



ADP - SOCIAL SCIENCE DATA ARCHIVES

Analyze data! Deposit study! Promote science!

Challenges of Open Science and Open Data

Vipavc Brvar Irena

Invited lecture at the course New Technologies and Social Science Data
Collection , Faculty of Social Sciences, 11. 1. 2023



Slovenian Social Science Data Archives (*ADP-Arhiv družboslovnih podatkov*)



- Founded in 1997
- Slovenian **national research data centre** for social sciences
- **Member of CESSDA** ERIC since 2017
- Status of a **trust-worthy archive** (CoreTrustSeal since 2018)
- involved in EU and national projects



CESSDA

- Consortium of European Social Science Data Archives (Service Providers).
- **European Research Infrastructure (ERIC)**
- 22 member and 1 observer countries

Consortia of universities departments
Departments within universities
Public Research bodies
Consortia along with SSH ERICs of a country

MISSION

- to provide a sustainable research infrastructure that enables the research community to conduct high-quality research in the social sciences
- to contribute to effective solutions to the major challenges facing society today.

● Members (22) / Observers (1)
● Partners (12)



1. Austria
2. Belgium
3. Croatia
4. Czech Republic
5. Denmark
6. Finland
7. France
8. Germany
9. Greece
10. Hungary
11. Iceland
12. Ireland
13. North Macedonia
14. Netherlands
15. Norway
16. Portugal
17. Serbia
18. Slovak Republic
19. Slovenia
20. Sweden
21. Switzerland (Observer)
22. UK
23. Italy

Countries aiming at membership:

- Bosnia and Herzegovina
- Bulgaria
- Estonia
- Lithuania
- Poland
- Romania
- Spain

Open Science Game: Open Up Your Research



OPEN UP YOUR RESEARCH

With this game, you follow Emma on her way to her PhD and decide for her to either practice science the traditional way or to follow a more open approach. While this game is intended to make researchers aware of the Open Science practices that could be applied in one's research workflow, not all of these practices might be equally suitable for all disciplines. What is more, it is not always easy to decide which parts of the research workflow should be open as there are many other factors at play that influence one's decision, such as funder requirements. Nevertheless, the game will give you an (albeit sometimes simplified) overview of the kind of open science practices that exist.

START

WORD SEARCH

<https://thewordsearch.com/maker/>

<https://bit.ly/3ZwdNBA>

The WordSearch

OpenScience_v1

N	O	I	T	A	C	I	L	P	E	R	L
E	P	P	I	R	L	I	A	N	A	E	A
D	D	U	U	L	I	M	N	G	C	I	N
O	R	B	M	A	C	A	O	I	C	M	R
C	P	L	C	N	E	N	N	S	E	P	U
H	D	I	I	R	N	U	Y	E	S	A	O
S	M	S	T	U	C	S	M	D	S	C	J
I	P	H	A	O	E	C	I	Y	D	T	D
L	I	D	T	J	N	R	Z	D	I	F	I
B	O	A	I	A	H	I	A	U	P	A	R
U	D	T	O	O	A	P	T	T	E	C	B
P	R	A	N	A	A	T	I	S	T	T	Y
R	E	P	O	S	I	T	O	R	Y	O	H
S	R	I	A	F	U	N	N	P	N	R	M

REPLICATION
REPOSITORY
PUBLISHDATA
HYBRIDJOURNAL
MANUSCRIPT
IMPACTFACTOR
OAJOURNAL
CITATION
FAIR
ANONYMIZATION
DOI
PID
LICENCE
PUBLISHCODE
STUDYDESIGN
DMP
ACCESS

[Download / Print Puzzle](#) [Puzzle Settings](#) [Game Theme](#) [Report a bug](#)



What is Open Science



unesco

Opening science involves **opening up the processes** of creating, evaluating, sharing, exploring and storing scientific knowledge, practices and perspectives.

The UNESCO Recommendation on Open Science are aiming to:

- make multilingual scientific knowledge **openly available**, **accessible** and **reusable** for **everyone**,
- increase scientific **collaborations** and **sharing** of information for the benefits of science and society, and to
- **open the processes** of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community.

Open does not mean '*uncontrolled*'

