ REPUBLIKY

CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR

NISPEZ

Archiving & Repositories

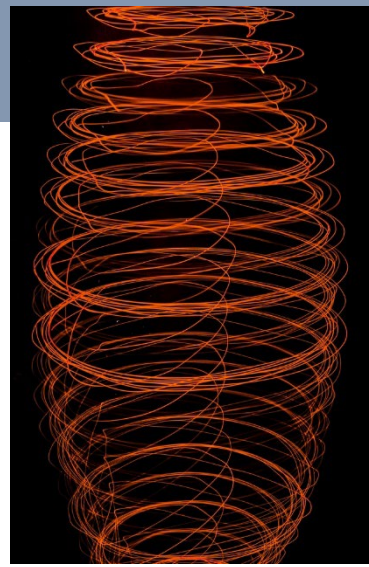
CONTENTS

- Memory of the scientific record
- Archiving–objectives
- Archiving vs. backup
- Selection of digital repository
- Trusted digital repositories
- Zenodo

Image: Digitalbevaring.dk:Jørgen Stamp, CC BY 2.5 DK, via Wikimedia Commons

Archiving

Archiving–memory of the scientific record



- **Archiving** → **continuity of scientific record**
- Research papers and datasets are **searched for and cited** (=necessary) even centuries after their publication

  Not all of them, of course, but no one can determine in advance which will be the "lucky ones".
- **In the past**: **hard copies** of the journal or research output records remained in the **library/archive**

– **a paper book can last for centuries**—longer than a CD-ROM **but** journals/newspapers made of inferior (acidic) paper will disintegrate after a few decades without replacement—if we want to preserve their content, it must be **digitized.**

- **The fragility of digital content (**articles, books, **research data)**

How to store electronic information so that it remains readable even decades later?

How to protect electronic information against unwanted modification, falsification, or destruction?

How to prevent Error 404... or → **long-term digital archiving**

Archiving–objectives

BASIC ARCHIVING OBJECTIVES

(a) document integrity,

(b) content availability,

(c) complete reproducibility,

(d) document authenticity,

(e) confidentiality of documents—in some cases

Preserving *archival materials* **for future use over tens or hundreds of years.**

Sources: 1, 2.

