DATA MANAGEMENT PLAN: STEP-BY-STEP

University of Maribor Open Science Summer School 2022

September 13, 2022

dr. Tea Romih

Central Technical Library at the University of Ljubljana

tea.romih@ctk.uni-lj.si

What is a data management plan (DMP)?

A DMP is a document that describes how data will be:

- **collected,**

- **analyzed,**

- **formatted,**

- **stored,**

- **preserved,**

- **protected,**

- **licensed**

- **and shared** throughout the research process.

Some DMPs also contain detailed information about the research group, project, funder, etc.

Why do we need data management plans?

1. Funder's requirement

- European Commission (Horizon Europe – incl. MSCA, Euratom)

- European Research Council

- other funding bodies, both European and non-European (e.g., FWF, FWO, SNSF, Wellcome Trust, US National Science Foundation, US National Institutes of Health …)

Why do we need data management plans?

1. Funder's requirement

- European Commission (Horizon Europe – incl. MSCA, Euratom)

- European Research Council

- other funding bodies, both European and non-European (e.g., FWF, FWO, SNSF, Wellcome Trust, US National Science Foundation, US National Institutes of Health …)

2. A DMP helps you plan and organize your data collection

European Commission | EN English | Search

Home > Research and innovation > Strategy > Strategy 2020-2024 > Our digital future > Open Science > European Open Science Cloud (EOSC)

European Open Science Cloud (EOSC)

What the cloud is, how it was developed and being implemented

What the European Open Science Cloud is

The ambition of the European Open Science Cloud (EOSC) is to provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and reuse data, tools and services for research, innovation and educational purposes.

This environment will operate under well-defined conditions to ensure trust and safeguard the public interest.

The EOSC enables a step change across scientific communities and research infrastructures towards

- seamless access
- FAIR ↗ (Findability, Accessibility, Interoperability and Reusability) management
- reliable reuse of research data and all other digital objects produced along the research life cycle (e.g. methods, software and publications)

The European Open Science Cloud (EOSC) ultimately aims to develop a 'Web of FAIR Data and services' for science in Europe upon which a wide range of value-added services can be built. These range from visualisation and analytics to long-term information preservation or the monitoring of the uptake of open science practices.

The EOSC is recognised by the Council of the European Union among the 20 actions of the policy agenda 2022-2024 of the European Research Area (ERA) with the specific objective to deepen open science practices in Europe. It is also recognised as the "science, research and innovation data space" which will be fully articulated with the other sectoral data spaces defined in the European strategy for data.

Full deployment of the EOSC will lead to higher research productivity, new insights and innovations, as well as improved reproducibility and trust in science.

Why do we need data management plans?

1. Funder's requirement

- European Commission (Horizon Europe – incl. MSCA, Euratom)

- European Research Council

- other funding bodies, both European and non-European (e.g., FWF, FWO, SNSF, Wellcome Trust, US National Science Foundation, US National Institutes of Health …)

2. A DMP helps you plan and organize your data collection

3. A DMP can help prevent adverse events

- miscommunication among project members

- unethical uses of data

- data errors

- data loss
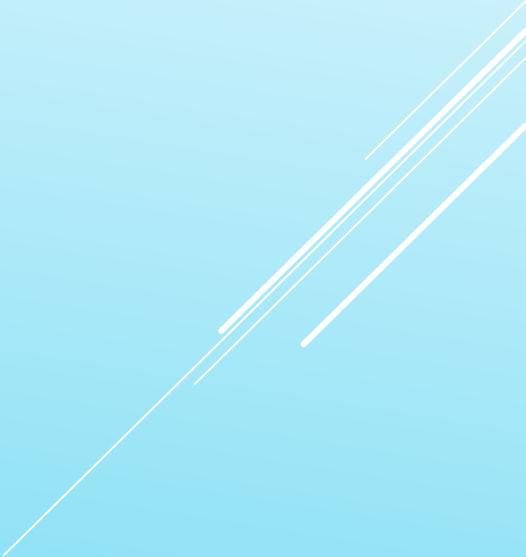
Stolen laptop contains cancer cure data

Distraught Oklahoma University researcher says her stolen MacBook contains years of valuable data on prostate cancer. And no, she didn't back up.

Leslie Katz
Jan. 13, 2011 2:01 p.m. PT

2 min read

No questions asked
$1000 reward

for anyone who leads to the safe return of
the stolen computer with all data intact

When: Jan 9 (Sunday) around Noon.
Where: My car (smashed car window) at Panera parking lot.
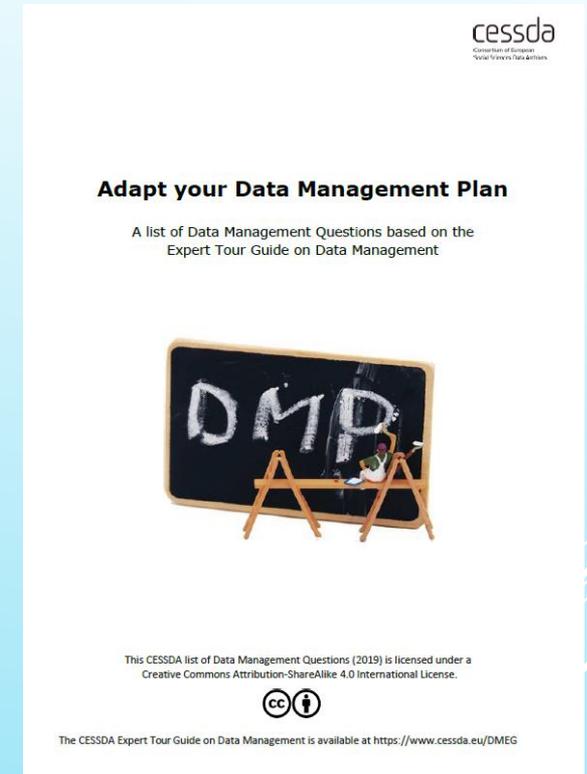What: 13-inch White Macbook (password protected)
    in dark orange computer bag.
Contact: sshin2@ouhsc.edu
        405-370-3099

PS. Thief, it is OK. Everybody makes mistake. Please return my computer safely for no questions asked-$1000 reward. If so, I would be forever grateful to you.

