Preparing your Horizon Europe Data Management Plan

- DeiC webinar 14 June 2023

Susanne den Boer · University of Copenhagen · susanne.denboer@adm.ku.dk

1 Horizon Europe's requirements for research data management

Research data management (RDM)

Research data:
Information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation.

E.g. statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images.

The focus is on research data that is available <u>in digital form</u>.

Horizon Europe's requirements for data management

In case of successful applications:

- Establish and maintain a **Data Management Plan (DMP)**
- **Manage data responsibly**, in accordance with legislation, contracts, information security requirements, etc.
- Manage data according to the **FAIR principles**
- **Make data accessible by default**, but "as open as possible, as closed as neccessary"

READ:

Horizon Europe (HORIZON) – Programme Guide
https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf

EU Grants AGA – Annotated Model Grant Agreement
https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/aga_en.pdf

Data management plan (DMP)

Document with questions about data management

Planning tool: checklist to thorough consider data management at project start

Collaboration tool: to align expectations and record decisions between project members

Manual: projects' guide for data management that project members can consult throughout the project

A DMP is not something you "sell" to others (e.g. administrators in the EC) - it is not a proposal!

- A DMP can create value for the research project. Audience for the DMP = project members.

- EC wants to see evidence of thoroughly considered data management, with consensus and clear division of responsibilities between project members.

Susanne P.A. den Boer

Requirements for DMPs in Horizon Europe

- **One DMP per funded project**, also in large collaborative projects, approx. 5 pages for individual fellowships, 15 pages for collaborative projects.

- Beneficiaries must **submit a DMP as a deliverable,** normally by month 6.

- An **updated DMP must also be delivered** mid-project (for projects longer than twelve months) and at the end of the project.

- DMP as a living document: **update the DMP whenever significant changes arise**. E.g. the generation of new data, changes in data access policies, decision to file for a   patent, changes in consortium composition.

- Beneficiaries are encouraged to **share their DMP openly**

How DMPs are reviewed (in Horizon 2020)

- REA project officers and/or external reviewers review the DMP.

- REA project officers use assessment grid corresponding to the H2020 template.

- DMP needs to demonstrate your strategy for data management, with consensus between project members. It is not a sales pitch!

- Project officer might give feedback (or not).

- Project officer might request you to resubmit the DMP, if information is missing or your strategy for data management is unclear.

1. DATA SUMMARY

1.a Is header information provided (action ID, acronym, DMP version, name of the DMP responsible)?   Yes ☐ N/A ☐ No ☐

1.b Are the purpose of the data collection and its relation to project objectives explained?
                                                                Yes ☐ N/A ☐ No ☐

1.c Are data types and formats specified?   Yes ☐ N/A ☐ No ☐

1.d Is the expected volume of the data estimated?   Yes ☐ N/A ☐ No ☐

1.e Is reuse of pre-existing data described including its origin?   Yes ☐ N/A ☐ No ☐

1.f Is data utility outlined (to whom will the data be useful)?   Yes ☐ N/A ☐ No ☐

Recommendations:

----

2. FAIR DATA
2.1. Making data findable, including provisions for metadata

2.1.a Will the data be assigned a unique and persistent identifier and registered in a searchable resource?   Yes ☐ N/A ☐ No ☐

2.1.b Are data naming conventions described?   Yes ☐ N/A ☐ No ☐

2.1.c Will the data be described with rich metadata (following standard practises in the field)?   Yes ☐ N/A ☐ No ☐

2 An introduction to HEU's data management plan template

Questions in the HEU DMP on:
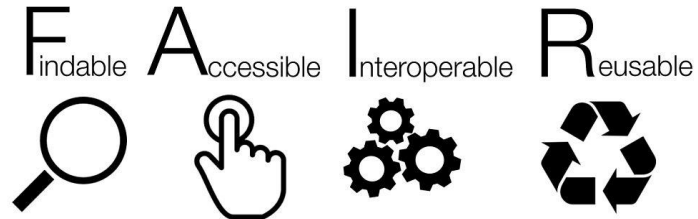
Data summary:
- Type and origin of data
- Reuse of existing data