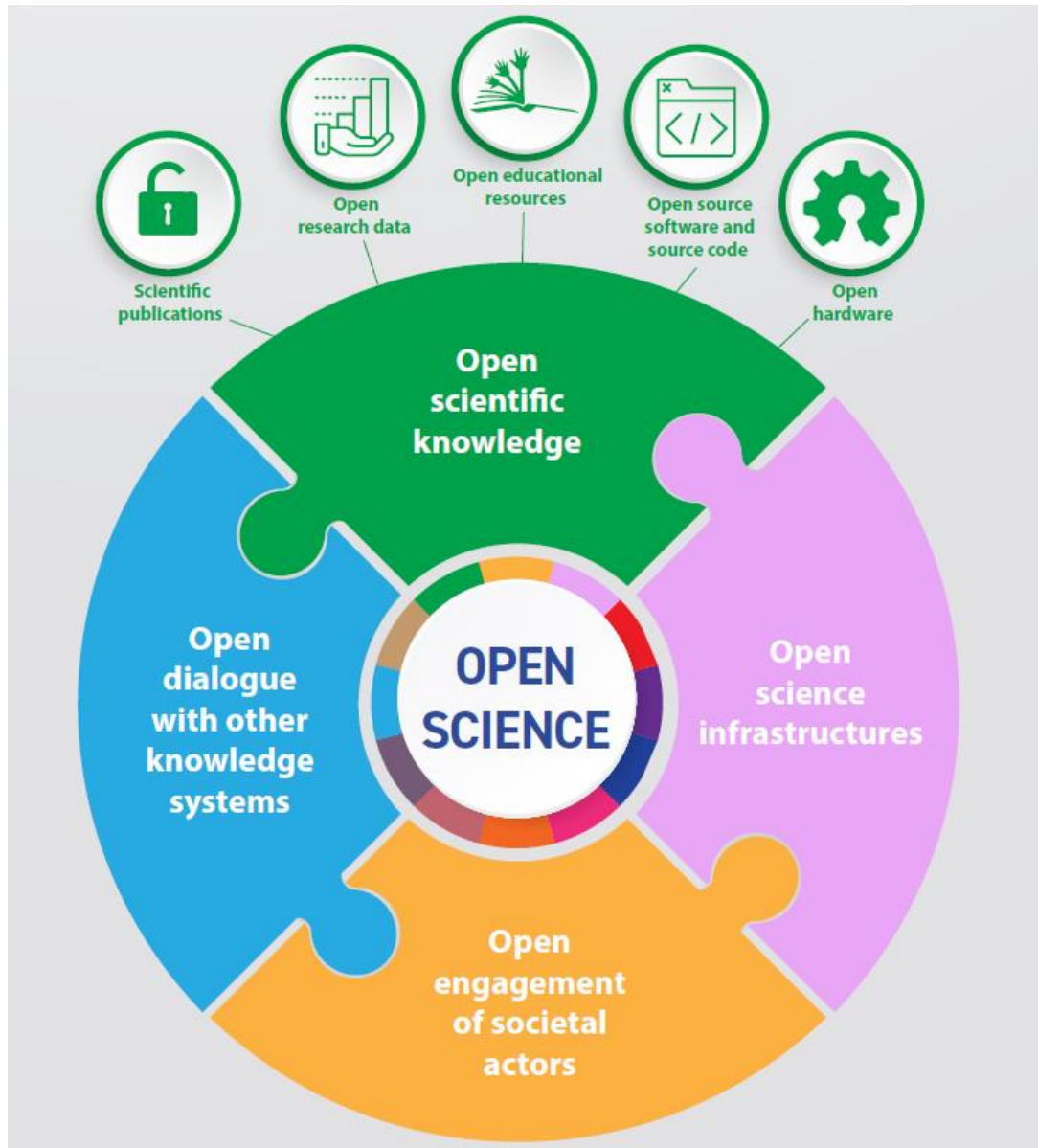
Open does not mean '*cost-free*'



Source: [Understanding Open Science](#)



Open Science knowledge



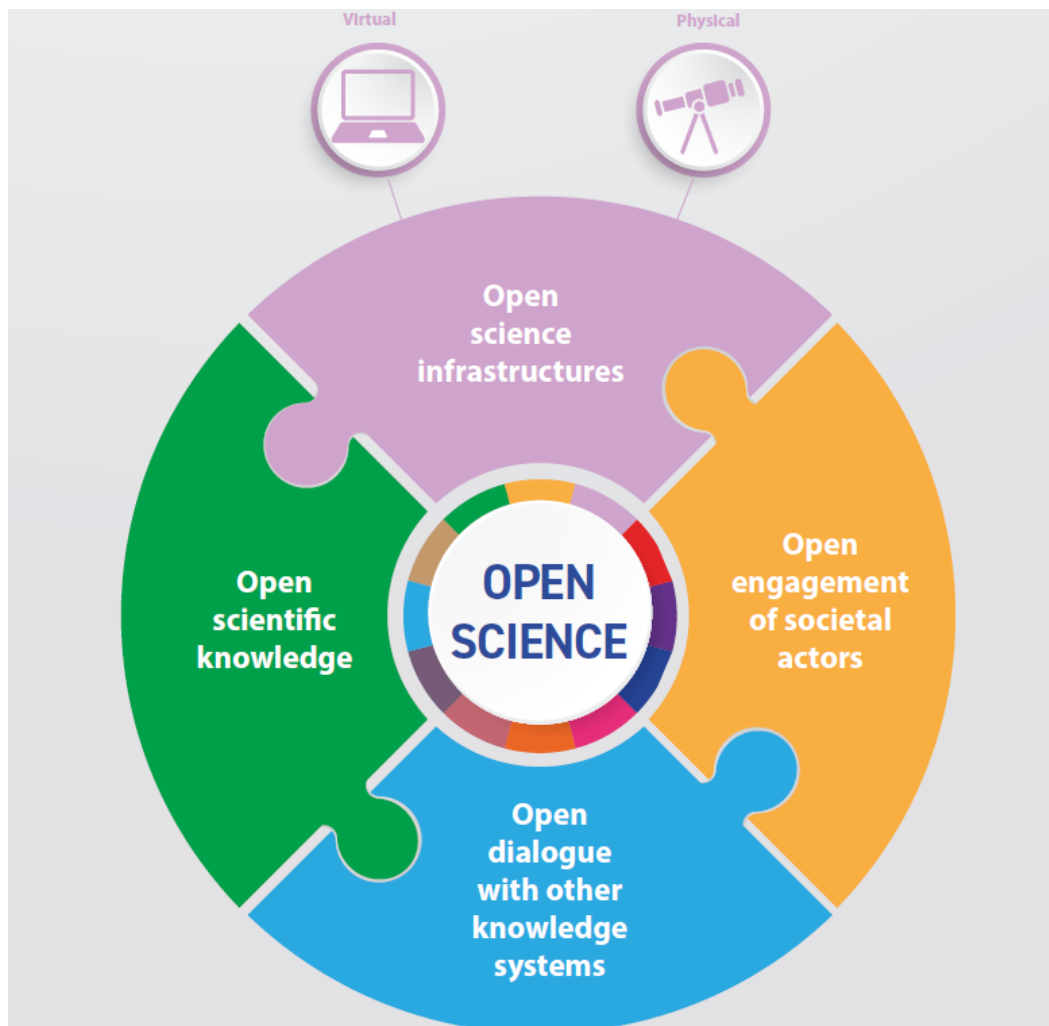
Refers to open access to scientific publications, research data, metadata, open educational resources, software, and source code and hardware that are available in the public domain or under copyright and licensed under an open licence that allows access, re-use, repurpose, adaptation and distribution under specific conditions. It also refers to the possibility of opening research methodologies and evaluation processes.



