Univerza v Mariboru

Computer Systems
Jožef Stefan Institute

How to address Open science in EU projects?

Prof Barbara Koroušić Seljak

Computer Systems Department

Jožef Stefan Institute, Ljubljana

Maribor, Slovenia,
12.–16. September 2022

Open SCIENCE Summer School

Outline

- Data-driven science in EU-funded projects
- What is data & meta data
- How to prepare data to be “deep”
- How to prepare data to be shareable and reusable (open)
- Open data strategies
- Researchers’ perception of open data
- Discussion

European projects

- **The main aim of EU-funded projects:**
  - To join forces of research groups from different disciplines and different geographical regions in **collecting and using various data** in a cooperative way.

Why data is so important?

- Science collects data in order to find answers to open research questions.
- **Wider importance of data:**
  - The PubMed Central statistics shows that one fourth of the daily users are academics, 40% are citizens, more than 15% are employed by companies, and the rest are policy makers and others.
  - Research and development (R&D) can contribute to productivity, however, R&D-intensive companies still report difficulties in accessing and re-using research data.

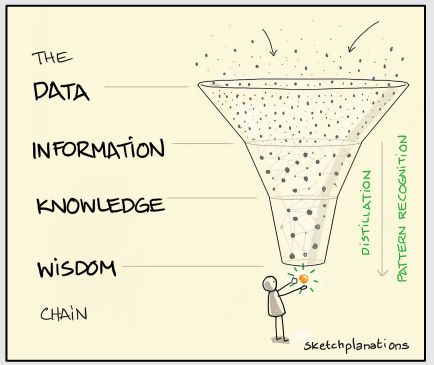
To share, or not to share data

- Research is expensive and it is almost always impossible to collect all the required data in one study.
- Solution:
  - Data sharing between researchers and re-using of existing data
- Fundamental question:
  - Shall we focus on collecting and sharing lots of data?
    - Big (large-scale) data does not necessarily mean that the data is valuable.
- Answer:
  - To be valuable, collected and shared data needs to be “deep”, meaning it distinguishes with **high quality and integrity**.

Data sharing

- Data sharing can accelerate scientific discovery while increasing return on investment beyond the researcher or group that produced them.
- Fundamental questions:
  - Technical questions:
    1. How to achieve high quality and integrity of data?
    2. How to share data?
  - Legal and ethical questions (out of the scope of this lecture)
  - What are the scientists' perception of data repositories and their perspectives on data management and sharing practices?

First, let's define data



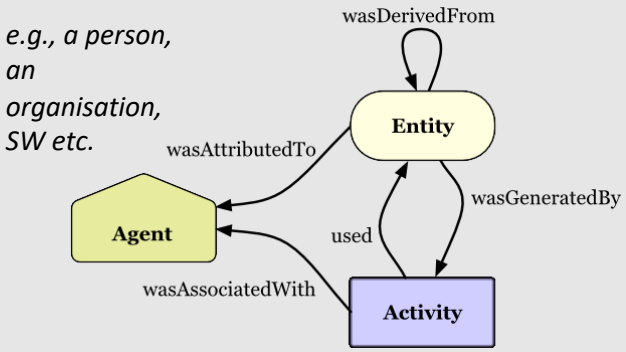
https://sketchplanations.com/dikw

- **Data** are factual records
- Data has no value without an information about its context and origin – needs to be enriched with meta data:
  - Example of data: F, M, … – what is the meaning of these data?
  - Could be: Sex/ gender (F for "Female", M for "Male") / Unit of measurement (F for "Fahrenheit", M for "Meter")
  - Once data is put into context, it becomes **information**.
  - As soon as we collect pieces of information about a particular fact or a person, we can start talking about **knowledge**. Without knowledge we cannot make informed decisions neither solve problems.

Meta data = Information about the context and the origin of data

- **Data provenance (origin):**
  - Describes entities, people and processes involved in the data production

e.g., a web page, a chart etc.



e.g., a person, an organisation, SW etc.

e.g., the 2nd version of data was generated by a translation from the 1st version of data in another language

- **Other meta data (context):**
  - Describes all other parameters of data
  - **Annotated data:** data enriched with meta data defined with respect to a given controlled vocabulary or ontology
    - Those meta data support the discovery of interesting datasets that can be linked using the meta data.

TQ1. How to achieve high quality and integrity of data?

- Data and its meta data are as complete as possible.
- Data is completely accurate – don' forget, data curation requires a lot of efforts and time!
- Data is consistent:
  - The data is inconsistent if we have collected more information on some issues than on others.
- Data is immediately actionable:
  - This depends on our ability to turn data into action quickly. If we have multiple types of systems within our infrastructure, each of them generating data in different ways, the data needs to be normalised before analysing it.