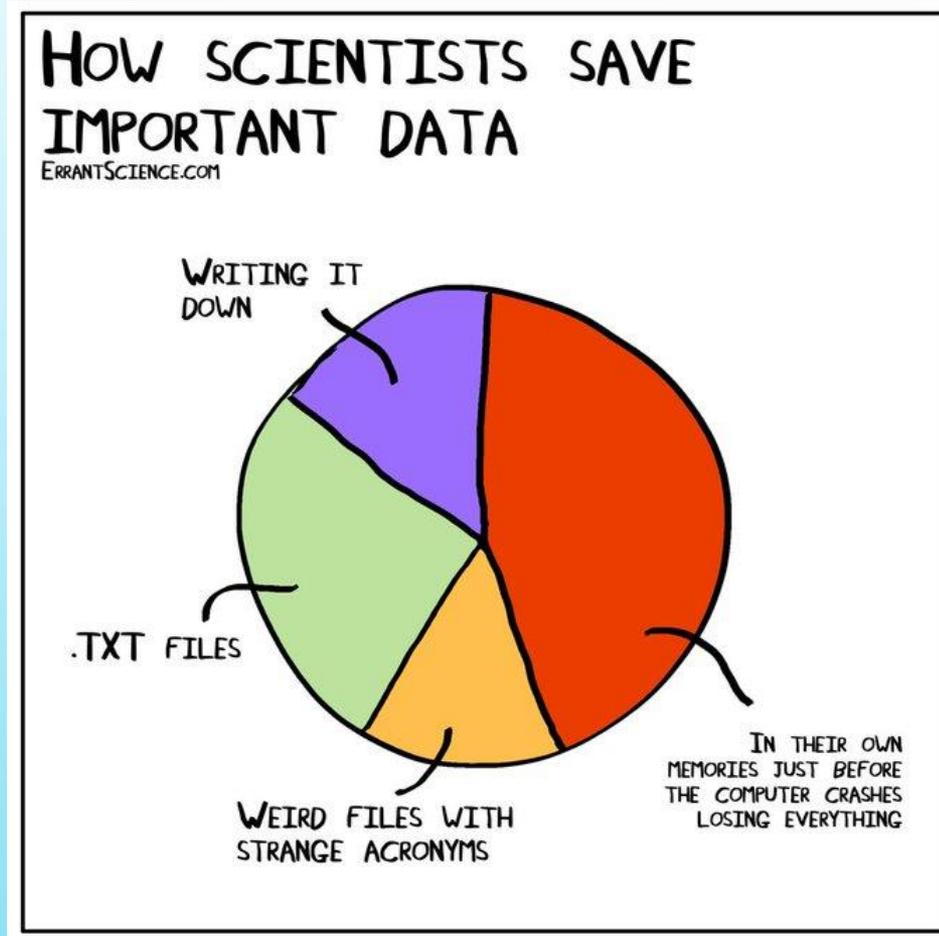https://www.cnet.com/culture/stolen-laptop-contains-cancer-cure-data/

The foundation for this lecture

Horizon Europe

Data Management Plan Template

Version 1.0
05 May 2021

European Research Council
Executive Agency

Established by the European Commission

ERC Data Management Plan

Template

cessda
(Consortium of European
Social Sciences Data Archives)

Adapt your Data Management Plan

A list of Data Management Questions based on the
Expert Tour Guide on Data Management

This CESSDA list of Data Management Questions (2019) is licensed under a
Creative Commons Attribution-ShareAlike 4.0 International License.

The CESSDA Expert Tour Guide on Data Management is available at https://www.cessda.eu/DMEG

FAIRification-based

Data lifecycle-based

STEP 0
(not in the DMPs)

Steps in the creation of a DMP

Step 1: **Basics**

STEP 2: **Dataset**

STEP 3: **Findability**

Step 4: **Accessibility**

Step 5: **Interoperability**

STEP 6: **Reusability**

STEP 7: **Costs & resources**

STEP 8: **Safety & security**

STEP 9: **Ethics**

STEP 1: Describe the basics

This step is missing from the Horizon Europe & ERC DMPs. However, it is useful to include it because you can recycle the information for the ReadMe file and/or data paper.

- Date of the DMP creation

- Project title

- Researchers & their contact information (incl. the main contact person, e.g., the project manager)

- Funder, if applicable (incl. grant number)

- Brief description of the project (incl. research question & project timeline)

- Data ownership and administrative responsibilities

STEP 2: Describe the dataset

This information can also be recycled in a ReadMe file and/or data paper.

- Will **new data** be generated or **exisiting data** will be reused?

STEP 2: Describe the dataset

This information can also be recycled in a ReadMe file and/or data paper.

- Will **new data** be generated or **exisiting data** will be reused?

- Expected **size** of the data

Generalist Repository Comparison Chart

This chart is designed to assist researchers in finding a generalist repository should no domain repository be available to preserve their research data. Generalist repositories accept data regardless of data type, format, content, or disciplinary focus. For this chart, we included a repository available to all researchers specific to clinical trials (Vivli) to bring awareness to those in this field.

