

# How to address **Open science** in EU projects?

Prof Barbara Koroušič Seljak  
Computer Systems Department  
Jožef Stefan Institute, Ljubljana

# 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

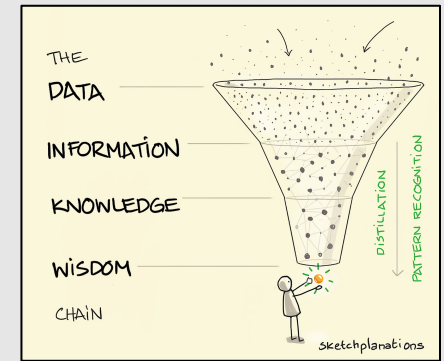
- Research is expensive and it is almost always impossible to collect all the required data in one study.
- Solution:
  - Data sharing between researchers and re-using of existing data
- Fundamental question:
  - Shall we focus on collecting and sharing lots of data?
    - Big (large-scale) data does 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

- **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.



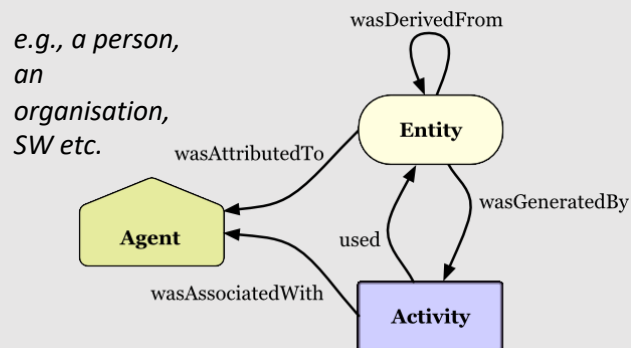
<https://sketchplanations.com/dikw>

# 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., the 2<sup>nd</sup> version of data was generated by a translation from the 1<sup>st</sup> 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.

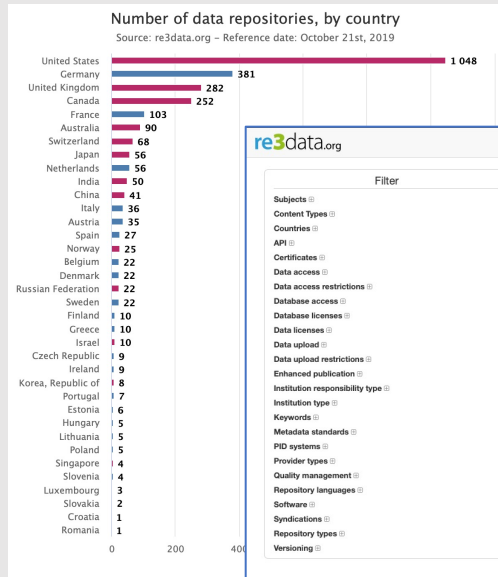
# TQ1. How to achieve high quality and integrity of data? Following the FAIR principles

- **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](https://www.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)



## Digitalni repozitorij raziskovalnih organizacij Slovenije

- Uvodnik
- Iskanje
- Brskanje
- Statistika
- Obvestila
- Kontakti
- Prijava

## Dobrodošli v repozitoriju DiRRROS!

A+ | A- | SLO | ENG

- The effect of dietary fibre content on skatole and indole production in faeces of immunocastrated male pigs
- Reduction of boar taint compounds in Kraski prsut through dry-curing process
- Carcass and meat quality traits of pig fatteners from Slovenian Breeding programme
- Influence of housing conditions on reliability of immunocastration and consequences for growth performance of male pigs
- Reduced use of nitrites and phosphates in dry-fermented sausages
- The physiological impact of GFLV virus infection on grapevine water status
- Comparative study of temporal changes in pigments and optical properties in sepals of *Helleborus odoratus* and *H. niger* from prebloom to seed production
- Effect of cluster thinning within the grapevine variety
- Resilience of dairy farms measured through production plan adjustments
- The effect of cultivation practices on agronomic performance, elemental composition and isotopic signature of spring oat (*Avena sativa* L.)

Število novih in skupno število gradiv po posameznih organizacijah, vključenih v DiRRROS. Zajeta so samo gradiva s celotnim datoteko, ne glede na jezik dokumenta. Kot nova so štetá gradiva, ki so bila objavljena v zadnjih 30-ih dneh. Polja z 0, sprožijo mentov posamezne organizacije, polja z 1 pa omogočajo naročanje na RSS.