The FAIR principles for data sharing:
- Making data and metadata findable
- Making data and metadata accessible
- Making data interoperable
- Increasing data reuse

Other data management actions:
- Other research outputs
- Resources, costs and responsibilities
- Data security and ethics

F indable   A ccessible   I nteroperable   R eusable

https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf

Data and Project summary

Connected to, e.g.:
Q2.2 Making data accessible
Q2.4 Making data reusable
Q4.1 Allocation of resources
Q5.1 Data security
…and more

Data type

Personal data
Confidential data
Data with commercial potential
Physical material
Big data

Collaborator 'type'

Within university
Outside university
Outside DK/EU
Industry partners
Citizens

Connected to, e.g.:
Q1.1 Reuse existing data
Q4.3 Responsibilities
Q5.1 Data security
…and more

Connected to, e.g.:
Q1.6 Data utility
Q2.4 Making data reusable
Q4.4 Long term preservation
…and more

Inside discipline
Outside discipline
Interest groups
Policy makers
General public

Target groups

Research Integrity
Legal compliance, incl. GDPR
Ethical requirements
Institutional Policies
Funder, publisher policies

External requirements

Can influence, e.g.:
Q6.1 Ethics
Q6.2 Informed consent
Q7.1 Procedures for data management
…and more

Data summary - a question and an answer

> *What types and formats of data will the project generate or re-use?*

> *What is the expected size of the data that you intend to generate or re-use?*

> *What is the origin/provenance of the data, either generated or re-used?*

WP	Data type	Format	Size	Origin
1A	Interviews: interview guide, audio recordings, transcripts, metadata	PDF, MP3, DOC, TXT	<1TB	22 staff members of the Social Democrats in Denmark, who previously agreed to participate in our study in response to a survey sent out in 2021. Interviews to be recorded one week before the elections
1B	Measurements: heart rate, blood pressure	CSV	<5MB	As above. Measurements to be collected one week before the elections.
2	UN Sustainable Development Goals (SDG) country profile for Denmark	CSV	<5MB	All data were downloaded from the UN SDG indicators website: https://unstats.un.org/sdgs/dataportal/countryprofiles/dnk, where the data are freely available for reuse (see UNdata's "Terms of Use").

The FAIR principles

http://www.lego.com/404
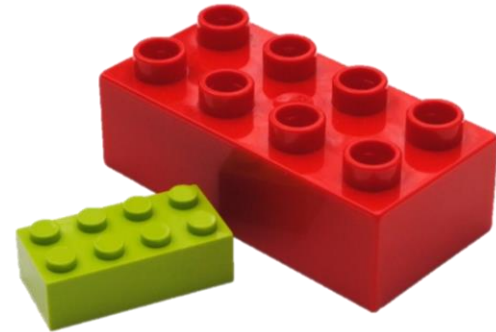
The FAIR principles

Drafted by researchers, funders, publishers in 2016, described in Nature article: *"The FAIR Guiding Principles for scientific data management and stewardship"* https://doi.org/10.1038/sdata.2016.18

The FAIR principles describe actions to ensure that humans (or machines)...

- can find out that your research / data exist (**F**indable)
- know how to get access to your data (**A**ccessible)
- can open your data and work with it (**I**nteroperable)
- understand how the data were created, and can be reused (**R**eusable)

Please note:

- FAIR data ≠ Open data!
- All open data must be FAIR, but not all FAIR data must be open.

Accessible

1. To provide access to data, upload your data in a DATA REPOSITORY

Discipline or data type specific repository/db

Generalist repository

Institutional repository

Browse repositories: www.re3data.org

Accessible

2. Indicate MODE OF ACCESS

> **Open Access**

> **Embargoed Access**

> **Restricted Access**

> **No Access**

This is not just about access to DATA but also access to METADATA

May 12, 2020 — Dataset — Embargoed Access

Transient Electromagnetic data from the Acoculco area in Mexico: raw data

Benediktsdóttir, Ásdís; Hersir, Gylfi Páll; Arango-Galván, Claudia; Romo, Jose Manuel

The dataset is composed of raw data from TerraTEM (from Monex GeoScope, single loop) from the Acoculco area, Mexico and their locations in a .csv file.