Archiving

ARCHIVING

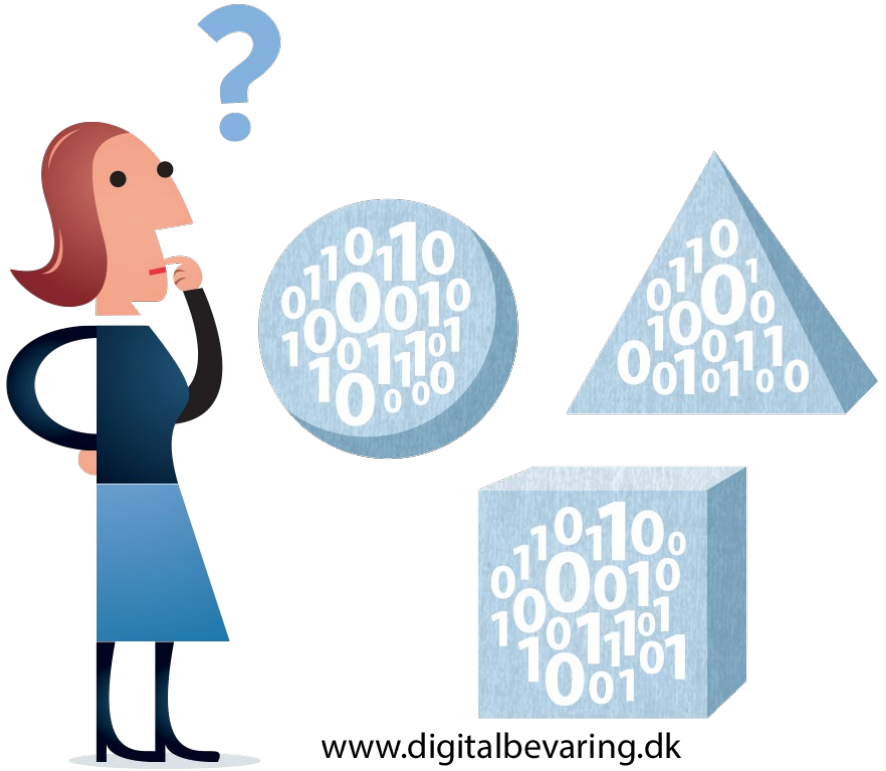
vs

BACK-UP

Backup vs Archive		
	Archive	**Backup**
Primary goal	Prolonged data retention	Disaster recovery
Data copied	Legacy data worth retaining	All data
File versions stored	One	Multiple
Budget	Low	Medium
Necessary storage size	Medium	High
Typical storage	Cloud, tape	Cloud, on-prem
Retention duration	Long	Medium
Entry discoverability	Easy	Hard in legacy backups. Easy in SpinOne.
Automation	No	Yes
Functionality	Data storing Data discovery Download	Data storing, Data discovery, Download Full/granular recovery, Data migration Automated data backup, Reports

Source: https://spin.ai/blog/what-is-data-archiving-definition-benefits-and-best-practices/

Where to go with all that digital content?



Where to go with all that digital content? Sensitivity of data

1. Data categorization

PUBLIC DATA: accessible without any restrictions to anyone

INTERNAL DATA: no special regulation or protection by law/contract required

DISCRETE DATA: regulation or protection required, usually protected by law or under a contract/licence.

SENSITIVE DATA: special regulation or special protection required, usually strictly protected by law or by contract/license.

2. Storage categorization

EXTERNAL PORTABLE MEDIA: Disk, Flash Drive, Memory Cards, DVD Disk – used by users transfer information;

LOCAL SOLUTIONS: desktop computers and laptops in the offices...

INTERNAL NETWORK DRIVES, CLOUD SOLUTIONS: document server

EXTERNAL CLOUD SOLUTION – Contractual: e.g. Team SharePoint - Microsoft

EXTERNAL CLOUD SOLUTION CLOUD – non-cntractual: Google Drive, Microsoft OneDrive, Dropbox, Amazon storage, GitHub repositories etc.

3. Suitability of storage for different types of data

PUBLIC DATA – any storage

INTERNAL DATA
External portable media – encryption recommended
Local solution – screen locks, passwords recommended
Internal network drives, cloud solutions – suitable
External solutions – not suitable

DISCRETE/SENSITIVE (CONFIDENTIAL) DATA
Only internal network drives/cloud solutions – encryption and other security measure might be neccessary

Zdroj: https://osc.mendelu.cz/en/36466-data-categories (*Masaryk University: Recommendations for the Usage of Storages*); https://tech.mendelu.cz/en/35253-storage-facilities-and-their-use; Data management plány a plánování (DMP), EOSC CZ, 2024.

Digital repositories

Repositories

REPOSITORY TYPES

- **Disciplinary/subject-based** – specific discipline/s (*e.g.* ***arXiv****, 1991)*
- **Institutional –***self-archiving + green open access* (DR Žilinskej univerzity/University of Zilina, repozitár UMB/Matej Bell University, Inštitucionálny repozitár SAV(Slovak Academy of Sciences, DR Univerzity Komenského/Comenius University); World Bank Group repository etc.
- **National** – preserve publication output of the **country**
- **Multi-disciplinary (universal/general/catch-all)**
- **Software**
- **DATA**
- **Other**

How to find a suitable data repository?

Obrázok: Image by Peggy und Marco Lachmann-Anke from Pixabay

Repositories

HOW TO FIND A SUITABLE DATA REPOSITORY

1) When selecting a suitable repository for research results, always verify if **there is any disciplinary repository/community repository,** in which the results can be deposited.