Source: [Understanding Open Science](#)



Open Science Infrastructures



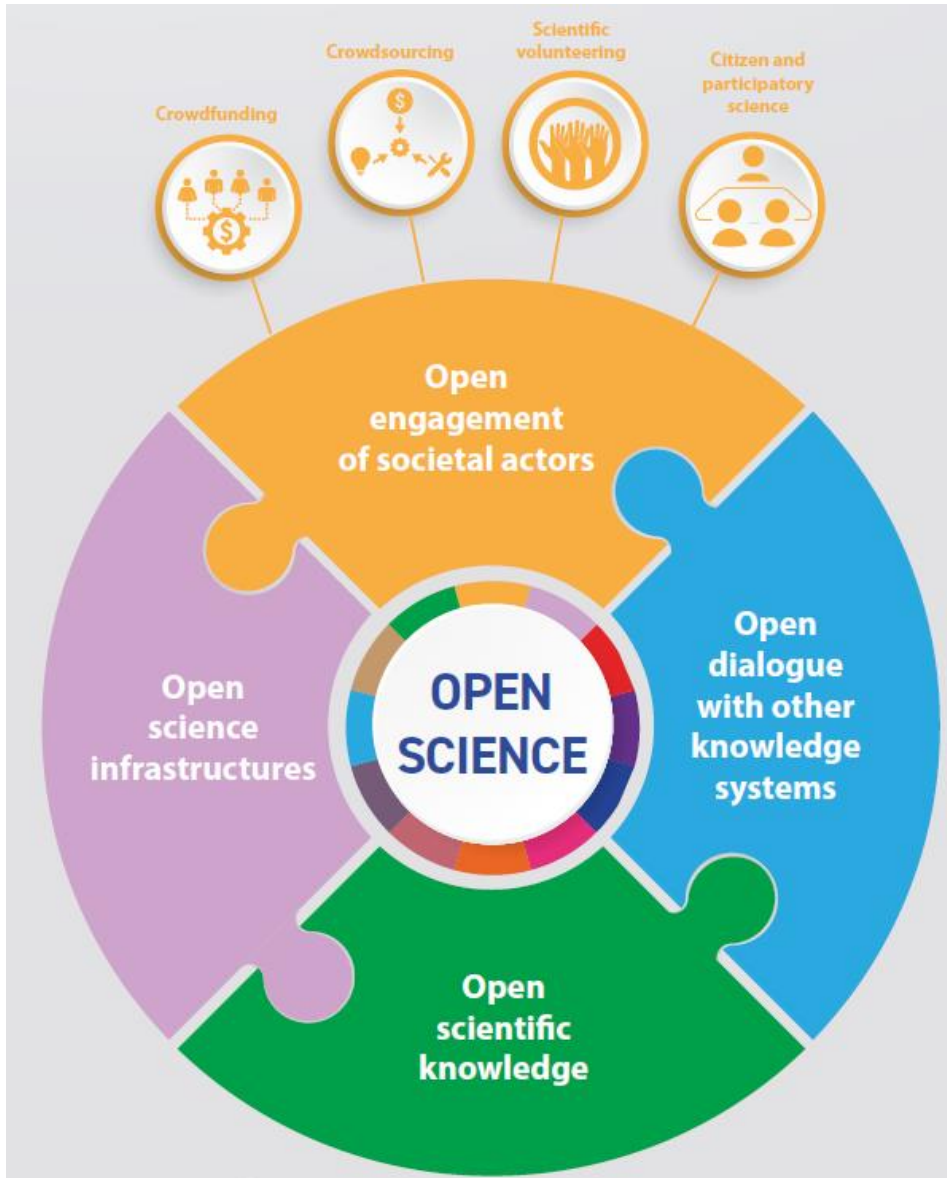
Refer to shared research infrastructures that are needed to support open science and serve the needs of different communities. Open science infrastructures can provide common access to physical facilities as well as essential open and standardized services and capabilities to manage and create access, portability, analysis and federation of data, scientific literature, thematic science priorities or community engagement.