https://fairsharing.org/collection/GeneralRepositoryComparison

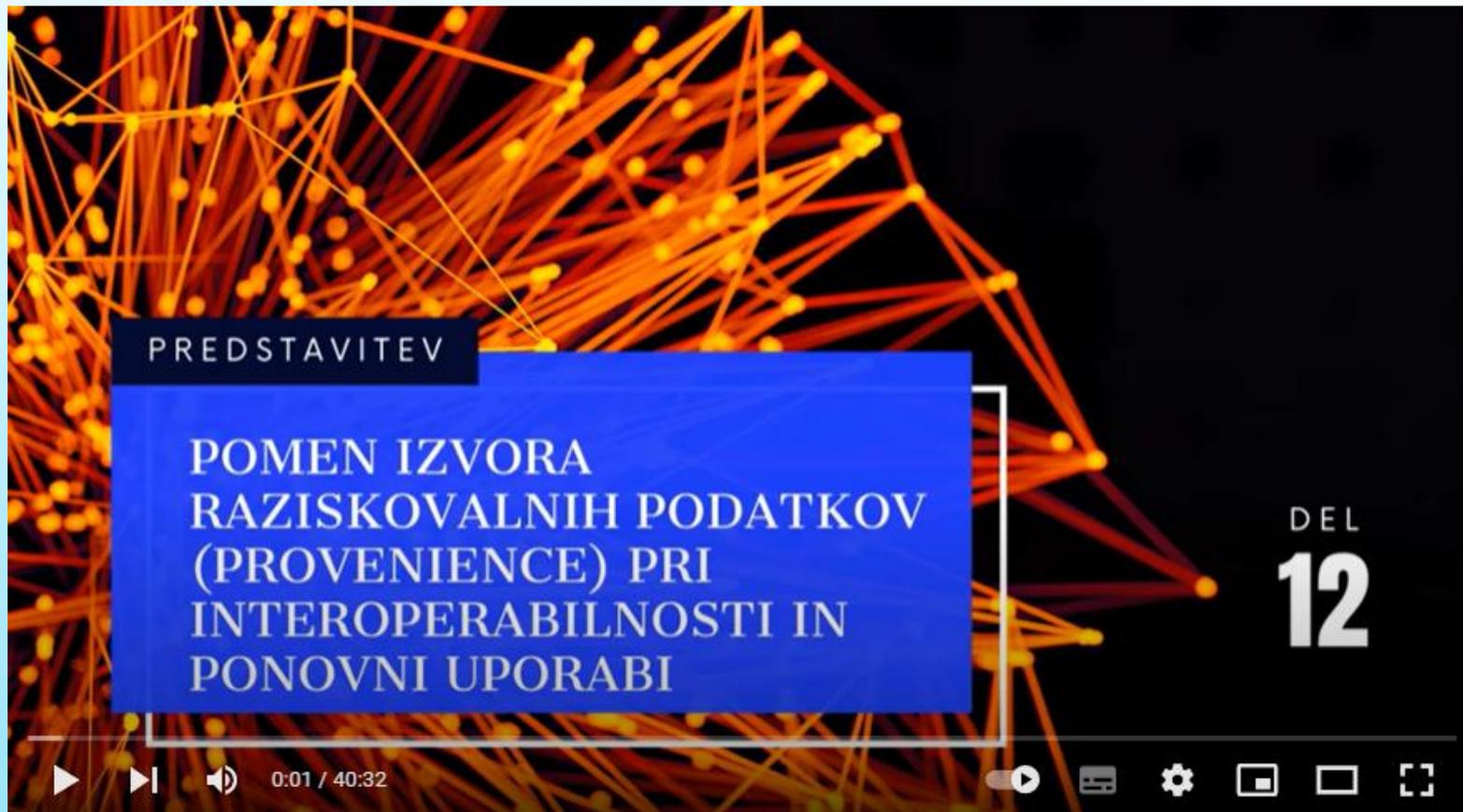| TOPIC	HARVARD DATAVERSE	DRYAD	FIGSHARE	MENDELEY DATA	OSF	VIVLI	ZENODO
Brief Description	Harvard Dataverse is a free data repository open to all researchers from any discipline, both inside and outside of the Harvard community, where you can share, archive, cite, access, and explore research data.	Open-source, community-led data curation, publishing, and preservation platform for CC0 publicly available research data. Dryad is an independent non-profit that works directly with: • researchers to publish datasets utlizing best practices for discovery and reuse • publishers to support the integration of data availability statements and data citations into their workflows • institutions to enable scalable campus support for research data managment best practices at low cost	A free, open access, data repository where users can make all outputs of their research available in a discoverable, reusable, and citable manner. Users can upload files of any type and are able to share diverse research products including datasets, code, multimedia files, workflows, posters, presentations, and more. With discoverable metadata supporting FAIR principles, file visualizations, and integrations, researchers can make their work more impactful and move research further faster.	Mendeley Data is a free repository specialized for research data. Search more than 20+ million datasets indexed from 1000s of data repositories and collect and share datasets with the research community following the FAIR data principles.	OSF is a free and open source project management tool that supports researchers throughout their entire project lifecycle in open science best practices.	Vivli is an independent, non-profit organization that has developed a global data-sharing and analytics platform. Our focus is on sharing individual participant-level data from completed clinical trials to serve the international research community.	Powering Open Science, built on Open Source. Built by reserachers for researchers. Run from the CERN data centre, whose purpose is long term preservation for the High Energy Physics discipline, one of the largest scientific datasets in the world
Size limits	No byte size limit per dataset. Harvard Dataverse currently sets a file size limit of 2.5GB.	300GB/dataset	Soft limit of 20GB/file for free accounts. System limit of 5000GB/file. Unlimited storage of public data but 20GB storage for private data for free accounts. Email info@figshare.com to have upload and storage limits raised.	10GB per dataset	Projects currently have not storage limit. There is a 5GB/file upload limit for native OSF Storage. There is no limit imposed by OSF for the amount of storage used across add-ons connected to a given project.	If more than 10GB per study data, reach out to us	50GB per dataset, contact us via https://zenodo.org/support for higher limits
Storage space per researcher	1 TB per researcher	No limit	No limit	No limit	No limit	No limit	No limit
Persistent, Unique Identifier Support	DOI, Handle	DOI	DOI	DOI	DOI	DOI	DOI

STEP 2: Describe the dataset

This information can also be recycled in a ReadMe file and/or data paper.

- Will **new data** be generated or **exisiting data** will be reused?

- Expected **size** of the data

- **Data types, file formats, naming conventions**

- **Organisation** of data (simple files vs. databases)

https://dirrosdata.ctk.uni-lj.si/raziskovalni-podatki/oblikovanje-podatkov-za-deljenje/

STEP 2: Describe the dataset

This information can also be recycled in a ReadMe file and/or data paper.

- Will **new data** be generated or **exisiting data** will be reused?

- Expected **size** of the data

- **Data types, file formats, naming conventions**

- **Organisation** of data (simple files vs. databases)

- **Origin/provenance** of the data (who, what, when, how, why)

https://dirrosdata.ctk.uni-lj.si/metapodatki/provenienca/

https://www.youtube.com/watch?v=E6kyKCR7Szw&ab_channel=CTKLjubljana

Google

Dataset Search

Search for Data Sets 🔍

Try coronavirus covid-19 or education outcomes site:data.gov.

Learn more about Dataset Search.

https://datasetsearch.research.google.com/

Last updated | Download format | Usage rights | Topic | Free

Saved data sets

100+ data sets found

Li-ion Battery Aging Datasets
data.nasa.gov
catalog.data.gov
+1more

application/rdfxml +5

Updated Jun 26, 2018

Panasonic 18650PF Li-ion Battery Data
data.mendeley.com
www.narcis.nl

Updated Jun 21, 2018

LG 18650HG2 Li-ion Battery Data and Example Deep Neur...
data.mendeley.com
search.datacite.org

Updated Mar 5, 2020

Li-ion Battery Aging Datasets

Explore at Rally - Open Data Portal

Explore at catalog.data.gov | **Explore at data.wu.ac.at**

221 scholarly articles cite this dataset (View in Google Scholar)

csv, application/rssxml, application/rdfxml, xml, json, tsv

Data set updated

Jun 26, 2018

Description

This data set has been collected from a custom built battery prognostics testbed at the NASA Ames Prognostics Center of Excellence (PCoE). Li-ion batteries were run through 3 different operational profiles (charge, discharge and Electrochemical Impedance Spectroscopy) at different temperatures. Discharges were carried out at different current load levels until the battery voltage fell to preset voltage thresholds. Some of these thresholds were lower than that recommended by the OEM (2.7 V) in order to induce deep discharge aging effects. Repeated charge and discharge cycles result in accelerated aging of the batteries. The experiments were stopped when the batteries reached the end-of-life (EOL) criteria of 30% fade in rated capacity (from 2 Ah to 1.4 Ah).

STEP 2: Describe the dataset

This information can also be recycled in a ReadMe file and/or data paper.

- Will **new data** be generated or **exisiting data** will be reused?

- Expected **size** of the data

- **Data types, file formats, naming conventions**

- **Organisation** of data (simple files vs. databases)

- **Origin/provenance** of the data (who, what, when, how, why)

- ***Quality assurance*** *at the data collection stage (only CESSDA)*

STEP 3: Findability aspects

- Will data be identified by a **persistent identifier**? → *Will they be deposited to a trusted repository?* (next step - Accessibility)

Persistent identifiers (PIDs)

- **Unique and permanent identifiers** of various digital objects (e.g., research papers, research data, registered reports …), non-digital objects (e.g., research projects, funders, books, paintings …) and people

Persistent identifiers (PIDs)

- **Unique and permanent identifiers** of various digital objects (e.g., research papers, research data, registered reports …), non-digital objects (e.g., research projects, funders, books, paintings …) and people

- Consist of 2 parts:

  - **A unique identification code** that enables differentiation between two entities (e.g., two researchers with the same name by their ORCiD),

  - **A service that locates the entity through time**, even if its location changes (e.g., the researcher changes institutions, a digital object gets transferred to a website with a different domain)

Persistent identifiers (PIDs)

- **Unique and permanent identifiers** of various digital objects (e.g., research papers, research data, registered reports …), non-digital objects (e.g., research projects, funders, books, paintings …) and people

- Consist of 2 parts:

  - **A unique identification code** that enables differentiation between two entities (e.g., two researchers with the same name by their ORCiD),

  - **A service that locates the entity through time**, even if its location changes (e.g., the researcher changes institutions, a digital object gets transferred to a website with a different domain)

- **Most common PIDs:** DOI, HNDL, URN, ARK

STEP 3: Findability aspects

- Will data be identified by a **persistent identifier**? → *Will they be deposited to a trusted repository?* (next step - Accessibility)

- What **metadata** will be created? Will they be machine-readable?