TQ1. How to achieve high quality and integrity of data? Following the FAIR principles

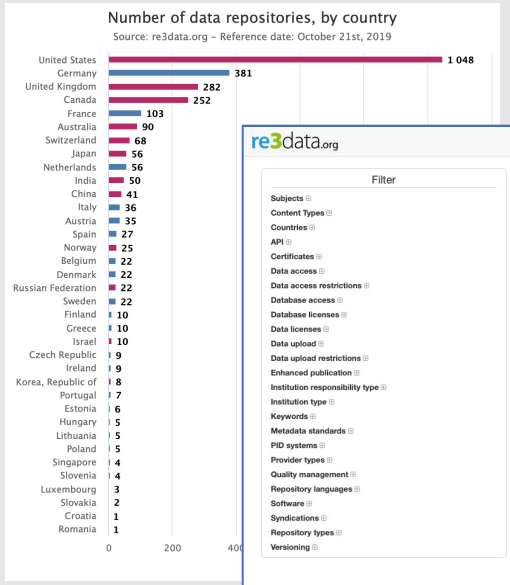
- **Findable**: Easy to find by both human experts and information systems and based on mandatory description of the meta data;
- **Accessible:** Stored for long term such that they can be easily accessed and/or downloaded with well-defined license and access conditions (Open Access when possible), whether at the level of meta data, or at the level of the actual data content;
- **Interoperable:** Ready to be combined with other datasets by humans as well as information systems;
- **Reusable:** Ready to be used for future research and to be processed further using computational methods.

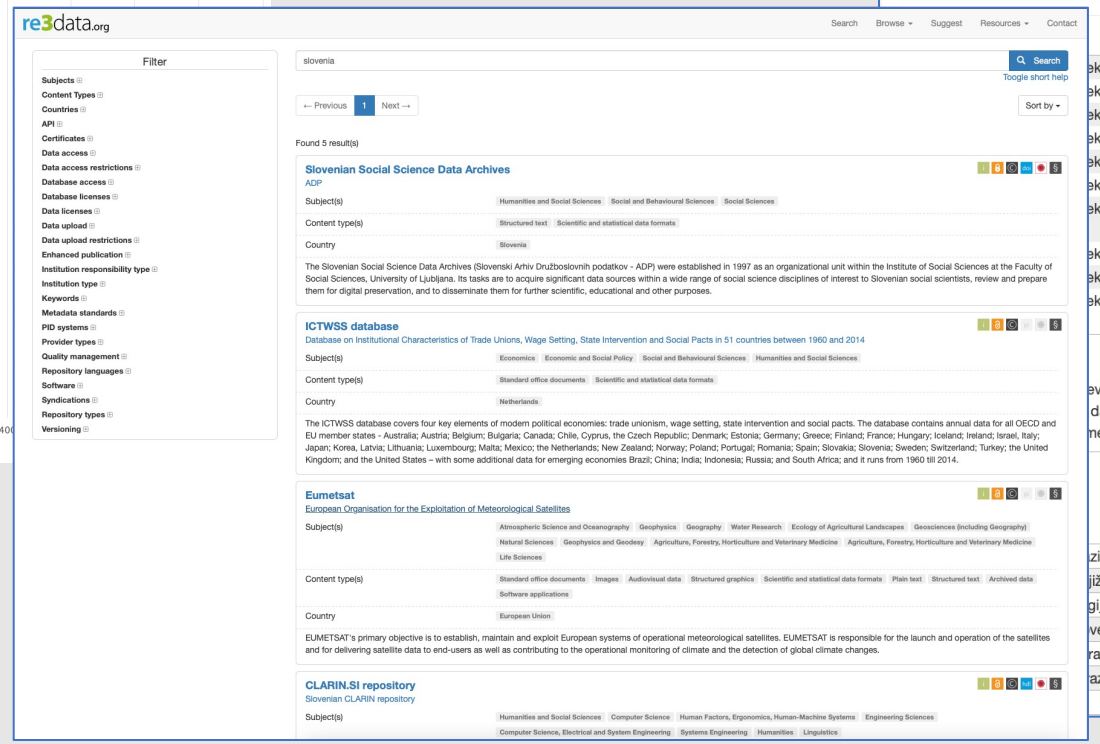
- Mandatory in EU projects!
- Fairsfair.eu

TQ2. How to share open data?

- **Using trusted open data repositories:**
  - To enable data sharing and preservation over the long term.
  - Domain-specific data repositories
  - Zenodo (DOI – Digital Object Identifier, connectivity with GitHub)
- **Following Open science strategies & policies:**
  - Open science is more than open access to **publications** or **data**.
  - Open science is a broader concept that includes
    - the interoperability of infrastructures,
    - open and shared research methodologies and tools (e.g., for text and data mining).

Data repositories in Slovenia (re3data.org)