2) **General rules** – regardless the chosen repository, always verify if it

- **is sustainable in the long term;**
- **stores data in a secure way;**
- **enables data access control (open access, limited acces, embargo)**
- **ensures that data remain retrievable**, **accessible and reusable** using **persistent identifiers;**
- **describes the data in a standard way using accepted metadata standards;**
- **allows the depositor to specify the licence(s)** governing access to and reuse of the data;
- **enables dataset versioning;**
- **prize.**

Data Repositories

REGISTRIES OF DATA and OPEN ACCESS REPOSITORIES

Registry of Research DATA REPOSITORIES: re3data, (https://www.re3data.org/)

Registry of Open Access Repositories, ROAR, (http://roar.eprints.org/)

Open Directory of Open Access Repositories, OpenDOAR (https://v2.sherpa.ac.uk/opendoar/)

Disciplinary or Subject-based OArepositories

Data Repository Guidance – Nature – different types

https://www.nature.com/sdata/policies/repositories#general

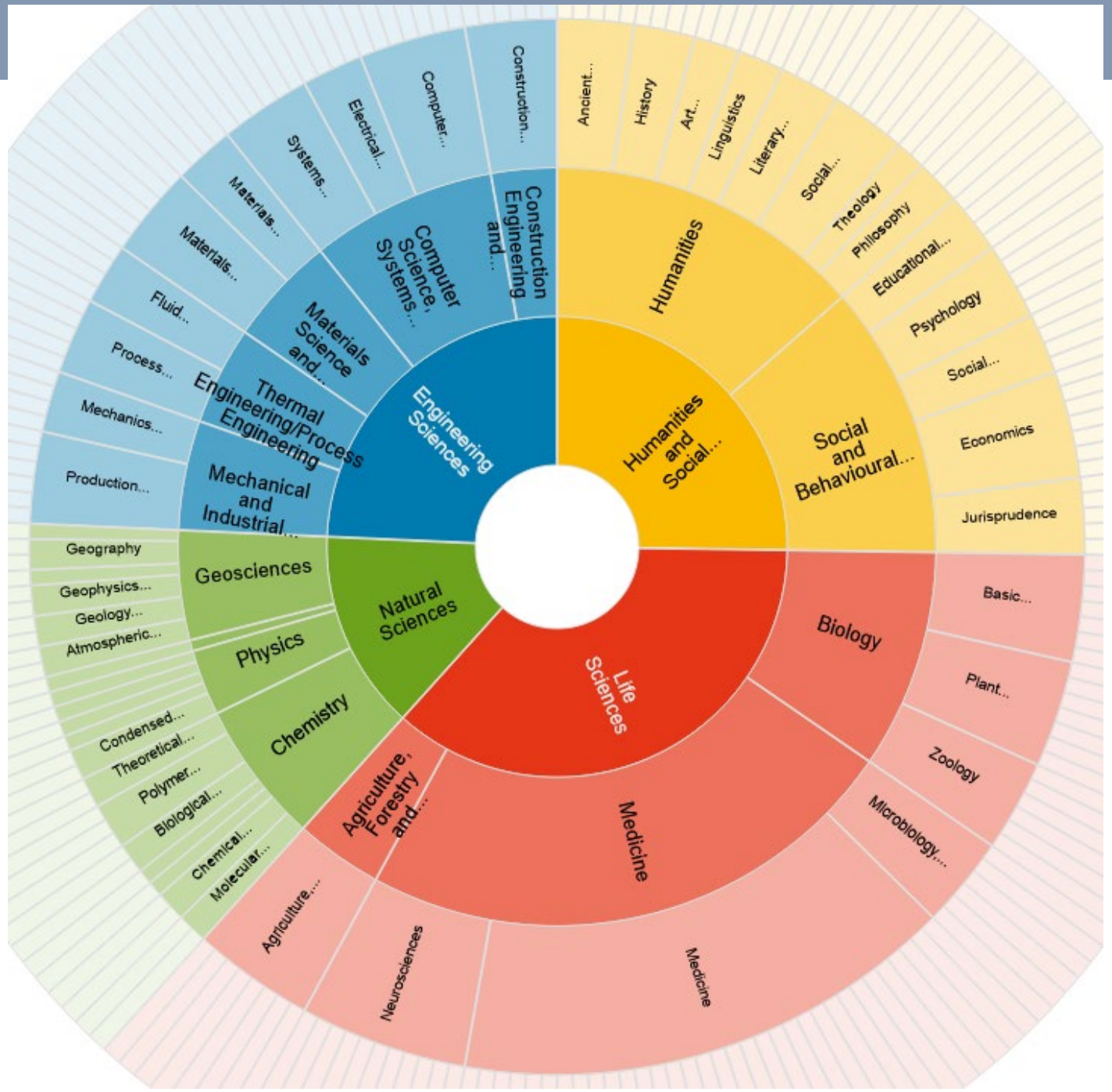
ELIXIR RDMkit| Data publicationEMBL-EBI'sdata submission wizard. ELIXIR DepositionDatabases.