Metadata schemas: structure
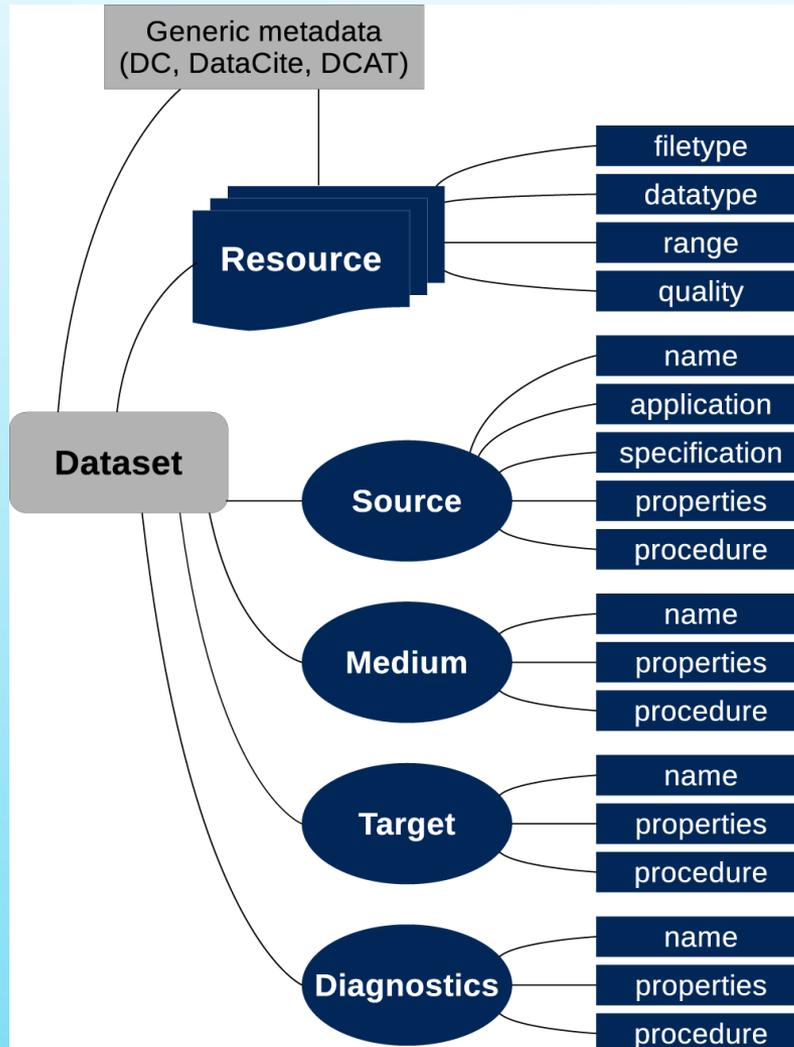
Example: a metadata schema for plasma physics

https://github.com/plasma-mds/plasma-metadata-schema

Metadata schemas: code

```xml
<article xmlns="http://champ-project.org/article"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:champ="http://champ-project.org/champ" xmlns:dcterms="http://purl.org/dc/terms/"
    xsi:schemaLocation="http://champ-project.org/article http://champ-project.org/schema/champ_article.xsd">
    <overview>
        <dcterms:title>Plasticized Poly(vinyl chloride)-Based Photonic Crystal for Ion Sensing</dcterms:title>
        <champ:analyticalFocus>Inorganic materials for ion analysis</champ:analyticalFocus>
        <dcterms:bibliographicCitation>Anal. Chem., 2014, 86 (24), pp 11986–11991 DOI:10.1021/ac503447m</dcterms:bib
    </overview>
    <champ:contact>
        <champ:person>Tatsuro Endo</champ:person>
        <champ:address>Department of Applied Chemistry, Osaka Prefecture University, 1-1 Gakuencho, Naka-ku, Sakai,
        <champ:email>endo@chem.osakafu-u.ac.jp</champ:email>
        <champ:phone>+81-72-254-9284</champ:phone>
        <champ:role>Corresponding Author</champ:role>
    </champ:contact>
    <champ:analyte>
        <champ:substance>
            <champ:inchiString>InChI=1S/K/p+1</champ:inchiString>
            <champ:inchiKey>NPYPAHLBTDXSSS-UHFFFAOYSA-N</champ:inchiKey>
            <champ:substanceName>Potassium ion</champ:substanceName>
        </champ:substance>
    </champ:analyte>
    <champ:matrix>Buffer Solution</champ:matrix>
    <champ:samplingCondition champ:name="temperature" champ:unit="degC">23.7</champ:samplingCondition>
    <champ:instrument>Polymer-based Optical Sensor</champ:instrument>
    <champ:instrument>Visible spectroscopy</champ:instrument>
    <champ:setting champ:name="wavelength" champ:unit="nm">580</champ:setting>
    <champ:concept>
        <champ:term>sensitivity</champ:term>
        <champ:scope>general</champ:scope>
        <champ:source>ChAMP Concept Vocabulary</champ:source>
    </champ:concept>
    <champ:concept>
        <champ:term>response time</champ:term>
        <champ:scope>general</champ:scope>
        <champ:source>ChAMP Concept Vocabulary</champ:source>
    </champ:concept>
</article>
```

Example: The Chemical Analysis Metadata Platform (ChAMP) Schema

https://champ.stuchalk.domains.unf.edu/journal-article-metadata-xml

List of Metadata Standards

ABCD - Access to Biological Collection Data

The Access to Biological Collections Data (ABCD) Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). The ABCD Schema attempts to be comprehensive and highly structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomised data and free-text can be accommodated.

Sponsored by Biodiversity Information Standards TDWG - the Taxonomic Databases Working Group, the current specification was last modified in 2007.

AgMES - Agricultural Metadata Element Set

A semantic standard developed by the Food and Agriculture Organization (FAO) of the United Nations, AgMES enables description, resource discovery, interoperability and data exchange of different types of information resources in all areas relevant to food production, nutrition and rural development.

Sponsored by the UN AIMS - Agricultural Information Managment Standards, the current standard was issued in November 2010.

STEP 3: Findability aspects

- Will data be identified by a **persistent identifier**? → *Will they be deposited to a trusted repository?* (next step - Accessibility)

- What **metadata** will be created? Will they be machine-readable?

- What **disciplinary or general standards** will be followed to create metadata? What if they do not yet exist?