Univerza v Mariboru

Maribor, Slovenia,
12.–16. September 2022

Open Science Summer School

Computer Systems
Jožef Stefan Institute

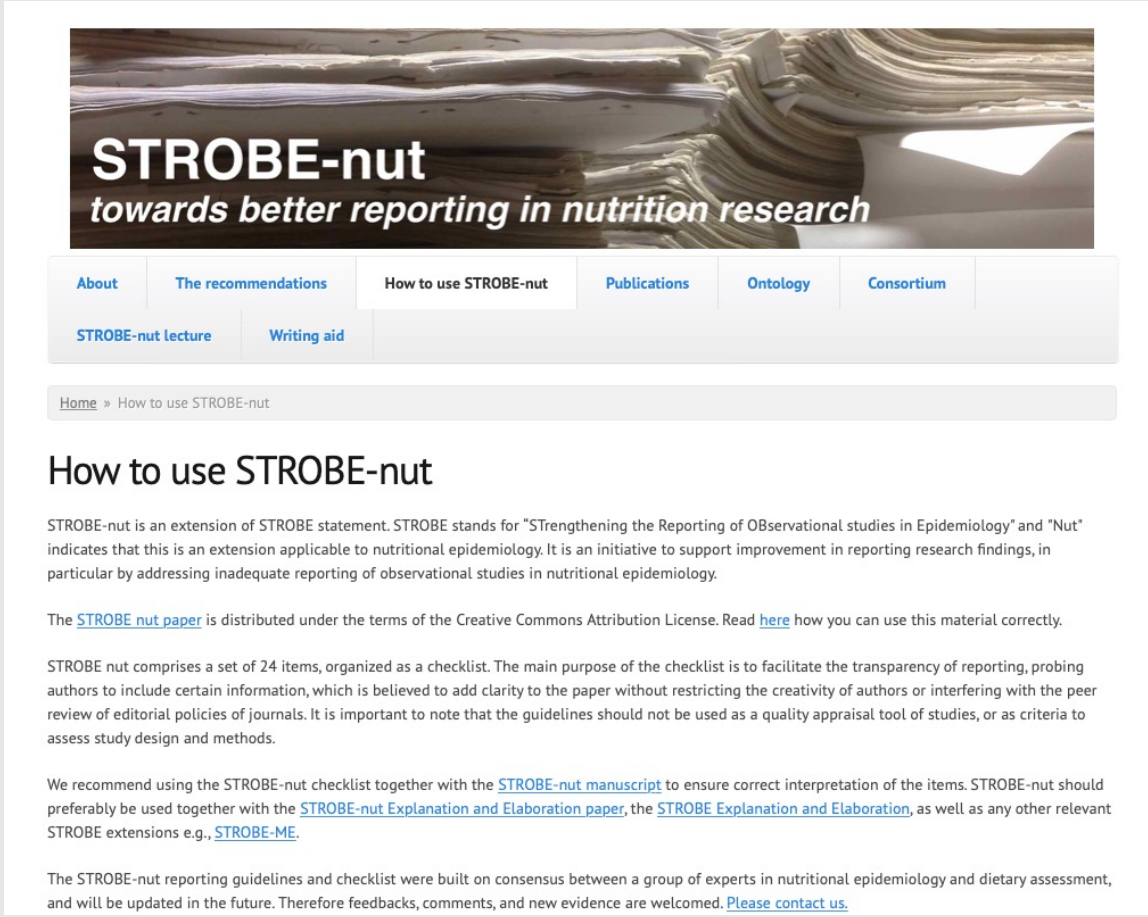
Open science strategies & policies

- The EU's Open science policy
- EOSC (European Open Science Cloud)
  - will enable researchers across disciplines and countries to store, curate and share data;
  - FNS Cloud project
- National (Slovenian)
  - Resolution on the Slovenian Scientific Research and Innovation Strategy 2030 - introduces Open science as an important integral part of the scientific research
  - UM – Open-access policy for RIs

How to prepare research data to comply with the Open science strategies?

- Start early - consider these aspects from the very beginning of your study
  - Talk to more experienced colleagues, read the project's Data Management Plan
  - Check your institution's and the project's internal policies/ rules (the Grant and Consortium Agreements)
  - If possible, join a training workshop (e.g., OPSI - https://podatki.gov.si)
- Carefully design data provenance and meta data models
  - Select standard schemes for defining data provenance and meta data, e.g.,
    - Dublin Core Metadata Initiative (general)
    - FAIRSharing.org (domain-specific)
    - Discuss this issue with other project partners!
- If more project partners contribute to the data collection, retain sufficient IPRs to comply with open access requirements.

Example for data and meta data preparation - STROBE-nut



Open-access scientific information

Open access: Practice of providing online access to scientific re-usable information that is free of charge to the end-user.

Scientific information:

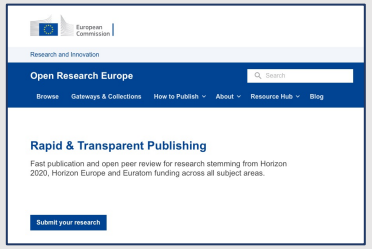
- Articles appearing in journals including online
- Book chapter(s) describing project's activities
- Software, source code, algorithms, dataset (as part of a publication or independently) including collaborative workspaces accessible by organisations that are not the project beneficiaries (recommended)
- Critical reviews or other publications appearing in peer-reviewed journals including online
- Doctoral and Master theses
- Patents
- Press releases and other media activities including television, radio and online (not including social media)
- Scientific publications prepared using data or other results derived from the project
- Technical reports (confidential and non-confidential)
- Webpages or other online publications
- Written or oral/ poster abstracts/ presentations given at conferences, symposia or meetings at the local, national, regional or international level

Important to know

- Open access still requires peer review academic process!
- You can still protect (pend a patent) before publishing scientific information as open access.

Routes to open-access scientific information

- Self-archiving / 'green' open access:
  - authors or a representative archive (deposits) published scientific information in an online repository before, at the same time as, or after publication.
  - Some publishers request that open access be granted only after an embargo period has elapsed!
- Open access publishing / 'gold' open access:
  - Scientific information is immediately published in an open-access mode.
- Publication costs:
  - Fees are reimbursable if publishing venue is full open access only (hybrid ones are excluded).
  - No cost for publishing in Open Research Europe.