Disciplinary repositories: https://eosc.eu/eosc-task-forces/

Data Repositories

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

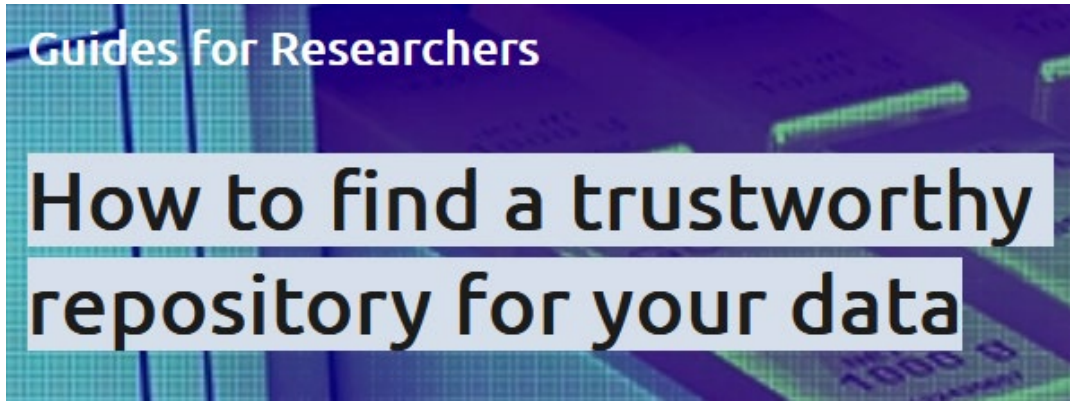
https://www.re3data.org/browse/by-subject/



How to find a trustworthy (trusted) digital repository?

Obrázok:
Jørgen Stamp, CC BY 2.5 DK, via Wikimedia Commons

Data Repositories



- What is required?
- The EC prefers **certified repositories**
- How to comply with the Open Data Pilot requirements?
- Which repository to use?
- Support on metadata, sensitive data, usage licences
- Are data publishing costs supported?
- What are these repository certifications based on?
- How can OpenAIRE help?

OpenAire: https://www.openaire.eu/find-trustworthy-data-repository

Repositories

TRUSTWORTHINESS—AUDIT & CERTIFICATION

ISO 16363 standard: Space data and information transfer systems — Audit and certification of trustworthy digital repositories

- **comprehensive external audit & certification**
- repositories with different content, **on-site**
- Slovak Standard: STN ISO 16363: 2014 (31 0591), Systémy prenosu vesmírnych údajov a informácií. Audit a certifikácia dôveryhodných digitálnych úložísk.
- Will be replaced with ISO/DIS 16363 (https://www.iso.org/standard/56510.html)