Source: [Understanding Open Science](#)



Engagement of Societal Actors



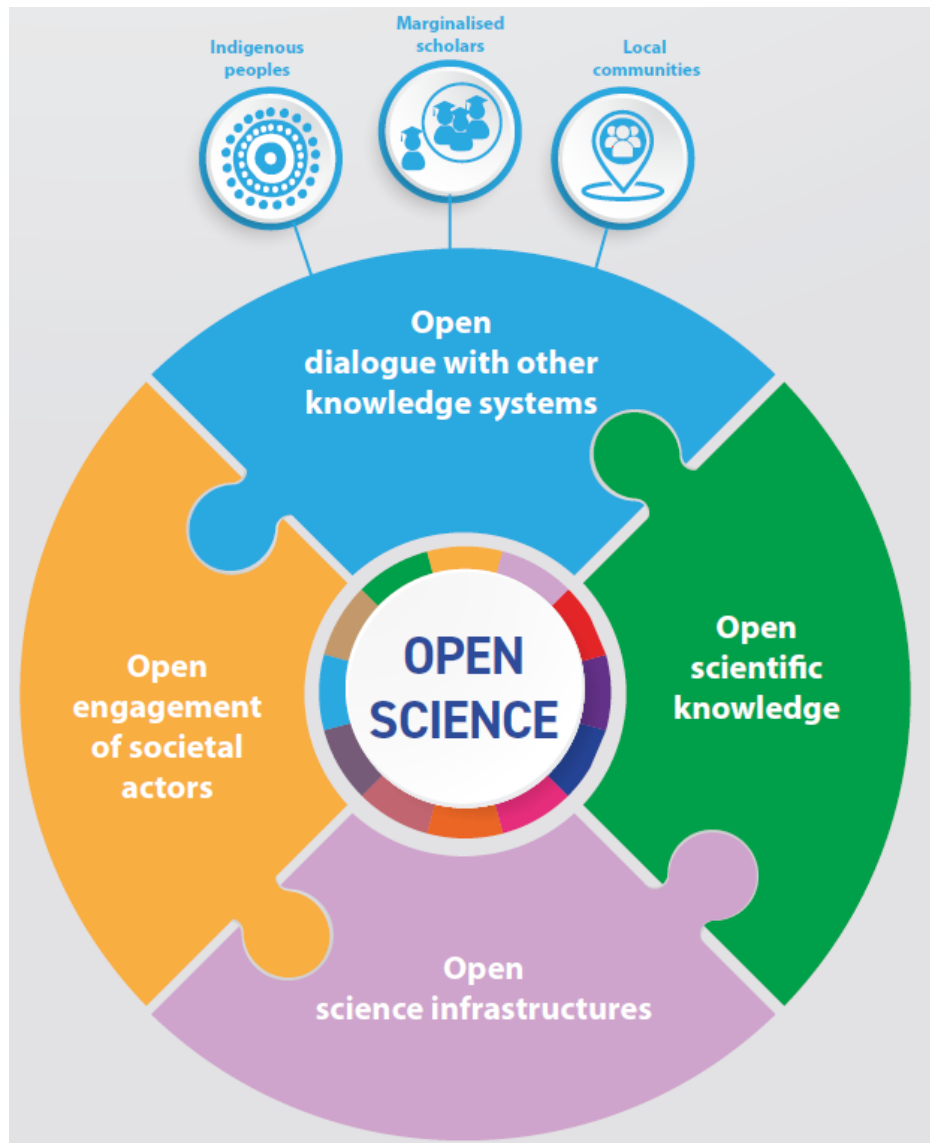
Refers to extended collaboration between scientists and societal actors beyond the scientific community, by opening up practices and tools that are part of the research cycle and by making the scientific process more inclusive and accessible to the broader inquiring society. New forms of collaboration and work can be used, such as crowdfunding, crowdsourcing, scientific volunteering and citizen science.



Source: [Understanding Open Science](#)