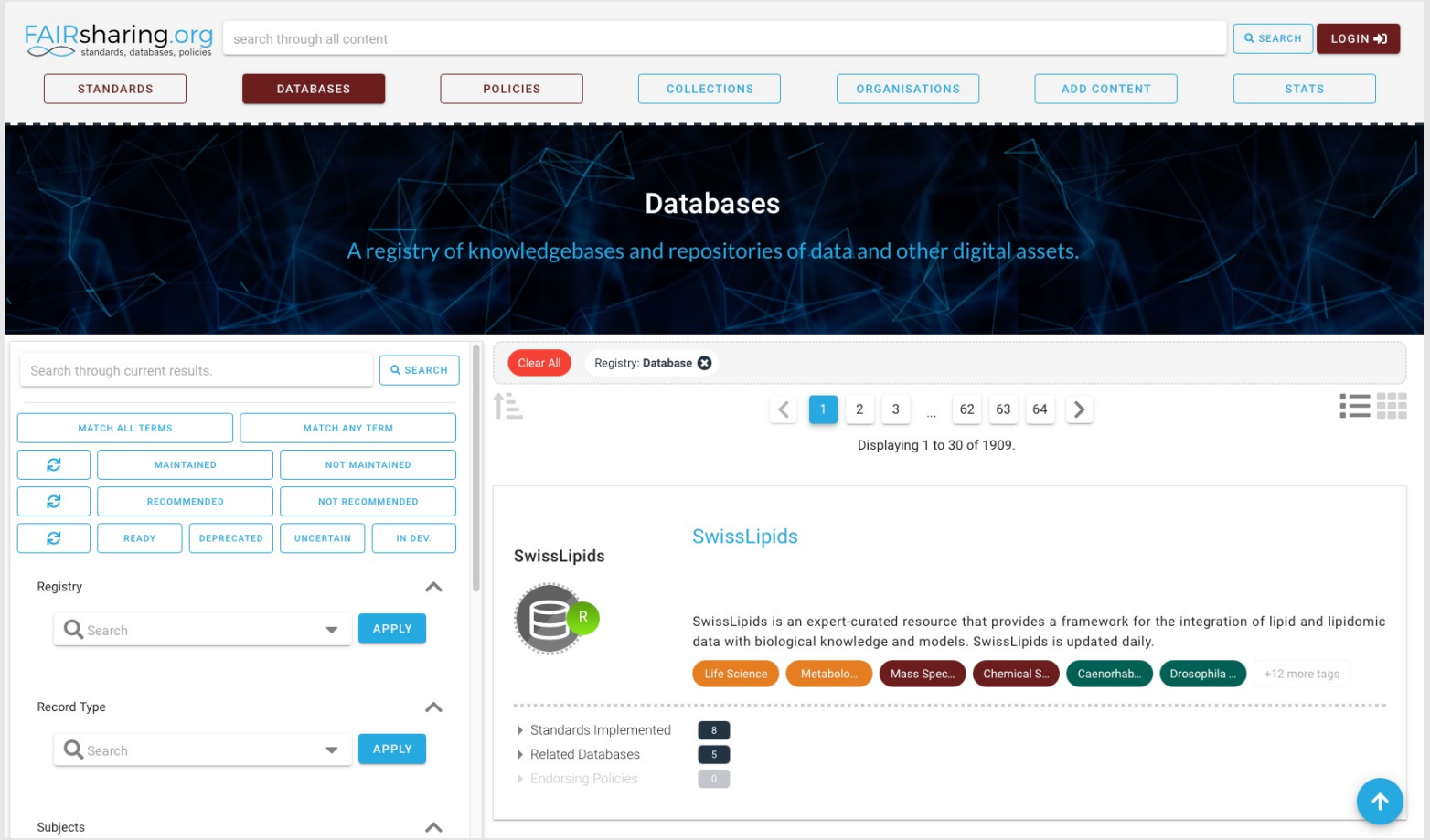
Don’t forget!

- In EU-funded projects projects, typically a **prior notice** of any planned open-access scientific information needs to be given!
- There is a contractual obligation to **acknowledge EU funding** in any open-access publication.
- Where possible the following **disclaimer** should also be included, i.e.,
  - *Information and views set out across this < publication type > are those of the Consortium and do not necessarily reflect the official opinion or position of the European Union. Neither European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use that may be made of the information contained herein.*

Check list (for each open-access publication)

- ❑ Drafted publication
- ❑ Agreed authorship
- ❑ Established notification period (the number of calendar days as agreed within the project)
- ❑ Completed notification form
- ❑ Informed Coordinator and Dissemination WP Leader
- ❑ Received comments
- ❑ Revised publication (if necessary)
- ❑ Submitted publication
- ❑ Determine date of publication
- ❑ Informed Coordinator and Dissemination WP Leader
- ❑ Publication is in the public domain
- ❑ Completed reporting
- ❑ Informed Coordinator and Dissemination WP Leader