The data were gathered under the European Union's Horizon 2020 research and innovation programme under grant agreement No. 727550, and by the Mexican Energy Sustainability Fund CONACYT-SENER, Project 2015-04-268074.

Files ⌄

⊘ Embargoed Access
Files are currently under embargo but will be publicly accessible after May 1, 2024.

March 2, 2019 — Dataset — Open Access

EvoBib: A Bibliographic Database for Historical Linguistics

Johann-Mattis List

The work on this bibliographical database started when I was a member of the EvoClass research project at Heinrich Heine University Düsseldorf. It is actually nothing more than a collection of all titles (currently 3335 entries, few have been removed, as they turned out to be invalid) which I stumbled over when pursuing my research in quantitative historical linguistics. The database thus mainly contains work on lexicostatistics, methodological aspects of the comparative method, and the application of biological methods in historical linguistics.

The most recent version of the database can be accessed at http://bibliography.lingpy.org, here, I offer the database file in BibTex format, to make sure that different versions are being tracked.

Files (1.4 MB) ⌄

Name	Size	
evobib-converted.bib	1.4 MB	⬇ Download

September 11, 2020 — Other — Closed Access

Multi-Point Compositional Measurements of Solar Wind and Transient Phenomena

Yeimy J. Rivera; Aleida Higginson; Susan T. Lepri; Nicholeen M. Viall

White paper for the Heliophysics 2050 workshop on the necessity for continuous 4π coverage of compositional measurements to study plasma between the Sun and heliosphere to gain a comprehensive understanding of the solar wind's method of release and its energization.

Files ⌄

🔒 Closed Access
Files are not publicly accessible.

Citations ❓ 0 ⌄

Show only: ☐ Literature (0) ☐ Dataset (0) ☐ Software (0) ☐ Unknown (0)
☐ Citations to this version   [Search]

No citations.

Other — Restricted Access

IPBES Long-term Vision on Data and Knowledge Management

IPBES task force on knowledge and data

Editor(s)

Debora Pignatari Drucker; Dave Thau; Rainer M. Krug; Joy A Kumagai; Aidin Niamir

The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) Plenary extended the mandate of the task force on knowledge and data for the implementation of objective 3 (a) of the IPBES work programme up to 2030, regarding advanced work on knowledge and data (decision IPBES-7/1, section IV). The Plenary, at its 8th session approved the interim workplan of the task force on knowledge and data for the intersessional period 2021–2022 (decision IPBES/8/11, section IV), which includes the development of a long-term vision on data management (IPBES/8/INF/12). Pursuant to that work plan, the IPBES task force on knowledge and data prepared a long-term vision document on knowledge and data management and delivery, in line with the approved IPBES rolling work programme.

The IPBES long-term vision on data and knowledge management describes an overarching vision supported by two essential tracks for knowledge and data management and delivery with corresponding targets outlining possible paths to reaching the vision.

Files ⌄

🔑 Restricted Access
You may request access to the files in this upload, provided that you fulfil the conditions below. The decision whether to grant/deny access is solely under the responsibility of the record owner.

Please provide your name, affiliation, email address, and reason for accessing the long-term vision on data and knowledge management.

[✎ Request access...]

Susanne P.A. den Boer

Accessible – a question and an answer

> *Will the data be deposited in a trusted repository?*
> *Have you explored appropriate arrangements with the identified repository where your data will be deposited?*
> *Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement?*
> *How long will the data remain available and findable?*

"Data and metadata will be made freely available via the repository Zenodo.

(Meta)data are retrievable using the open, free and universally implementable protocols OAI-PMH and REST API.

The metadata will be licensed under public domain (CC0) and no authorization is necessary to retrieve the metadata.