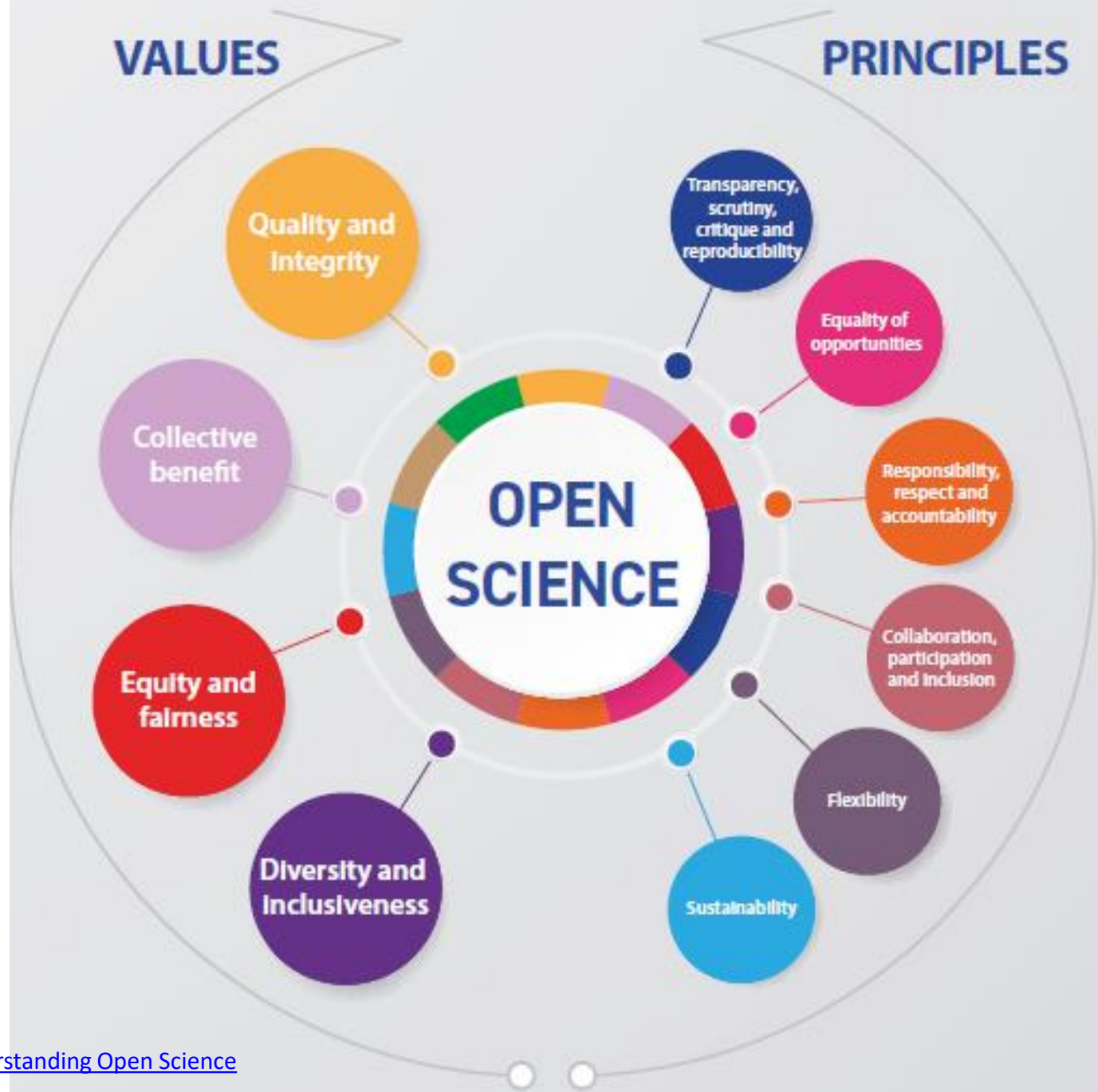


Open dialog with Other Knowledge Systems



Refers to the dialogue between different knowledge holders, that recognizes the richness of diverse knowledge systems and epistemologies and diversity of knowledge producers, in line with the 2001 [UNESCO Universal Declaration on Cultural Diversity](#).





What is research data?

DEFINITIONS

Research data constitute **primary sources that underpin scientific research** and enable derivation of theoretical or applied findings.

([Preparing research data for open access : guide for data producers](#), 2015)

The tangible forms this 'material' may take are e.g. "facts, observations, interviews, recordings, measurements, experiments, simulations, and software; numerical, descriptive and visual; raw, cleaned up and processed" (Van Berchum & Grootveld, 2017).



INFORMATION TYPES



Types of data: format

General
format

- **Numeric** Data consisting largely of values expressed as digits from 0 to 9 and, optionally, signs for negative values, decimal points, or letters only when intended to represent numbers (for example, A-F or a-f in hexadecimal).
- **Text** Data consisting largely of text, including letters, numbers, and special characters or symbols used in writing for punctuation, abbreviation, etc. For example, interview transcriptions, narratives or essays written by study participants, newspaper articles, etc.
- **Still Image** Static images, such as graphs, drawings, photographs, diagnostic/medical images like X-rays, etc.
- **Geospatial** ...
- **Audio** Recorded sound, including voice, music, etc.
- **Video** Moving images.
- **Software** ...
- **Interactive Resource** ...
- **Three D** ...
- **Other** ...

Research data in Social Sciences

Type of data

Quantitative data

Qualitative data



General description

In quantitative research, the gathered information is in numerical form. Quantitative research is used to quantify behaviour, attitudes or opinions. The goal of quantitative research is often to test ideas stated at the start of the research, to formulate facts and uncover patterns.

Qualitative research is primarily exploratory research. It gathers information that is not in numerical form. The goal of qualitative research is often to develop (new) ideas and a deeper understanding not achievable by numerical scores.

Data life cycle



CESSDA Training Team (2017 - 2020). *CESSDA Data Management Expert Guide*. Bergen, Norway: CESSDA ERIC. Retrieved from <https://www.cessda.eu/DMGuide>

As open as possible, as closed as needed

When there are **justified ethical or legal reasons for the protection of data**, such **sensitive data may be accessible under limited conditions**.

The ADP offers several such options:

- **data files with different levels of protection**, prepared for different users (PUF - public use file, SUF - scientific use file, ScUF - Secure use file, CUF - campus use file),
- **different types of users**, based on their purpose of use: scientific, public, educational, commercial,
- **regimes of access**: standard access, access under special conditions, access in a safe room.

https://www.adp.fdv.uni-lj.si/eng/usposobi/odprti_podatki/

Benefits of sharing your data



<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/1.-Plan/Benefits-of-data-management>

Research data management

... refers to how you handle, organise, and structure your research data throughout the research process.