Organizacija	Članki in drugi sestavni deli		Monografije in druga dela		Izvedena dela (dogodki)		Skupaj	
	Vseh	Novih	Vseh	Novih	Vseh	Novih	Vseh	Novih
raziskovalno dejavnost RS	3	0					3	0
Knjižnica Univerze v Ljubljani	340	0	8	0			348	0
Evropska unija	1	0					1	0
raziskovanja	45	4	1	0			46	4
raziskave	98	0	2	0			100	0

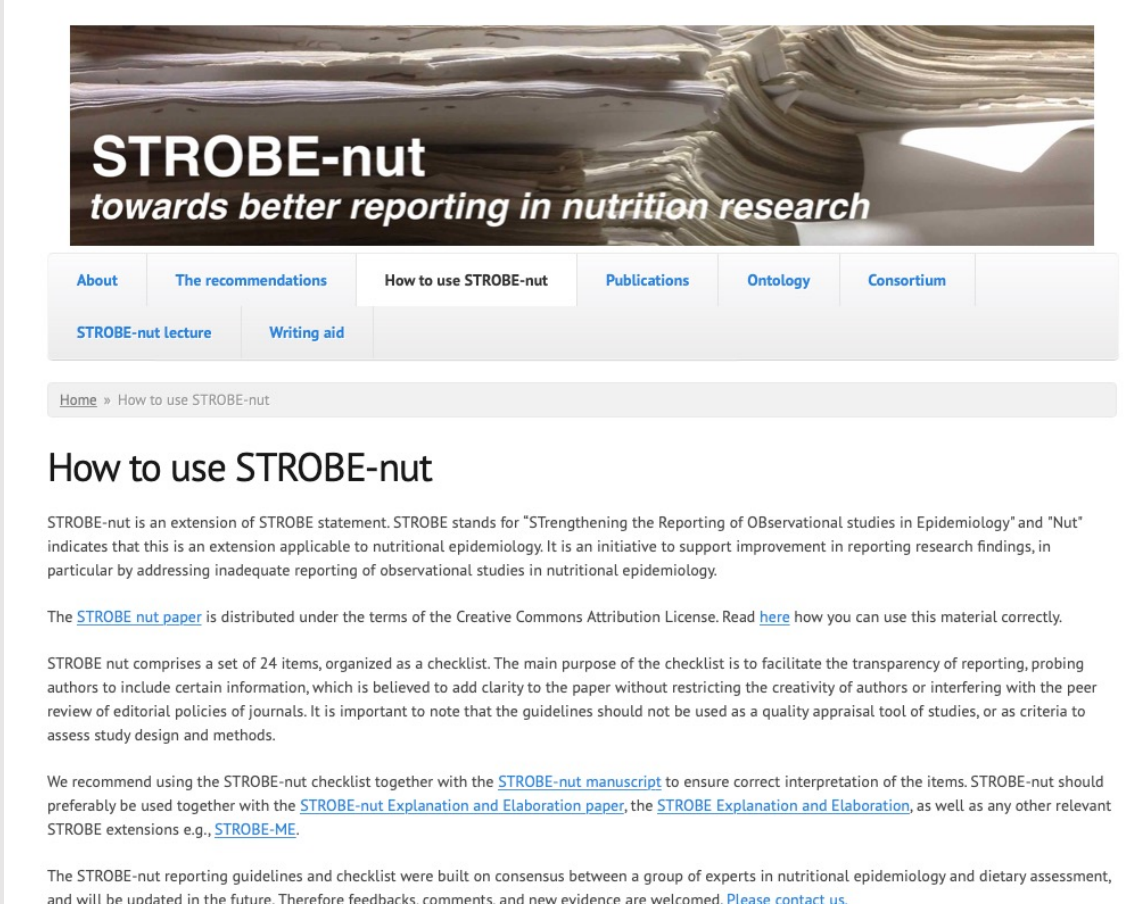
# Open science strategies & policies

- [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



**STROBE-nut**  
*towards better reporting in nutrition research*

[About](#) [The recommendations](#) [How to use STROBE-nut](#) [Publications](#) [Ontology](#) [Consortium](#)

[STROBE-nut lecture](#) [Writing aid](#)

[Home](#) » [How to use STROBE-nut](#)

## How to use STROBE-nut

STROBE-nut is an extension of STROBE statement. STROBE stands for "STrengthening the Reporting of OBservational studies in Epidemiology" and "Nut" indicates that this is an extension applicable to nutritional epidemiology. It is an initiative to support improvement in reporting research findings, in particular by addressing inadequate reporting of observational studies in nutritional epidemiology.