Metadata will be accessible even if the data are no longer available, stored in high availability database servers at CERN. CERN guarantees a minimum storage period of 20 years for data and metadata."

Fictional answer

Accessible – a question and an answer

> *Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions.*

"The different types of data that are generated during the project are open by default with the following general exceptions:

• <u>copyright and permissions for reusing third-party data sets</u>
We will process and combine input data from many different sources. Such repurposed data (e.g. model output data) can only be made open if any of the underlying data (e.g. model input data) is open, too.

• <u>personal data treatment and confidentiality issues</u>
Datasets referring to the quality and quantity of certain elements at risk, such as people and critical infrastructures, are not open by default as their publication may pose privacy, ethical or security risks.

• <u>data-driven business model</u>
Data that is exploited commercially through the MyClimateService.eu marketplace will not be made open."

Answer adapted from DOI: 10.5281/zenodo.3970982

Susanne P.A. den Boer

Findable

1. **METADATA: Use rich metadata to describe your data**

Metadata help humans / machines understand your data

Metadata allow carrying out searching, sorting, prioritising tasks

2. Assign PERSISTENT IDENTIFIERS (PIDS) to your data sets:

Unique, unbreakable internet links to data sets and other objects
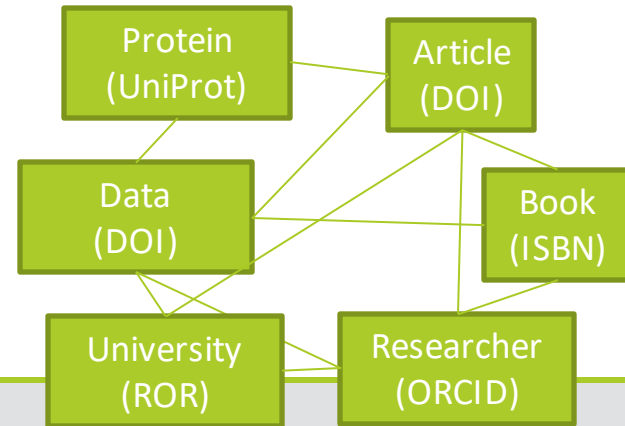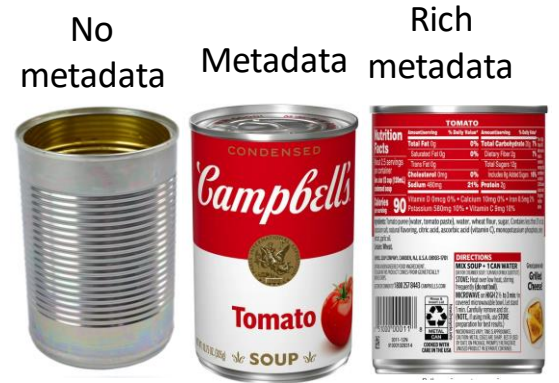
Why are they important?

Unique identifiers remove doubt as to what is meant

Others can use identifiers to cite your work

Identifiers can link research objects together, improving findability.

No metadata   Metadata   Rich metadata

Protein (UniProt)

Article (DOI)

Data (DOI)

Book (ISBN)

University (ROR)

Researcher (ORCID)

Susanne P.A. den Boer

Findable

3. MAKE METADATA SEARCHABLE: Add metadata to the data set in a repository where they are indexed by search engines

zenodo

Dialogue and Argumentation for Cultural Literacy Learning in Schools: Multilingual Data Corpus

Data set title →

Rapanta, Chrysi; Cascalheira, Dilar; Gil, Beatriz; Gonçalves, Cláudia; Garcia, D'Jamila; Morais, Rita; Pereira, João Rui; Čermáková, Anna; Maine, Fiona; Peck, Julia; Brummernhenrich, Benjamin; Jucks, Regina; Petronytė, Miglė; Valančienė, Daina; Juskiene, Vaiva; Badaukienė, Ramunė; Eigminienė, Dainora; Stonkuviene, Irena; Zaleskienė, Irena; Garcia-Mila, Mercè; Remesal, Ana; Castells, Nuria; Gilabert, Sandra; Miralda-Banda, Andrea; Luna, Jose; Vrikki, Maria; Evagorou, Maria; Chatzianastasi, Maria; Karousiou, Christiana; Papanastasiou, Elena; Stylianou-Georgiou, Agni; Rodosthenous, Marina; Talli, Cedar; Cohen, Irit; Shalom Greenberg, Chaim; Bar, Noa; Sarfati, Neta; Schwarz, Baruch