- A good data management strategy takes into account technical, organisational, structural, legal, ethical and sustainability aspects.
- Makes your research time-efficient, reproducible and safe as possible, if your data management is well thought through, structured, and documented.



FAIR principles



Findable

To aid automatic discovery of relevant datasets, (meta)data should be easy to find by both humans and machines and be assigned a persistent identifier.

Accessible

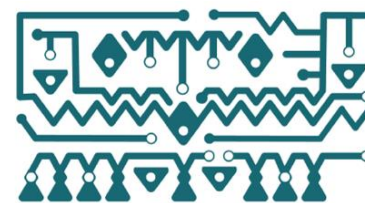
Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

Interoperable

(Meta)data should use standardised terms (controlled vocabularies), have references to other (meta)data and be machine actionable.

Reusable

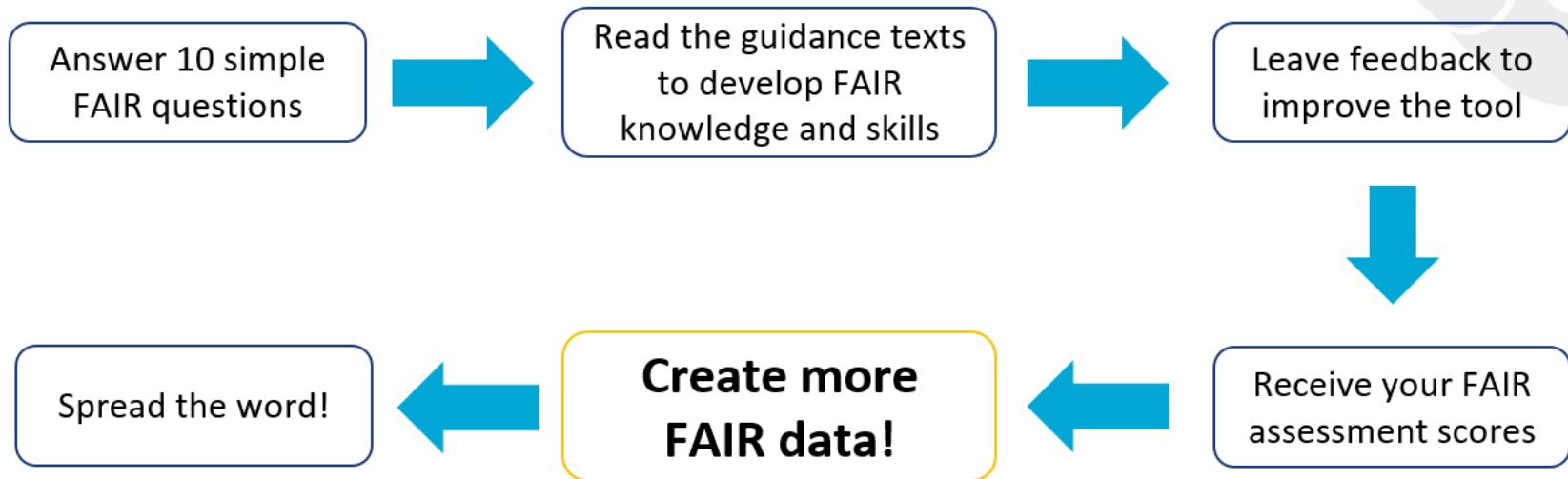
(Meta)data are sufficiently well described for both humans and computers to be able to understand them and have a clear and accessible data usage license.



**CARE Principles
for Indigenous
Data Governance**

FAIR | Aware

Your first step towards your FAIR data(set)



Data Archiving and Networked Services
DANS

Source: <https://doi.org/10.5281/zenodo.7180525>

[FAIR Aware questions](#)



European-wide alignment of RDM requirements

- Science Europe Initiative (<https://www.scienceeurope.org/policy/policy-areas/research-data/rdm-initiative/>)
 - S.E. is European organization of 36 Research Funding and Research Performing Organisations from 27 European countries.
 - [Practical Guidelines to the International Alignment of Research Data Management published](#) in January 2021



Guide developed by CESSDA Archives

Training / Training Resources / Data Management Expert Guide



Data Management Expert Guide

This guide is designed by European experts to help social science researchers make their research data Findable, Accessible, Interoperable and Reusable (FAIR).

You will be guided by different European experts who are - on a daily basis - busy ensuring long-term access to valuable social science datasets, available for discovery and reuse at one of the [CESSDA social science data archives](#).

**Self-study for researchers
(15 hours of online content)**

www.cessda.eu/DMEG

The General Data Protection Regulation (GDPR)

- The GDPR applies from the 25 May 2018.
- The GDPR applies to any data controller or data processor in the EU **who collects personal data about a data subject** of any country, anywhere in the world.
- A data controller or data processor that **is based outside the EU but collects personal data on EU citizens** will also be covered by the GDPR.
- This means that **a researcher** (data controller) **based within the EU** who collects personal data about a participant, from any other country within the EU, or the world, needs to comply with the GDPR.