The [STROBE nut paper](#) is distributed under the terms of the Creative Commons Attribution License. Read [here](#) how you can use this material correctly.

STROBE nut comprises a set of 24 items, organized as a checklist. The main purpose of the checklist is to facilitate the transparency of reporting, probing authors to include certain information, which is believed to add clarity to the paper without restricting the creativity of authors or interfering with the peer review of editorial policies of journals. It is important to note that the guidelines should not be used as a quality appraisal tool of studies, or as criteria to assess study design and methods.

We recommend using the STROBE-nut checklist together with the [STROBE-nut manuscript](#) to ensure correct interpretation of the items. STROBE-nut should preferably be used together with the [STROBE-nut Explanation and Elaboration paper](#), the [STROBE Explanation and Elaboration](#), as well as any other relevant STROBE extensions e.g., [STROBE-ME](#).

The STROBE-nut reporting guidelines and checklist were built on consensus between a group of experts in nutritional epidemiology and dietary assessment, and will be updated in the future. Therefore feedbacks, comments, and new evidence are welcomed. [Please contact us](#).



# 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

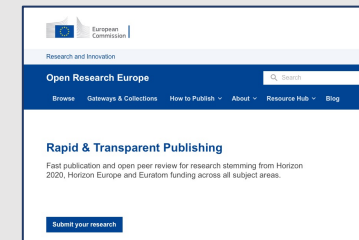


# Important to know

- 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

FAIRsharing.org standards, databases, policies search through all content SEARCH LOGIN

STANDARDS DATABASES POLICIES COLLECTIONS ORGANISATIONS ADD CONTENT STATS

## Databases

A registry of knowledgebases and repositories of data and other digital assets.

Search through current results. SEARCH

MATCH ALL TERMS MATCH ANY TERM

MAINTAINED NOT MAINTAINED

RECOMMENDED NOT RECOMMENDED

READY DEPRECATED UNCERTAIN IN DEV.

Registry

Search APPLY

Record Type

Search APPLY

Subjects

Clear All Registry: Database

1 2 3 ... 62 63 64

Displaying 1 to 30 of 1909.

### SwissLipids

SwissLipids is an expert-curated resource that provides a framework for the integration of lipid and lipidomic data with biological knowledge and models. SwissLipids is updated daily.

Life Science Metabolo... Mass Spec... Chemical S... Caenorhab... Drosophila ... +12 more tags

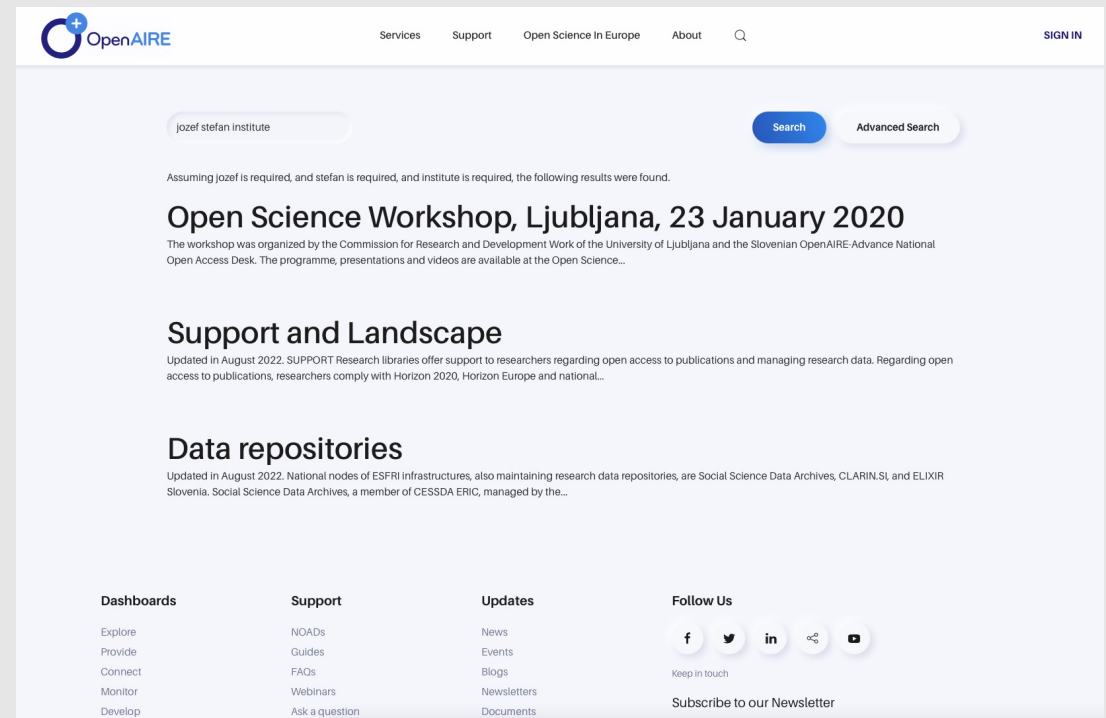
- Standards Implemented 8
- Related Databases 5
- Endorsing Policies 0