Ontologies   Search   Annotator   Recommender   Mappings

Login   Support

Browse

Browse the library of ontologies   ?

https://bioportal.bioontology.org/ontologies

Search...

Showing 1004 of 1182   Sort: Popular

Submit New Ontology

Entry Type

☑ Ontology (1004)

☐ Ontology View (178)

Uploaded in the Last

Category

☐ All Organisms (38)

☐ Anatomy (77)

☐ Animal Development

☐ Animal Gross Anatom

☐ Arabidopsis (3)

☐ Biological Process (59

Medical Dictionary for Regulatory Activities Terminology (MedDRA) (MEDDRA)

MedDRA is an international medical terminology with an emphasis on use for data entry, retrieval, analysis, and display

Uploaded: 6/10/22

projects	classes
11	77,295

notes
1

SNOMED CT (SNOMEDCT)

SNOMED Clinical Terms

Uploaded: 6/10/22

projects	classes
23	361,907

notes
3

RxNORM (RXNORM)

RxNorm Vocabulary

Uploaded: 6/10/22

projects	classes
7	106,962

National Drug Data File (NDDF)

National Drug Data File Plus Source Vocabulary

Uploaded: 6/10/22

projects	classes
1	31,026

STEP 3: Findability aspects

- Will data be identified by a **persistent identifier**? → *Will they be deposited to a trusted repository?* (next step - Accessibility)

- What **metadata** will be created? Will they be machine-readable?

- What **disciplinary or general standards** will be followed to create metadata? What if they do not yet exist?

- *Will metadata be added **directly into the files** or will they be produced in **another program or document**? (Only CESSDA)*

STEP 3: Findability aspects

- Will data be identified by a **persistent identifier**? → *Will they be deposited to a trusted repository?* (next step - Accessibility)

- What **metadata** will be created? Will they be machine-readable?

- What **disciplinary or general standards** will be followed to create metadata? What if they do not yet exist?

- *Will metadata be added **directly into the files** or will they be produced in **another program or document**? (Only CESSDA)*

- *Will search **keywords** be provided in the metadata? (Only HE)*

Figures   References   **Related**   Information

Dataset Search

| battery | 🔍 |

📄 U.S. quarterly **battery** electric vehicle sales by model 2020-2021

📄 Electric vehicles - **battery** range of select models in U.S. 2022

📄 **Battery** electric vehicles in use - worldwide 2016-2020

📄 Electric vehicles - global lithium-ion **battery** pack costs 2011-2030

📄 **Battery** Market Size, Share & Trends Analysis Report By Product (Lead Acid, Li-ion, Nickle Metal Hydride, Ni-cd), By Application (Automotive, Industrial, Portable), By Region, And Segment Forecasts, 2020 - 2027

📄 Projected **battery** demand worldwide by application 2020-2030

📄 Lithium-ion **Battery** Market Size, Share & Trends Analysis Report By Product (LCO, LFP, NCA, LMO, LTO, NMC), By Application (Consumer Electronics, Energy Storage Systems, Industrial), By Region, And Segment Forecasts, 2022 - 2030

📄 Lithium ion **battery** production capacity by company 2028

📄 Share of the lithium-ion **battery** production capacity worldwide by country 2020&2025

📄 Predicted average **battery** capacities in EVs worldwide 2017-2025

Save Word

bat·tery | \ ˈba-t(ə-)rē 🔊 \

plural **batteries**

Definition of *battery*

1 **a** : the act of beating someone or something with successive blows **:** the act of battering (see BATTER entry 1 sense 1)

   b *law* : an offensive touching or use of force on a person without the person's consent
   // evidence that supports a charge of *battery*
   — compare ASSAULT entry 1 sense 2a

2 [Middle French *batterie*, from *battre* to beat] *military*

   a : a grouping of artillery pieces for tactical (see TACTICAL sense 1a(1)) purposes
   // a *battery* of cannon from the Revolutionary War

   b : the guns of a warship
   // the starboard *battery*

3 *military* : an artillery (see ARTILLERY sense 2b) unit in the army equivalent to a company (see COMPANY entry 1 sense 2b)

4 **a** : a combination of apparatus for producing a single electrical effect
   // a *battery* of generators

   b : a group of two or more cells (see CELL sense 5) connected together to furnish electric current
   also : a single cell that furnishes electric current
   // need to replace the flashlight's *batteries*

   c **batteries** *plural* : level of energy or enthusiasm
   // needs a vacation to recharge her *batteries*

5 **a** **(1)** : a number of similar articles, items, or devices arranged, connected, or used together **:** SET, SERIES
   // ran through a *battery* of tests
   // a *battery* of filing cabinets

      (2) : a series of cages or compartments for raising or fattening poultry —often used before another noun
   // *battery* chickens
   // *battery* farming

   b : a usually impressive or imposing group **:** ARRAY
   // a *battery* of specialists

6 : the position of readiness of a gun for firing
   // the gun would not return to *battery*
   — *Infantry Journal*

7 *baseball* : the pitcher and catcher of a team
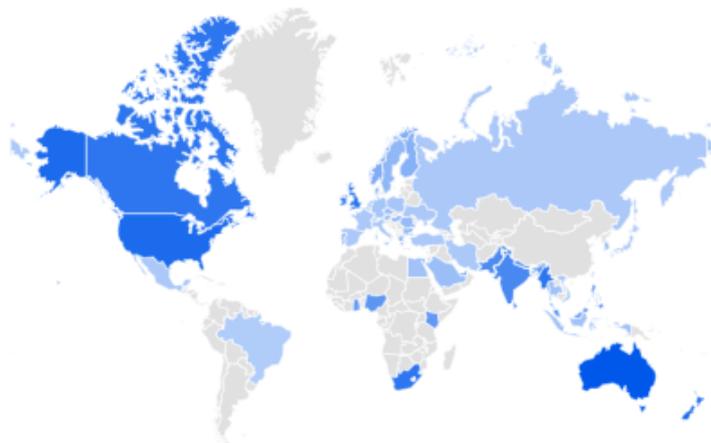   // one of the greatest *batteries* in baseball history

battery

Zanimanje po območjih (?)

Območje ▾

☐ Vključi območja z nizkim obsegom iskanja

	Sorodne teme (?)	Naraščajoče ▾	
1	Litij-železov-fosfatni akumulator ...	+ 60%
2	Električno vozilo - Tema	+ 50%
3	Kilovatna ura - Merska enota: en...	+ 40%
4	Electric vehicle battery - Tema	+ 40%