Short definition “personal data”

- *Personal data is any information that may be used to identify a person directly or indirectly*

- **Directly identifying personal data**

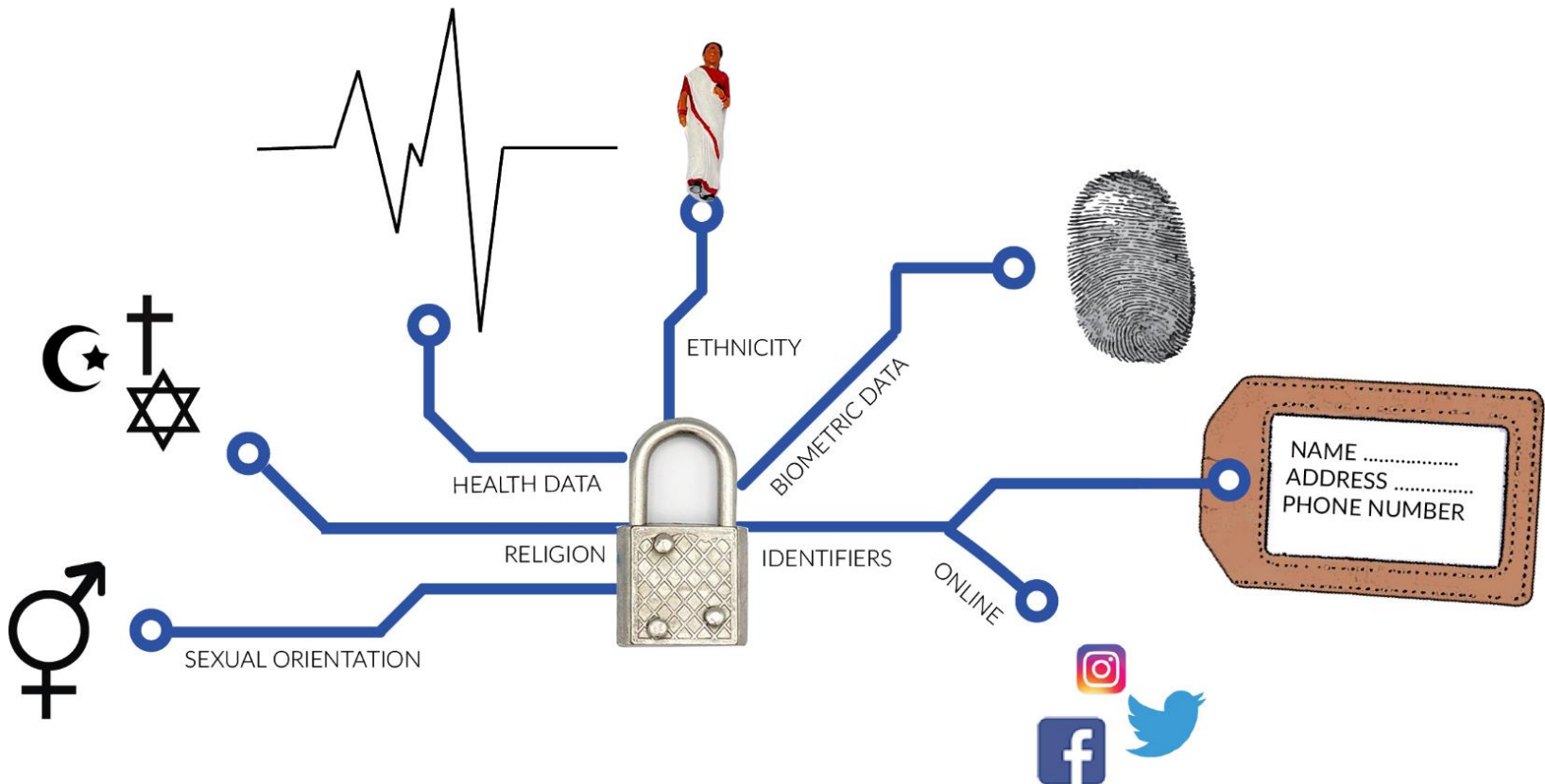
through full name, personal identification number

- **Indirectly identifying personal data**

through a combination of background information



Sensitive personal data



The Grounds for Processing Personal Data

There are 6 grounds for the processing of personal data, and one of these must be present in order to process a data subject's personal data:

1. Consent of the data subject
2. Necessary for the performance of a contract
3. Legal obligation placed upon controller
4. Necessary to protect the vital interests of the data subject
5. Carried out in the public interest or is in the exercise of official authority
6. Legitimate interest pursued by controller

Strategy for Sharing Data

- Obtain **informed consent**, also for data sharing and preservation or curation
- Protect identities e.g. **anonymisation** and not collecting personal data if not necessary
- **Regulate access** where needed (all or part of data) e.g. by group, use or time period
- **Securely store** and protect personal and sensitive data



Assessing risk when sharing data: a guide

February 2022

Open Data Institute

Contents

Contents	1
About	2
Purpose of this guide	3
Minimising risk at the point of data creation	4
Managing risks when sharing data	6
Risk categories:	6
Consider likelihood and severity of any impact	7
Legal and regulatory risks	9
Does the data contain any personal data?	9
Examples of personal and sensitive personal data:	10
Does the data contain third party data (data created by someone else)?	10
Do you have the legal permissions to share the data?	10
Examples of data typically sourced from third parties:	11
Examples of data licences and data sharing agreements:	11
Are there any other legal or regulatory considerations relevant to this data?	12
Examples of other legal or regulatory considerations:	12
Ethical risks	13
Are there any cultural considerations relevant to this data?	13
Is sharing the data likely to impact people or communities?	14
Will sharing the data impact the environment?	14
Does the data contain anything that, if made available, could impact national security?	15
Examples of data that, if shared, could impact national security:	15
Does the data contain anything that, if made available, could impact the security of the organisation, or it's staff?	15
Examples of data that could impact the security of an organisation or it's staff:	15
Reputational risks	16
Will anyone be surprised by you holding, sharing or using this data?	17
Is a data quality caveat required?	17
Are there any free-text or comment fields in the dataset?	17
Commercial risks	18
Does the data contain anything commercially sensitive?	18
Examples of commercially sensitive data:	18
Next steps	19
Further resources	20