# 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



The screenshot shows the OpenAIRE website interface. At the top, there is a navigation bar with the OpenAIRE logo, links for Services, Support, Open Science In Europe, About, and a search icon, along with a SIGN IN button. Below the navigation bar is a search bar containing the text 'jozef stefan institute'. To the right of the search bar are buttons for 'Search' and 'Advanced Search'. The main content area displays search results for 'jozef stefan institute'. It starts with a note: 'Assuming jozef is required, and stefan is required, and institute is required, the following results were found.' The first result is 'Open Science Workshop, Ljubljana, 23 January 2020', with a brief description: 'The workshop was organized by the Commission for Research and Development Work of the University of Ljubljana and the Slovenian OpenAIRE-Advance National Open Access Desk. The programme, presentations and videos are available at the Open Science...'. Below this is a section titled 'Support and Landscape' with a note: 'Updated in August 2022. SUPPORT Research libraries offer support to researchers regarding open access to publications and managing research data. Regarding open access to publications, researchers comply with Horizon 2020, Horizon Europe and national...'. The next section is 'Data repositories' with a note: 'Updated in August 2022. National nodes of ESFRI infrastructures, also maintaining research data repositories, are Social Science Data Archives, CLARIN.SI, and ELIXIR Slovenia. Social Science Data Archives, a member of CESSDA ERIC, managed by the...'. At the bottom of the page, there are four columns of links: 'Dashboards' (Explore, Provide, Connect, Monitor, Develop), 'Support' (NOADs, Guides, FAQs, Webinars, Ask a question), 'Updates' (News, Events, Blogs, Newsletters, Documents), and 'Follow Us' (Facebook, Twitter, LinkedIn, RSS, YouTube). Below the social media icons is the text 'Keep in touch' and 'Subscribe to our Newsletter'.

# Examples of Slovenian OpenAIRE compatible repositories

PRVA STRAN

## Dobrodošli v digitalni knjižnici Univerze v Mariboru!

Nova dela v DKUM:

06.09.2022	FS	dip.	<a href="#">Koncept orodja za avtomatizirano pritrjevanje mavčnih plošč na stropno konstrukcijo</a>
06.09.2022	FVV	dip.	<a href="#">Vloga in povezanost digitalne fotografije s policijskim delom</a>
06.09.2022	FVV	dip.	<a href="#">Nezakonite posvojitve kot oblika trgovine z ljudmi</a>
06.09.2022	EPF	dip.	<a href="#">Analiza notranjega okolja kot del planiranja na ravni strateškega managementa</a>
05.09.2022	FKKT	dip.	<a href="#">Funkcionalizacija zaščitnih mask z uporabo naravnih snovi</a>
05.09.2022	PEF	mag.	<a href="#">Izkušnje nadarjenih mladostnikov s svojo »oznako nadarjenosti«</a>
05.09.2022	PEF	mag.	<a href="#">Spoprijemanje s čustvenimi stiskami učencev s poudarkom na pomoči z umetnostjo</a>
05.09.2022	PEF	dip.	<a href="#">Plesno ustvarjanje otrok prvega in drugega starostnega obdobja na različne zvrsti glasbe</a>
05.09.2022	PEF	dip.	<a href="#">Strateške matematične igre s pravilom "3 v vrsto" v predšolskem obdobju</a>
05.09.2022	PEF	dip.	<a href="#">Pompejsko omizje Marija Preglja v likovnih delih predšolskih otrok</a>

V tabeli so navedene organizacije, ki so vključene v digitalno knjižnico UM. Zajeta so samo gradiva s celotnim besedilom (z datoteko) v vseh jezikih. Kot nova so šteta tista gradiva, ki so bila objavljena v zadnjih 30-ih dneh. Polja z sprožijo iskanje, polja z pa omogočajo naročanje na RSS.