	Sorodne poizvedbe ...	Naraščajoče ▾	
1	battery percentage iphone 13	Izjemna rast
2	iphone 13 pro battery	+ 4.600%
3	show battery percentage iphone ...	+ 4.350%
4	iphone 13 battery mah	+ 4.350%

battery,accumulator

1800 - 2019 ▾   English (2019) ▾   Case-Insensitive   Smoothing ▾
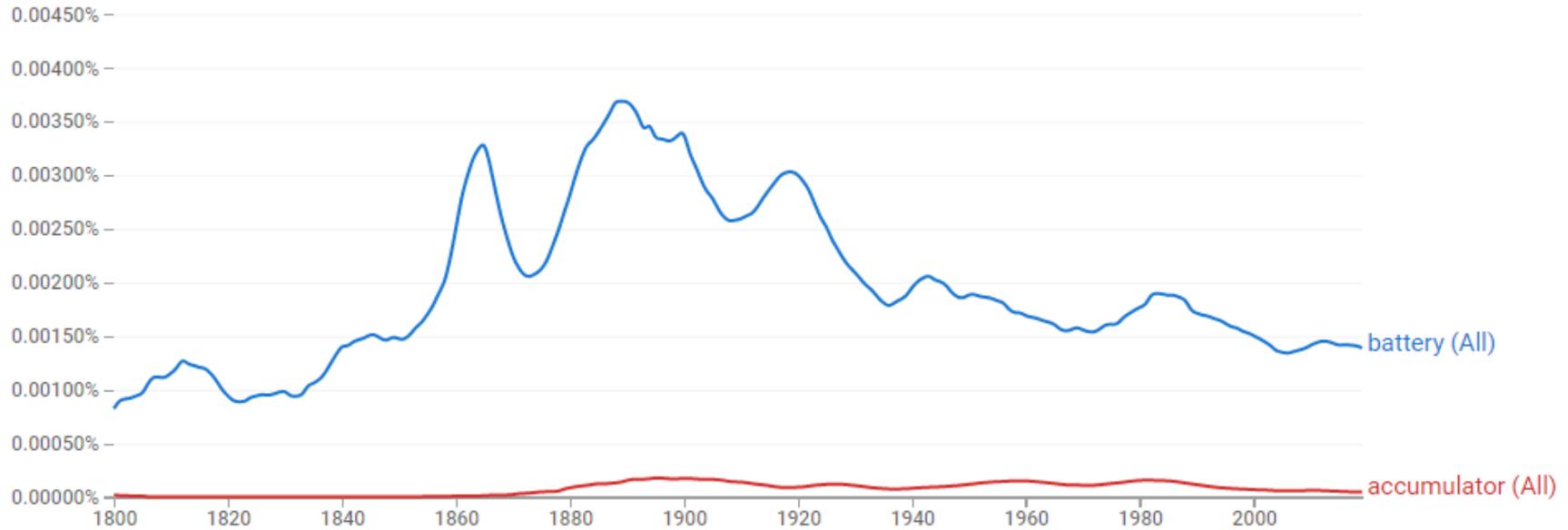
(click on line/label for focus, right click to expand/contract wildcards)

Search in Google Books

battery > 1800 - 1835   1836 - 1888   1889 - 1901   1902 - 1987   1988 - 2019   English (2019)

accumulator > 1800 - 1885   1886 - 1956   1957 - 1967   1968 - 1992   1993 - 2019   English (2019)

STEP 4: Accessibility aspects

- Will data be deposited in a **trusted repository**?

A repository that is either:

- **certified** (e.g., CoreTrustSeal, DIN 31644, ISO 16363),

- **domain-specific and endorsed by the target research community** (e.g., HEPData, Crystallography Open Database, PubChem …),

- **general or institutional repository that has characteristics of a trusted repository** (e.g., Zenodo).

https://dirrosdata.ctk.uni-lj.si/repozitoriji/zaupanja-vredni-repozitoriji/

STEP 4: Accessibility aspects

- Will data be deposited in a **trusted repository**?

- Does the repository issue a **persistent identifier**?

- Does it use a **free and standardized access protocol** (e.g., HTTP, FTP, SMTP, OAI-PMH …)?

STEP 4: Accessibility aspects

- Will data be deposited in a **trusted repository**?

- Does the repository issue a **persistent identifier**?

- Does it use a **free and standardized access protocol**?

- Will access to the data be **fully open** or **(partially) restricted**?

- What are the **conditions** for embargo/restricted access?

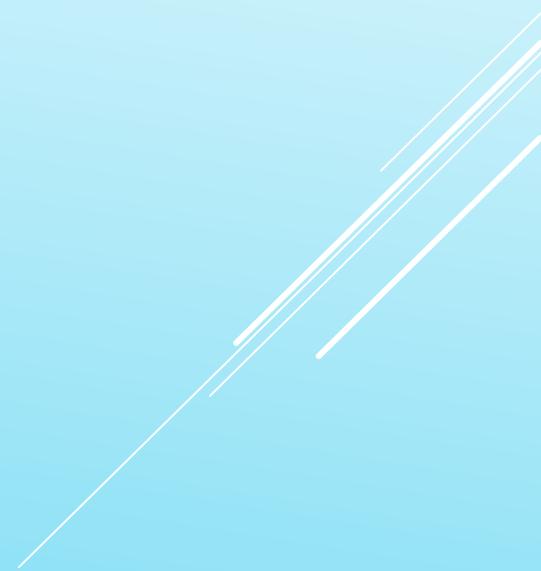https://dirrosdata.ctk.uni-lj.si/raziskovalni-podatki/upravicene-izjeme-od-odprtosti/

Justified exceptions from the openness principle

1. Protection of results because of **legitimate interests** or **other constraints** (confidentiality, trade secrets, security rules, EU competitive interests or intellectual property rights)

Justified exceptions from the openness principle

1. Protection of results because of **legitimate interests** or **other constraints** (confidentiality, trade secrets, security rules, EU competitive interests or intellectual property rights)

2. Protection of **personal data** (GDPR)

Justified exceptions from the openness principle

1. Protection of results because of **legitimate interests** or **other constraints** (confidentiality, trade secrets, security rules, EU competitive interests or intellectual property rights)

2. Protection of **personal data** (GDPR)

3. Data under **third party license**

Justified exceptions from the openness principle

1. Protection of results because of **legitimate interests** or **other constraints** (confidentiality, trade secrets, security rules, EU competitive interests or intellectual property rights)

2. Protection of **personal data** (GDPR)

3. Data under **third party license**

Accessibility constraint: **large data**

Justified exceptions from the openness principle

1. Protection of results because of **legitimate interests** or **other constraints** (confidentiality, trade secrets, security rules, EU competitive interests or intellectual property rights)

2. Protection of **personal data** (GDPR)

3. Data under **third party license**

Accessibility constraint: **large data**

„As open as possible, as closed as necessary"

Possibilities for restricted access

1. **Embargo** → possible in, e.g., Zenodo, Dryad, Figshare …

Possibilities for restricted access

1. **Embargo** → possible in, e.g., Zenodo, Dryad, Figshare …

2. **„Right to be forgotten"** → metadata must remain even if the data is deleted/restricted

Possibilities for restricted access

1. **Embargo** → possible in, e.g., Zenodo, Dryad, Figshare …

2. **„Right to be forgotten"** → metadata must remain even if the data is deleted/restricted

3. **Restriction of access by physical or IT means**

Possibilities for restricted access

1. **Embargo** → possible in, e.g., Zenodo, Dryad, Figshare …

2. **„Right to be forgotten"** → metadata must remain even if the data is deleted/restricted

3. **Restriction of access by physical or IT means**

The publications resulting from such data must include:

- a **description of the restrictions** on the data,

- **all necessary information** required to apply for access to the data and the conditions under which access will be granted.

STEP 4: Accessibility aspects

- Will data be deposited in a **trusted repository**?