DPIA - Data Protection Impact Assessment

- Sensitive data
- Consent not possible
- Long term processing/ archiving
- Vulnerable group
- Very identifiable data
- Combination of the above
- The DPIA is a written document to be formally approved by the University and DPO

DATA Publication (P!!!)



PUBLICATIONS AND DATA

It is expected that a Data Publication will ensure that data will potentially be considered as a first-class research output (Knowledge Exchange, 2013).

For a dataset to “count” as a publication should be:

- Properly documented with metadata;
- **Reviewed for quality;**
- Searchable and discoverable in catalogues (or databases);
- Citable in articles.

Where to publish?



⊕ Journal supplementary material service

⊕ Institutional data repository

Repository of the
University of Ljubljana

⊕ General purpose repository

zenodo

⊕ Domain specific data repository

DARIAH-SI

⊕ Trusted domain specific data repository

CLARIN.SI

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES



Other EU research data infrastructures

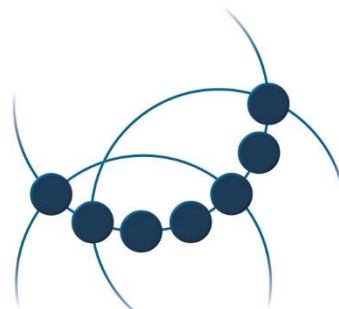
CLARIN - European Research Infrastructure for Language Resources and Technology

CLARIN.SI



CLARIN

Common Language Resources and
Technology Infrastructure



A stylized blue flower-like logo with five petals.
DARIAH-EU
Digital Research Infrastructure
for the Arts and Humanities

A stylized purple flower-like logo with five petals.
DARIAH-SI



A grid of small squares in various colors (red, green, blue, yellow, grey).
E-RIHS
EUROPEAN RESEARCH INFRASTRUCTURE
FOR HERITAGE SCIENCE

ELIXIR | A distributed infrastructure for life-science information



Domain specific data repositories & DP



PUBLICATIONS AND DATA

Advantages

- Offers **specialist domain knowledge** and data management expertise, e.g. to create a catalogue record and documentation;
- More likely to **accept complete datasets**;
- Provides preservation and curation to **community standards**, e.g. file formats migration;
- Ability to **control access** of (sensitive) personal data;
- May handle data re-use queries;
- May make your data visible via dissemination and promotion.

Disadvantages

- Most likely to be selective about what kind of data they accept;
- Requires advance planning of the effort needed to meet high standards for metadata and documentation.

Journal supplementary ≠ Data publication



PUBLICATIONS AND DATA

Advantages

- Most likely to comply with the journal or publisher's requirements;
- Data readily available alongside published findings

Disadvantages

- May claim copyright over the data;
- May keep data behind a subscription wall;
- Unlikely to offer a data repository's functionality or long-term solution;
- May not apply user-friendly or preservation formats;
- **More likely to accept subsets rather than complete datasets.**

([CESSDA DMEG](#), 2017-2020)



Map a research topic

Get an overview - Find documents - Identify relevant concepts



- PubMed (life sciences)
- BASE** (all disciplines)

[Refine your search](#) ▾

xebovat

GO

Try out: [digital education](#) [climate change AND impact](#)



BA Students' Perceptions of Standardized English Tests ;
Percepciones de los estudiantes de licenciatura sobre las pruebas
estandarizadas de inglés (2022-08-16)



EOSC Portal - A gateway to information and resources in EOSC

Browse by research activity

Browse by scientific domain



Discover research outputs

Find datasets, scientific publications, software...



Publish research outputs

Store, backup, archive your data, publications, software



Process and analyse

Verify, organise, transform, integrate, and extract data in an appropriate output format software service



Access computing and storage resources

Find HPC, IT centres for science, cloud computing, cloud container computing, online storage



Find instruments & equipment

Find research instruments and scientific equipment



Access training material

Find lessons, courses, videos



Access research infrastructures



Manage research data



Why publish research data?



Data Sharing and Management Snafu in 3 Short Acts

Karen Hanson, Alisa Surkis and Karen Yacobucci (2012) NYU Health Sciences Library:
<https://www.youtube.com/watch?v=N2zK3sAtr-4>

University of Ljubljana
Faculty of Social Sciences
Social Science Data Archive
Kardeljeva ploščad 5
1000 Ljubljana
Slovenia

 www.adp.fdv.uni-lj.si

 arhiv.podatkov@fdv.uni-lj.si

 [Arhiv.Druzboslovnih.Podatkov](https://www.facebook.com/Arhiv.Druzboslovnih.Podatkov)

 [@ArhivPodatkov](https://twitter.com/ArhivPodatkov)