FAIRsharing.org



Univerza v Mariboru

Maribor, Slovenia,
12.–16. September 2022

Open SCIENCE Summer School

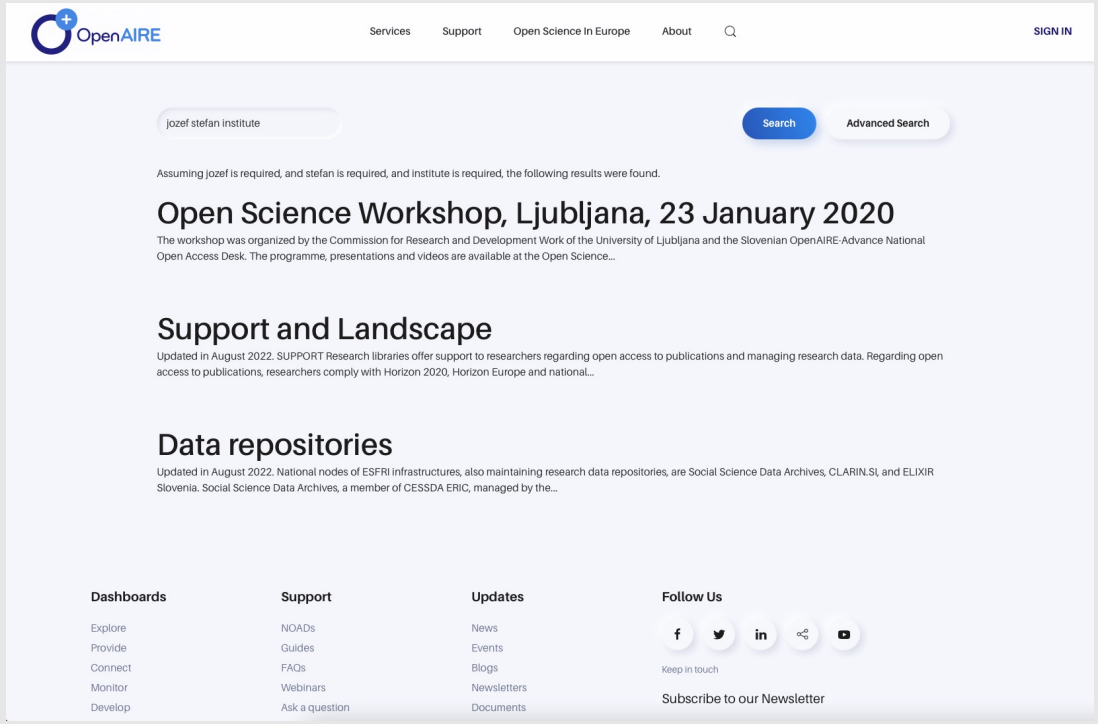
Computer Systems
Jožef Stefan Institute

How to store open research data?

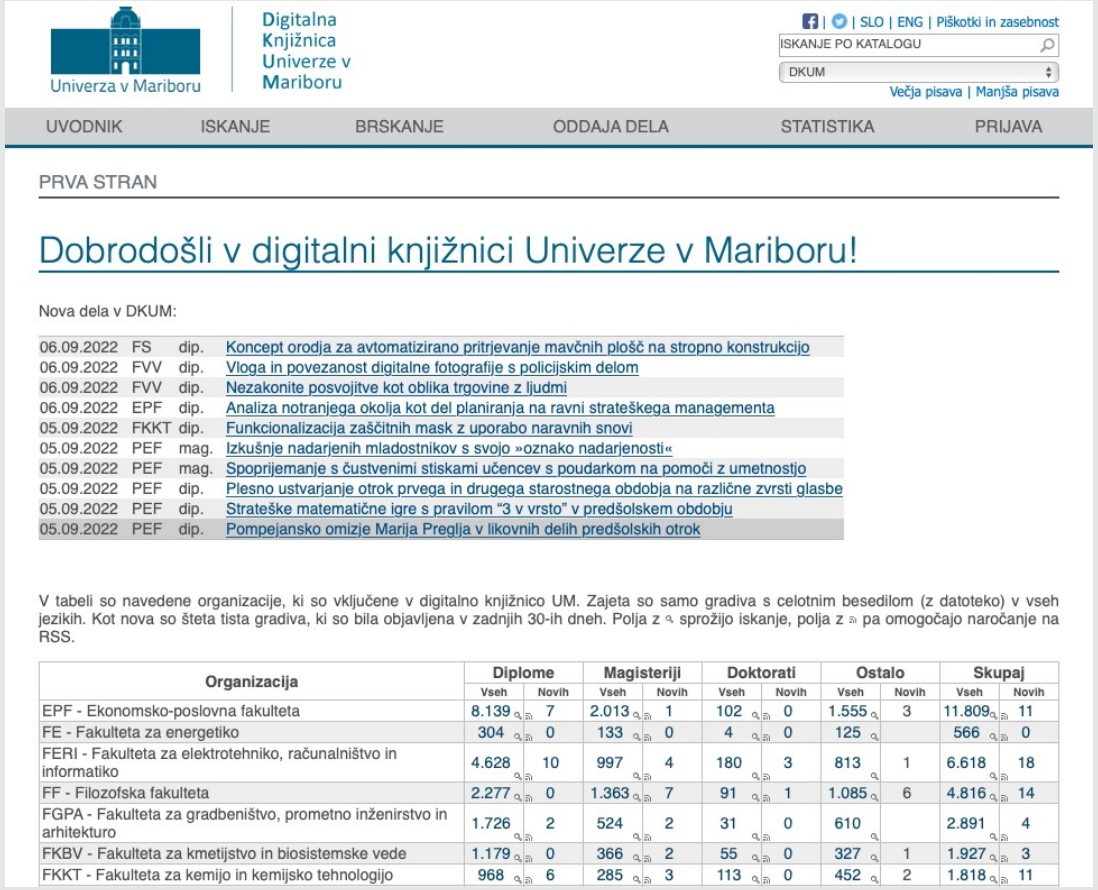
- Check the FAIRness of your research data:
  - There exist various FAIR data assessment tools.
  - Check with the project partners