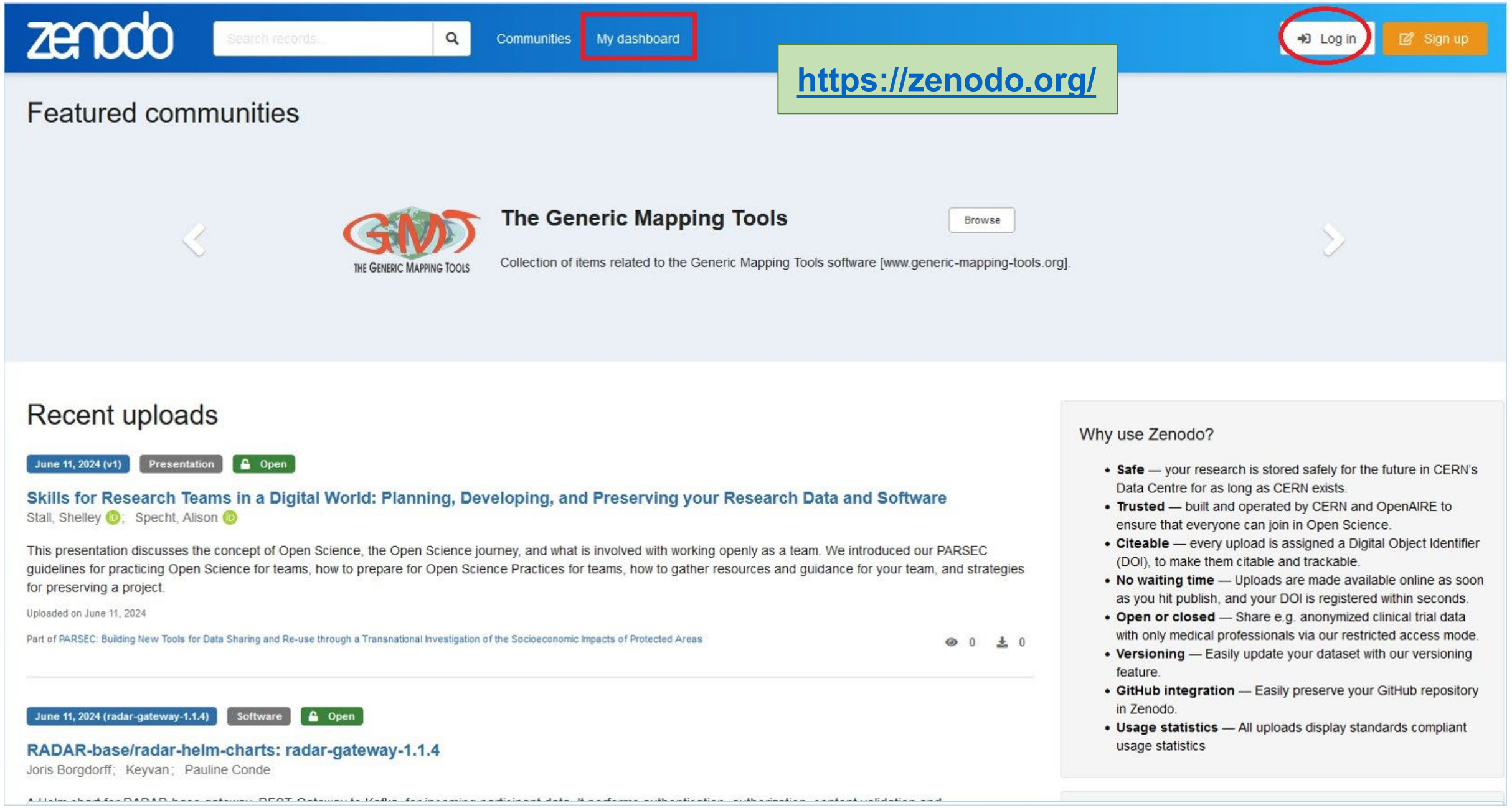
Nestor Seal – verification compliant with the German standard DIN 31644

CoreTrustSeal

- List of Current CoreTrustSeal certified data repositories
- fewer metrics, shorter duration, takes place every three years, off-site