- Does the repository issue a **persistent identifier**?

- Does it use a **free and standardized access protocol**?

- Will access to the data be **fully open** or **(partially) restricted**?

- What are the **conditions** for embargo/restricted access?

- Is there a **specific software** needed to access or read the data? Is the software or a reference to it included with the data?

STEP 5: Interoperability aspects

- Will you use **established software, hardware and computer code** to collect your data?

STEP 5: Interoperability aspects

- Will you use **established software, hardware and computer code** to collect your data?

- Will you provide **references to other data** (datasets from your previous research or other projects, datasets of other authors)?

STEP 5: Interoperability aspects

- Will you use **established software, hardware and computer code** to collect your data?

- Will you provide **references to other data** (datasets from your previous research or other projects, datasets of other authors)?

- Will you describe your data and metadata using established **standard or field-specific terminologies/ontologies** (i.e. structured controlled vocabularies)?

- If you generate **project-specific ontologies or vocabularies**, will you provide mappings to more commonly used ontologies?

- Will you **openly publish** the generated ontologies or vocabularies to allow reusing, refining or extending them?

STEP 6: Reusability aspects

- What **documentation** will be created to validate data and facilitate their re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

STEP 6: Reusability aspects

- What **documentation** will be created to validate data and facilitate their re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

- Will **domain-specific standards** be used to document the provenance of the data?

Provenance standards: examples

Joule   ⬡ **CellPress**

Editorial
Standardized Battery Reporting Guidelines

Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative

BMJ 2003 ; 326 doi: https://doi.org/10.1136/bmj.326.7379.41 (Published 04 January 2003)
Cite this as: *BMJ* 2003;326:41

| Article | Related content | Metrics | Responses |

*Patrick M Bossuyt, professor of clinical epidemiology [a] (stard@amc.uva.nl), Johannes B Reitsma, clinical epidemiologist [a],
David E Bruns, editor [b], Constantine A Gatsonis, professor of medical science (biostatistics) and applied mathematics [c],
Paul P Glasziou, professor of evidence based practice [d], Les M Irwig, professor of epidemiology [e],
Jeroen G Lijmer, clinical epidemiologist [a], David Moher, director [f], Drummond Rennie, deputy editor [g],
Henrica C W de Vet, professor of epidemiology [h]*

for the STARD steering group

Published: 01 December 2001
Minimum information about a microarray experiment (MIAME)—toward standards for microarray data

Alvis Brazma ✉, Pascal Hingamp, John Quackenbush ✉, Gavin Sherlock, Paul Spellman, Chris Stoeckert,
John Aach, Wilhelm Ansorge, Catherine A. Ball, Helen C. Causton, Terry Gaasterland, Patrick Glenisson,
Frank C.P. Holstege, Irene F. Kim, Victor Markowitz, John C. Matese, Helen Parkinson, Alan Robinson, Ugis
Sarkans, Steffen Schulze-Kremer, Jason Stewart, Ronald Taylor, Jaak Vilo & Martin Vingron

STEP 6: Reusability aspects

- What **documentation** will be created to validate data and facilitate their re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

- Will **domain-specific standards** be used to document the provenance of the data?

- How will the data be **licensed**?

https://dirrosdata.ctk.uni-lj.si/raziskovalni-podatki/licenciranje-podatkov/
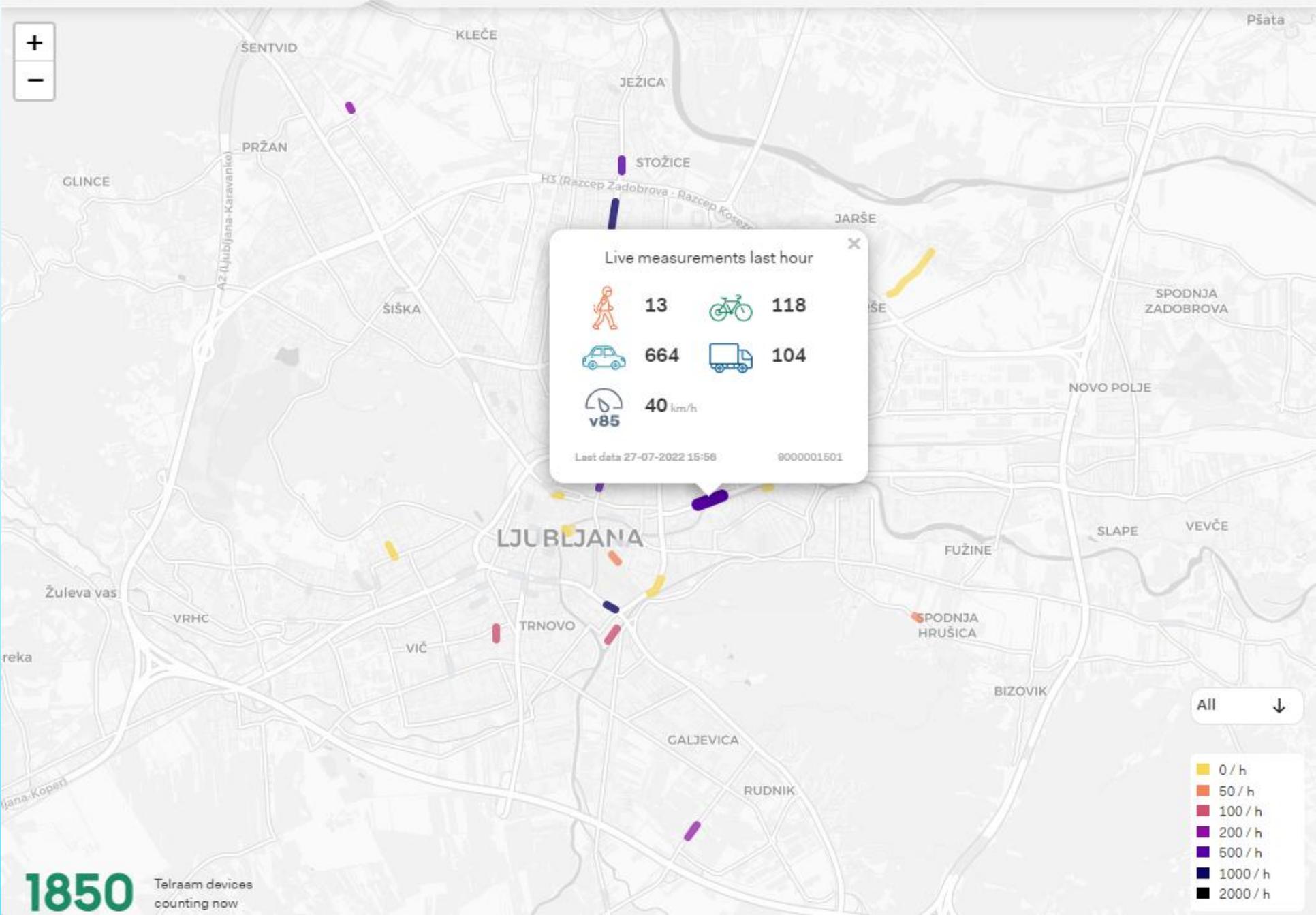
STEP 6: Reusability aspects

- What **documentation** will be created to validate data and facilitate their re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

- Will **domain-specific standards** be used to document the provenance of the data?