How to store open research data?

- Example (IJS):
  - Inform your institutional library about new dataset or publication, providing raw data and meta data
  - An information about new data is stored in Cobiss/ SICRIS
  - The meta data is uploaded to a Slovenian OpenAIRE compatible repository
  - Follow the project's rules



Examples of Slovenian OpenAIRE compatible repositories

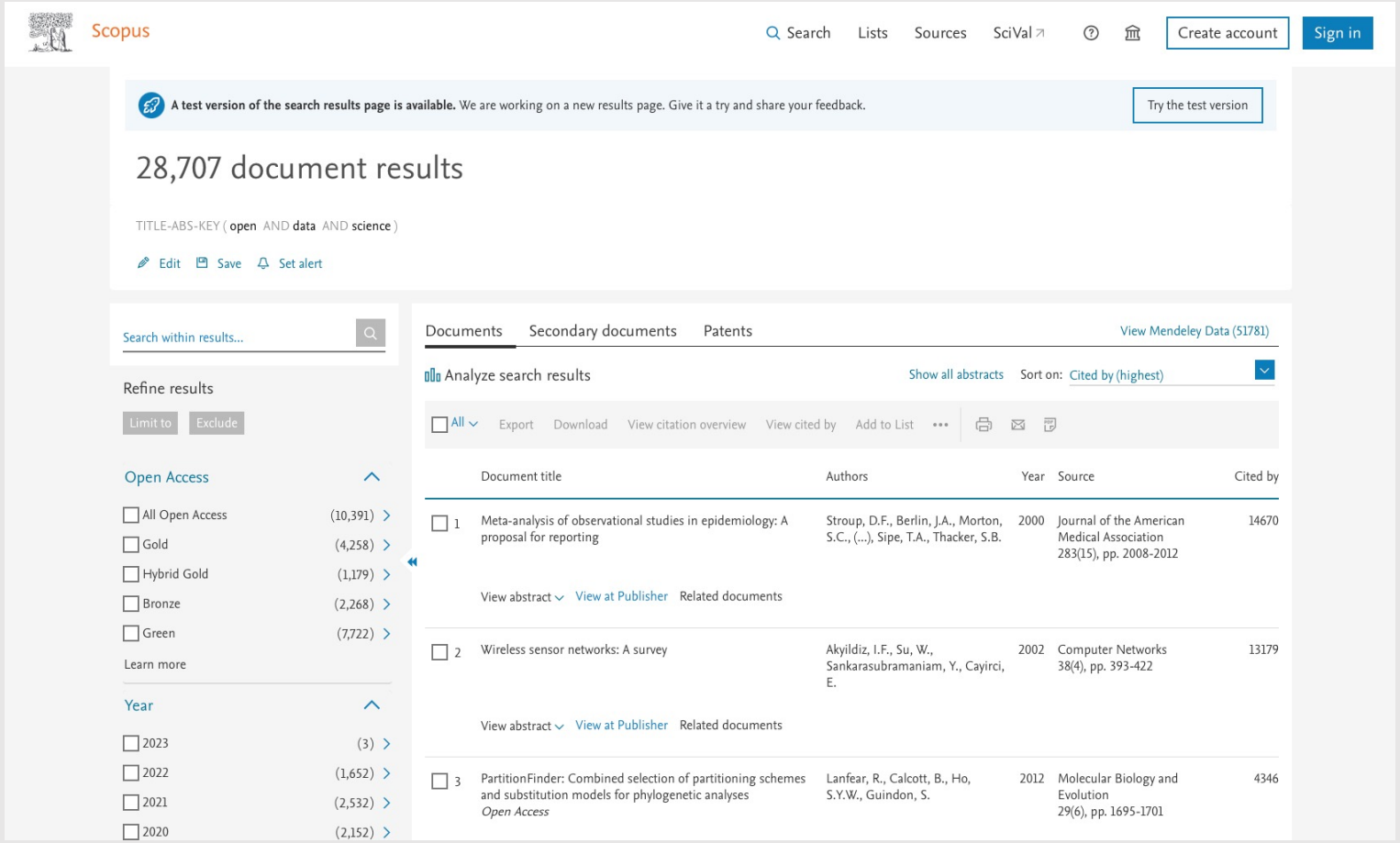


Univerza v Mariboru

Maribor, Slovenia,
12.–16. September 2022
Open SCIENCE Summer School

Computer Systems
Jožef Stefan Institute

Scopus (Elsevier) - meta data of scholarly publications



Univerza v Mariboru

Maribor, Slovenia,
12.–16. September 2022

Open SCIENCE Summer School

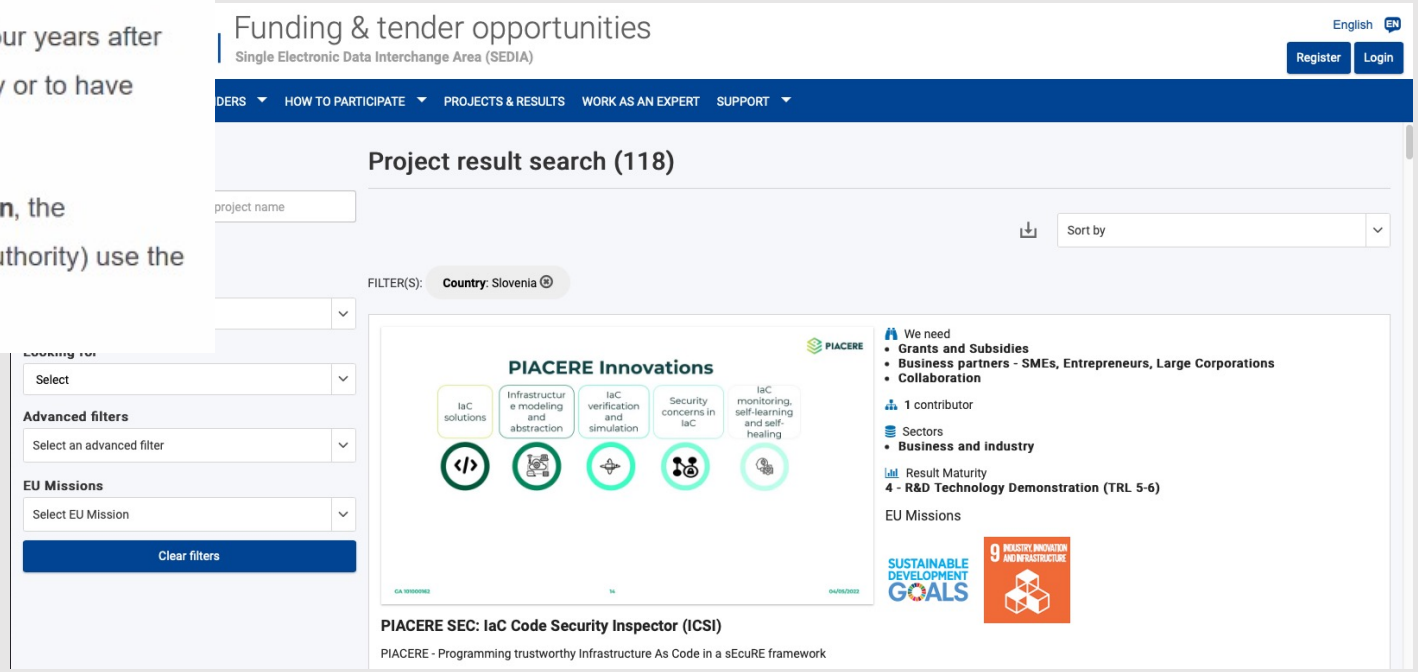
Computer Systems
Jožef Stefan Institute

Horizon Results Platform – data exploitation

Follow up of results after the end of the project

Through the reporting tools

- **Beneficiaries** which have received funding under the grant must —up to four years after the end of the action — use their best efforts to exploit their results directly or to have them exploited indirectly by another entity
- If the results are not exploited **within one year after the end of the action**, the beneficiaries must (unless otherwise agreed in writing with the granting authority) use the Horizon Results Platform to find interested parties to exploit the results.



How to address Open science while preparing a European project proposal?

- Assessment of Open science practices through the **Excellence** award criteria for proposal evaluation.
  - **Previous experience** on Open science practices are evaluated positively.
- Be careful:
  - Specific Work Programmes may oblige to adhere to Open science practices such as **involvement of citizens**, or to use the **European Open Science Cloud**.
  - Mandatory Data Management Plan for FAIR (Findable, Accessible, Interoperable, Reusable) research data

How to give contribution?

"Furggelen afterglow" by Lukas Schlagenhauf is licensed under CC BY-ND 2.0.

creativecommons.org

Attribution-NoDerivs 2.0 Generic (CC BY-ND 2.0)

This is a human-readable summary of (and not a substitute for) the license. Disclaimer.