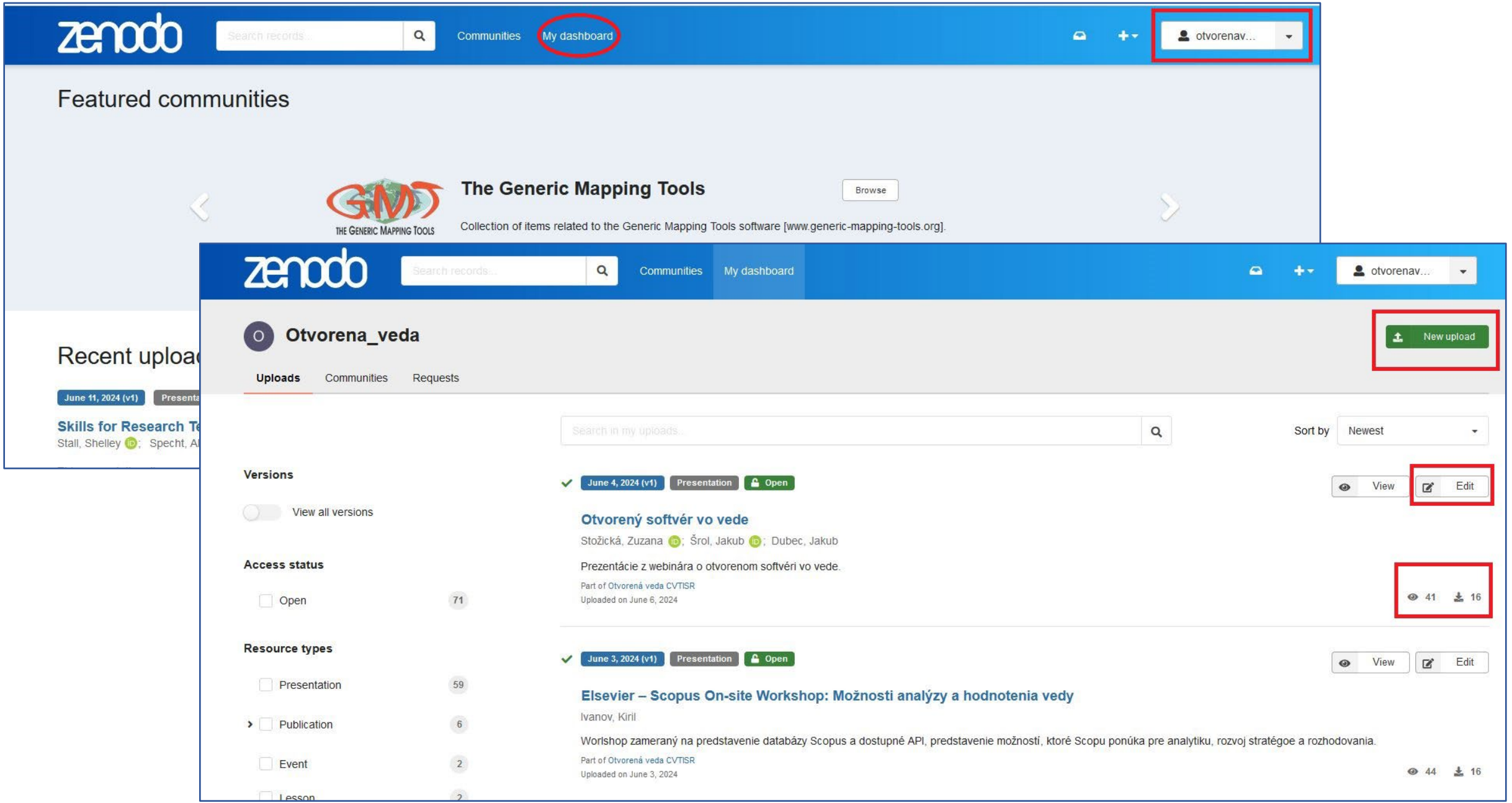
More information at: https://www.dpconline.org/handbook/institutional-strategies/audit-and-certification