Organizacija	Diplome		Magisteriji		Doktorati		Ostalo		Skupaj	
	Vseh	Novih	Vseh	Novih	Vseh	Novih	Vseh	Novih	Vseh	Novih
EPF - Ekonomsko-poslovna fakulteta	8.139	7	2.013	1	102	0	1.555	3	11.809	11
FE - Fakulteta za energetiko	304	0	133	0	4	0	125		566	0
FERI - Fakulteta za elektrotehniko, računalništvo in informatiko	4.628	10	997	4	180	3	813	1	6.618	18
FF - Filozofska fakulteta	2.277	0	1.363	7	91	1	1.085	6	4.816	14
FGPA - Fakulteta za gradbeništvo, prometno inženirstvo in arhitekturo	1.726	2	524	2	31	0	610		2.891	4
FKBV - Fakulteta za kmetijstvo in biosistemske vede	1.179	0	366	2	55	0	327	1	1.927	3
FKKT - Fakulteta za kemijo in kemijsko tehnologijo	968	6	285	3	113	0	452	2	1.818	11



# Scopus (Elsevier) - meta data of scholarly publications

Scopus

Search Lists Sources SciVal ? Create account Sign in

A test version of the search results page is available. We are working on a new results page. Give it a try and share your feedback. Try the test version

28,707 document results

TITLE-ABS-KEY (open AND data AND science)

Edit Save Set alert

Search within results...

Refine results

Limit to Exclude

Open Access

- All Open Access (10,391)
- Gold (4,258)
- Hybrid Gold (1,179)
- Bronze (2,268)
- Green (7,722)

Learn more

Year

- 2023 (3)
- 2022 (1,652)
- 2021 (2,532)
- 2020 (2,152)

Documents Secondary documents Patents View Mendeley Data (51781)

Analyze search results Show all abstracts Sort on: Cited by (highest)

All Export Download View citation overview View cited by Add to List

	Document title	Authors	Year	Source	Cited by
1	Meta-analysis of observational studies in epidemiology: A proposal for reporting	Stroup, D.F., Berlin, J.A., Morton, S.C., (...), Sipe, T.A., Thacker, S.B.	2000	Journal of the American Medical Association 283(15), pp. 2008-2012	14670
	View abstract View at Publisher Related documents				
2	Wireless sensor networks: A survey	Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.	2002	Computer Networks 38(4), pp. 393-422	13179
	View abstract View at Publisher Related documents				
3	PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses	Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S.	2012	Molecular Biology and Evolution 29(6), pp. 1695-1701	4346
	Open Access				

# 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.

The screenshot displays the 'Funding & tender opportunities' section of the Single Electronic Data Interchange Area (SEDIA). The page title is 'Project result search (118)'. A search bar for 'project name' is visible. A filter is applied for 'Country: Slovenia'. The main content area features a graphic for 'PIACERE Innovations' with five categories: 'IaC solutions', 'Infrastructure modeling and abstraction', 'IaC verification and simulation', 'Security concerns in IaC', and 'IaC monitoring, self-learning and self-healing'. To the right, a list of details includes: 'We need' (Grants and Subsidies, Business partners - SMEs, Entrepreneurs, Large Corporations, Collaboration), '1 contributor', 'Sectors' (Business and industry), 'Result Maturity' (4 - R&D Technology Demonstration (TRL 5-6)), and 'EU Missions' (Sustainable Development Goals 9: Industry, Innovation and Infrastructure). The bottom of the graphic is titled 'PIACERE SEC: IaC Code Security Inspector (ICSI)' with the subtitle 'PIACERE - Programming trustworthy Infrastructure As Code in a sEcuRE framework'.

# 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?



“Furgelen 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](#).