You are free to:

Share — copy and redistribute the material in any medium or format
for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NoDerivatives — If you remix, transform, or build upon the material, you may not distribute the modified material.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

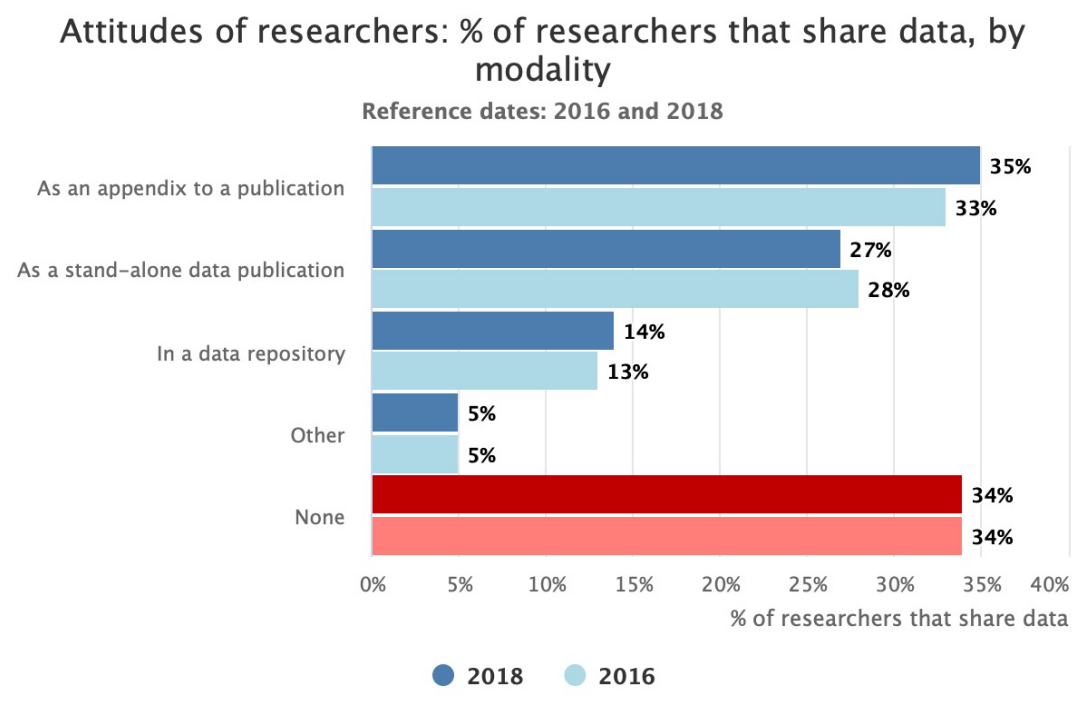
You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

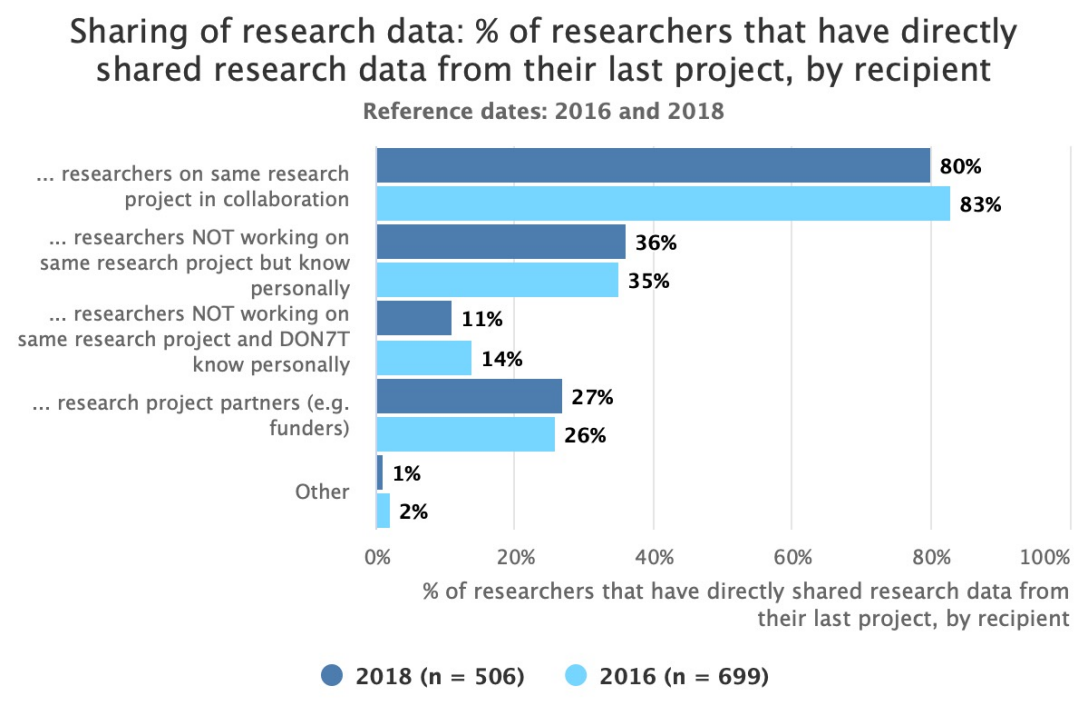
No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

Our perceptions of opening data

- Open science is mandatory in EU-funded projects!
- Discussion

Researchers' attitude towards sharing data





https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor/facts-and-figures-open-research-data_en#funderspolicies

Scientists' perceptions of open data

- Scientific Data, Nature (https://www.nature.com/articles/s41597-022-01428-w):
  - Using focus groups with scientists from five disciplines, questions about data management were discussed.
  - Participants identified **metadata quality control** and **training** as problem areas in data management.

Discussion

- My contacts:
  - Email address: barbara.korousic@ijs.si
  - Phone number: 01 4773-363