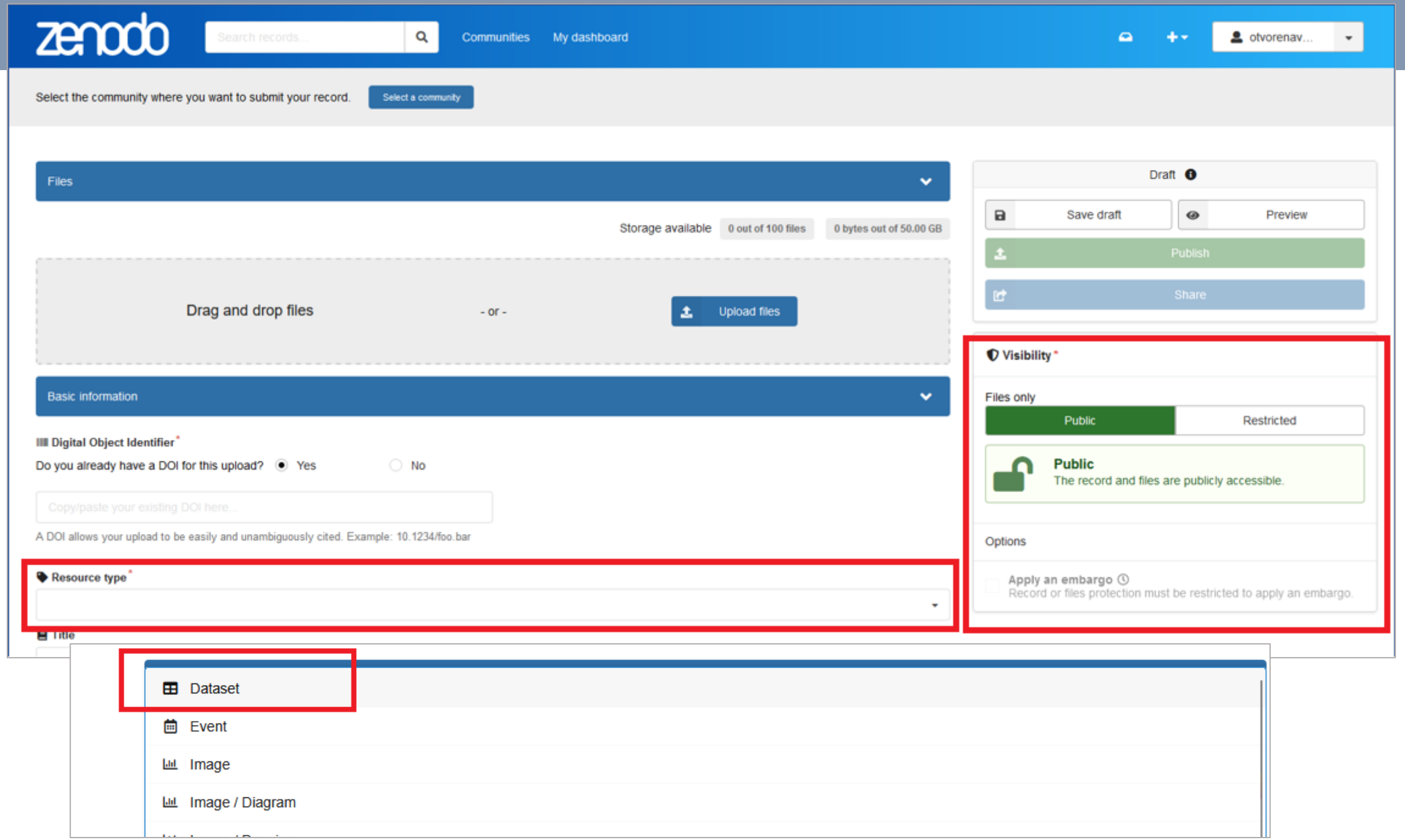
ZENODO


zenodo
Search records...
Communities
My dashboard
Log in
Sign up
https://zenodo.org/
Featured communities
The Generic Mapping Tools
Browse
Collection of items related to the Generic Mapping Tools software [www.generic-mapping-tools.org].
Recent uploads
June 11, 2024 (v1)
Presentation
Open
Skills for Research Teams in a Digital World: Planning, Developing, and Preserving your Research Data and Software
Stall, Shelley; Specht, Alison
This presentation discusses the concept of Open Science, the Open Science journey, and what is involved with working openly as a team. We introduced our PARSEC guidelines for practicing Open Science for teams, how to prepare for Open Science Practices for teams, how to gather resources and guidance for your team, and strategies for preserving a project.
Uploaded on June 11, 2024
Part of PARSEC: Building New Tools for Data Sharing and Re-use through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas
0
0
June 11, 2024 (radar-gateway-1.1.4)
Software
Open
RADAR-base/radar-helm-charts: radar-gateway-1.1.4
Joris Borgdorff; Keyvan; Pauline Conde
Why use Zenodo?
Safe — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
Trusted — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.
Citeable — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.
No waiting time — Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds.
Open or closed — Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode.
Versioning — Easily update your dataset with our versioning feature.
GitHub integration — Easily preserve your GitHub repository in Zenodo.
Usage statistics — All uploads display standards compliant usage statistics