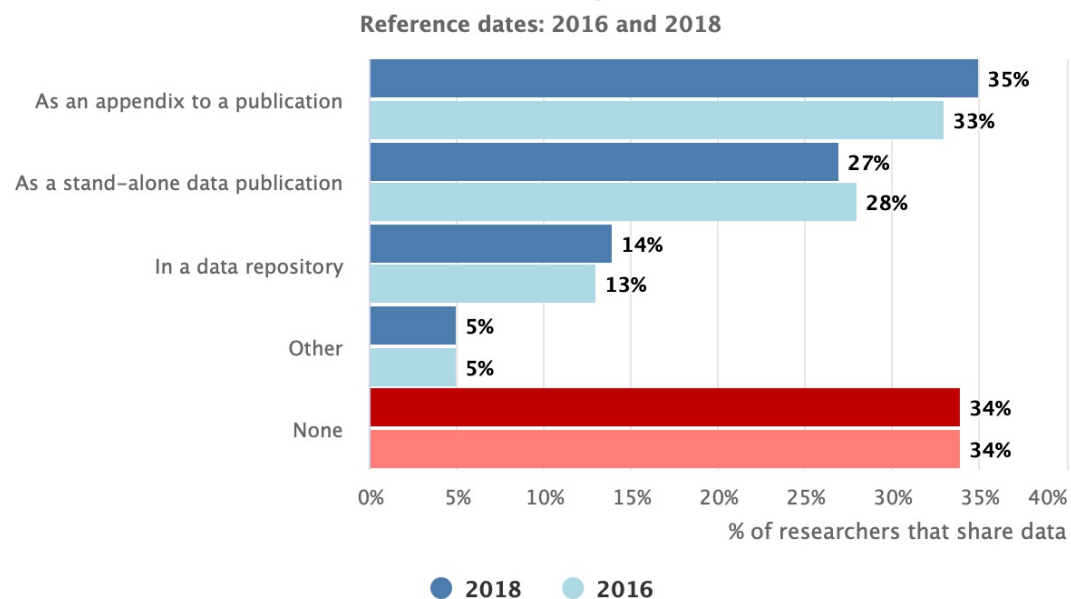
No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as [publicity, privacy, or moral rights](#) may limit how you use the material.

# Our perceptions of opening data

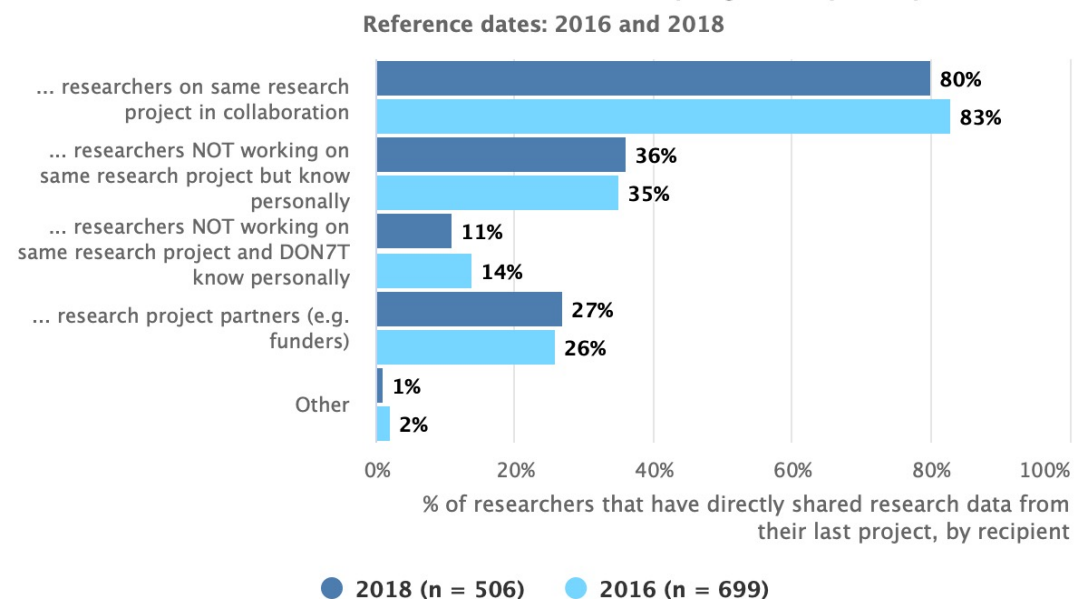
- Open science is mandatory in EU-funded projects!
- Discussion

# Researchers' attitude towards sharing data

Attitudes of researchers: % of researchers that share data, by modality



Sharing of research data: % of researchers that have directly shared research data from their last project, by recipient



[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](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](mailto:barbara.korousic@ijs.si)
  - Phone number: 01 4773-363