- How will the data be **licensed**?

- How will **data quality assurance** be performed? → *Quality assurance at the post-collection stage*

STEP 6: Reusability aspects

- What **documentation** will be created to validate data and facilitate their re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

- Will **domain-specific standards** be used to document the provenance of the data?

- How will the data be **licensed**?

- How will **data quality assurance** be performed? → *Quality assurance at the post-collection stage*

- *What is your strategy for **versioning** your data files? (CESSDA)*

STEP 6: Reusability aspects

- What **documentation** will be created to validate data and facilitate their re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

- Will **domain-specific standards** be used to document the provenance of the data?

- How will the data be **licensed**?

- How will **data quality assurance** be performed? → *Quality assurance at the post-collection stage*

- *What is your strategy for **versioning** your data files? (CESSDA)*

- *How should data be **cited** when reused? (CESSDA)*

Telraam

Live measurements last hour

🚶 13   🚲 118

🚗 664   🚚 104

v85   40 km/h

Last data 27-07-2022 15:58   9000001501

All ↓

- 🟨 0 / h
- 🟧 50 / h
- 🟥 100 / h
- 🟪 200 / h
- 🟪 500 / h
- ⬛ 1000 / h
- ⬛ 2000 / h

1850 Telraam devices counting now

STEP 7: Costs & resources

- What will be **direct and indirect costs** for making data or other research outputs FAIR (e.g. costs related to management, storage, archiving, re-use, security, etc.)?

- How will these be **covered**?

STEP 7: Costs & resources

- What will be **direct and indirect costs** for making data or other research outputs FAIR (e.g. costs related to management, storage, archiving, re-use, security, etc.)?

- How will these be **covered**?

- Who will be **responsible** for data management in your project?

STEP 7: Costs & resources

- What will be **direct and indirect costs** for making data or other research outputs FAIR (e.g. costs related to management, storage, archiving, re-use, security, etc.)?

- How will these be **covered**?

- Who will be **responsible** for data management in your project?

- How will **long-term preservation** be ensured?

STEP 7: Costs & resources

- What will be **direct and indirect costs** for making data or other research outputs FAIR (e.g. costs related to management, storage, archiving, re-use, security, etc.)?

- How will these be **covered**?

- Who will be **responsible** for data management in your project?

- How will **long-term preservation** be ensured?

- *What is the **potential value** of long-term data preservation? (ERC)*

STEP 8: Data safety & security

- What provisions are or will be in place for data security (including **data recovery** as well as **secure storage/archiving and transfer** of sensitive data)?

- Will the data be safely stored in **trusted repositories** for long term preservation and curation?

STEP 8: Data safety & security

- What provisions are or will be in place for data security (including **data recovery** as well as **secure storage/archiving and transfer** of sensitive data)?

- Will the data be safely stored in **trusted repositories** for long term preservation and curation?

- *How, where and at what intervals will the data & metadata be* **backed-up***? (CESSDA)*

STEP 8: Data safety & security

- What provisions are or will be in place for data security (including **data recovery** as well as **secure storage/archiving and transfer** of sensitive data)?

- Will the data be safely stored in **trusted repositories** for long term preservation and curation?

- *How, where and at what intervals will the data & metadata be **backed-up**? (CESSDA)*

- *How will **sensitive data & metadata** be protected? (if applicable) How will **access** to this information be managed? (CESSDA)*

STEP 9: Ethics

The ERC DMP does not include a section on ethics. It is the responsibility of the PI to inform the ERCEA Ethics Team of any ethics issues regarding the collection, processing, sharing and storage of data in relation to the project. The PI can also be asked to submit a separate Ethics DMP.

- Are there, or could there be, any **ethics or legal issues** that can have an impact on data sharing?

- Does your project require **approval by a local ethics committee**?

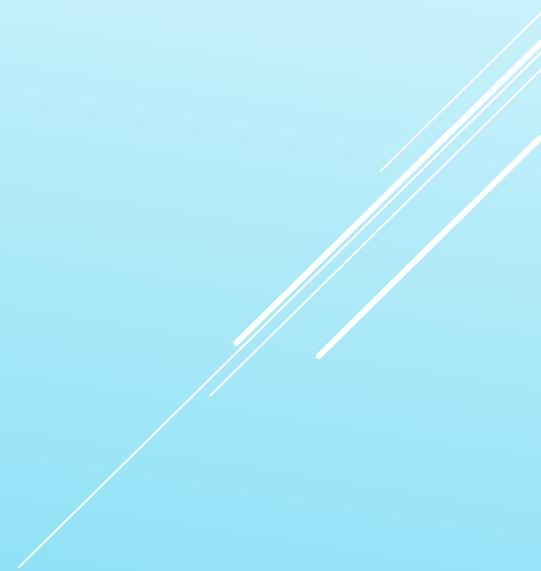- If **informed consent** is needed, how will you obtain it?

STEP 9: Ethics

The ERC DMP does not include a section on ethics. It is the responsibility of the PI to inform the ERCEA Ethics Team of any ethics issues regarding the collection, processing, sharing and storage of data in relation to the project. The PI can also be asked to submit a separate Ethics DMP.
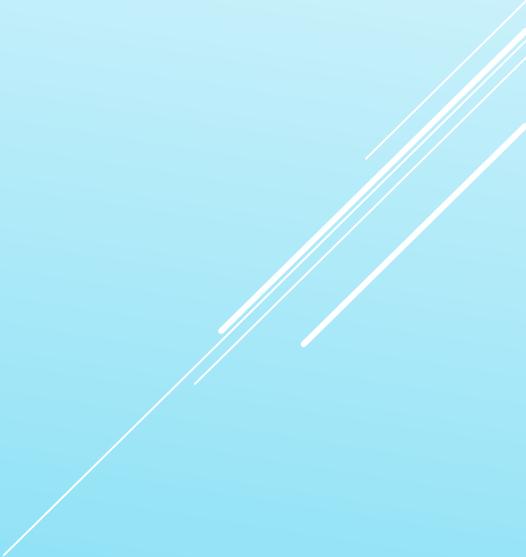
- Are there, or could there be, any **ethics or legal issues** that can have an impact on data sharing?

- Does your project require **approval by a local ethics committee**?

- If **informed consent** is needed, how will you obtain it?

- Is there any **confidential information** that requires special treatment and/or limits the access to it during/after the project? How will the material be protected during/after the project?

STEP 9: Ethics

The ERC DMP does not include a section on ethics. It is the responsibility of the PI to inform the ERCEA Ethics Team of any ethics issues regarding the collection, processing, sharing and storage of data in relation to the project. The PI can also be asked to submit a separate Ethics DMP.

- Are there, or could there be, any **ethics or legal issues** that can have an impact on data sharing?

- Does your project require **approval by a local ethics committee**?

- If **informed consent** is needed, how will you obtain it?

- Is there any **confidential information** that requires special treatment and/or limits the access to it during/after the project? How will the material be protected during/after the project?

- What are the agreements with **other stakeholders/collaborators**?

Questions?

Thank you for your attention!