ZENODO

https://zenodo.org/




zenodo
Search records...
Communities
My dashboard
otvorenav...
Select the community where you want to submit your record.
Select a community
Files
Storage available
0 out of 100 files
0 bytes out of 50.00 GB
Drag and drop files
- or -
Upload files
Basic information
Digital Object Identifier *
Do you already have a DOI for this upload?
Yes
No
Copy/paste your existing DOI here...
A DOI allows your upload to be easily and unambiguously cited. Example: 10.1234/foo.bar
Resource type *
Title
Draft
Save draft
Preview
Publish
Share
Visibility *
Files only
Public
Restricted
Public
The record and files are publicly accessible.
Options
Apply an embargo
Record or files protection must be restricted to apply an embargo.
Dataset
Event
Image
Image / Diagram

Archiving & Repositories

WHAT TO REMEMBER

- Archiving is essential for continuity of scientific record
- Arching and backup are not the same
- Look for trustworthy repositories
- Ask your colleagues or librariens

REFERENCES & RECOMMENDED SOURCES

List of references:

1.Mertl T. Dlouhodobé uchovávání elektronických dokumentů. 2015. https://is.ambis.cz/th/x6xcx/DP_dlouhodobe_uchovavani_elektronickych_dokumentu.pdf

2.Michaela Kmeťová: DLHODOBÁ ARCHIVÁCIA DIGITÁLNYCH OBJEKTOV V INŠTITUCIONÁLNYCH REPOZITÁROCH. ZBORNÍK FILOZOFICKEJ FAKULTY UNIVERZITY KOMENSKÉHO, Ročník XXVII KNIŽNIČNÁ A INFORMAČNÁ VEDA Bratislava 2017. Dostupné online na https://fphil.uniba.sk/fileadmin/fif/katedry_pracoviska/kkiv/Publikacie/KaIV/KIV27_100.pdf

3.Milena Tetřevová Maťašovská: Čas zmien v komunikácii vedeckých poznatkov. Dostupné na: https://itlib.cvtisr.sk/%c4%8cl%c3%a1nky/clanek1757/

4.Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report: *https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf*

Recommended sources for further study:

1. **EOSC CZ – training and webinars (in Czech and English)**
2. **NTK –Research Data Management Guide (in Czech and English)**
3. **RDMkit (in English)**
4. **OpenAIRE (in English)**
5. **otvorenaveda.cvtisr.sk/ (in Slovak)**
6. https://en.wikipedia.org/wiki/Research_data_archiving
7. Diego Menchaca. (2019). *6 repositories to share your research data*. https://www.teamscopeapp.com/blog/6-repositories-to-share-your-research-data
8. ASTERA. *Data Repository: Importance, Challenges, and Best Practices*. https://www.astera.com/type/blog/data-repository/
9. Metadata repositories**:** https://en.wikipedia.org/wiki/Metadata_repository
10. https://www.dataversity.net/what-is-a-metadata-repository/
11. https://www.dataversity.net/metadata-repository-basics-from-database-to-data-architecture/
12. File formats and standards: https://www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards
13. Digital preservation: https://ufs.libguides.com/c.php?g=1113411&p=8118652
14. E. P. Mclellan. Selecting formats for Digital Preservation. Lessons Learned during the Archivematica Project: https://www.niso.org/sites/default/files/stories/2019-11/IP_McLellan_Selecting_Formats_isqv22no2.pdf.
15. https://link.springer.com/referenceworkentry/10.1007/978-0-387-39940-9_909
16. EC Guide to HE programme: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf , Repositories:page 48, 49

Images – sources if not given directly in the slide with an image:

Slide 3: 1) Scrolls: Photo by Wilhelm Gunkel on Unsplash // 2) Archive Shelfs at Sächsisches Staatsarchiv in Dresden, Saxony, Germany: Photo by C M on Unsplash // 3) Computer with a code. Photo by Markus Spiske on Unsplash

Slide 9: Free computer server room image, public domain CC0 photo. Autor: rawpixel.com | Zdroj: rawpixel.com

Thank you.