Author names →

Unique identifier (ORCID)

This dataset is the Multilingual Corpus of the DIALLS (Dialogue and Argumentation for Literacy Learning in Schools) project (dialls2020.eu/) and consists of a set of transcripts of classroom interactions of students from ages 5 to 15 years old. These classroom interactions took place in seven DIALLS participant countries (UK, Portugal, Germany, Lithuania, Spain, Cyprus, and Israel).

Data set description →

The corpus is a set of 202 transcripts in the participant countries' native language (English, Portuguese, German, Lithuanian, Catalan, Cypriot Greek, and Hebrew). The transcripts in each native language range from a maximum of 35 for Hebrew (more than 17% of the overall corpus) to a minimum of 19 transcripts for Cypriot Greek (10% of the corpus). More than 50% of the transcripts in a language different from English (90 transcripts) have associated their English translation. The topic of the project is cultural literacy through dialogue and argumentation in school children.

Files (12.0 MB)		⌄
Name	Size	
1.DIALLS_Multilingual_Corpus_description.V2.pdf	408.5 kB	👁 Preview ⬇ Download
md5:5afd69a0de34782cfd3cca14fa4b9422 ⓘ		

Data set names →

Publication date:
May 7, 2021

DOI:
DOI 10.5281/zenodo.4742176 ← Unique identifier (DOI)

Keyword(s):
Multilingual corpus Education Classroom Dialogues European Identity Cultural Literacy Learning Dialogical teaching Argumentation Cultural Texts Empathy Dialogicity

→ Keywords

Grants:
European Commission:
● DIALLS - DIalogue and Argumentation for cultural Literacy Learning in Schools (770045)

Related identifiers:
Cites
https://dialls2020.eu/wp-content/uploads/2019/06/D3.1-final.pdf (Project deliverable)
https://dialls2020.eu/wp-content/uploads/2019/01/D2.3_final-submitted.pdf (Project deliverable)

Susanne P.A. den Boer

Findable – a question and an answer

> *Will data be identified by a persistent identifier?*
> *Will rich metadata be provided to allow discovery?*
> *Will metadata be offered in such a way that it can be harvested and indexed?*

"All research data are stored in the repository Zenodo in the community "ROMSOC H2020-MSCA-ITN". Zenodo allows us to describe our data with metadata (including keywords), which it registers and indexes in searchable resources. Zenodo also automatically assigns a DOI (Digital Object Identifier) to every published record and each of its versions, to allow a globally unique and persistent identification of the data. Research data can be linked to the corresponding publications and vice versa via their DOIs."

Answer inspired by DMP: https://zenodo.org/record/3459510

Interoperable

1. Use OPEN FILE FORMATS as much as possible

= Formats that can be used and implemented by anyone

Examples
CSV for tabular data
RTF for textual data
TIFF for images
NetCDF for geospatial data

Interoperable

1. Use OPEN FILE FORMATS as much as possible

= Formats that can be used and implemented by anyone

2. Use STANDARDS common in your discipline

For example

- **Minimum information:** metadata that as a minimum should be included to describe a data set

Minimum information about a microarray experiment (MIAME)–toward standards for microarray data

1. Experimental design: the set of hybridisation experiments as a whole

This section describes the experiment, which may consist of one or more hybridisations, as a whole. Normally 'experiment' should include a set of hybridisations which are inter-related and address a common question. For instance, it may be all the hybridisations related to research published in a single paper.

a) author (submitter), laboratory, contact information, links (URL), citations
b) type of the experiment - maximum one line, for instance:
normal vs. diseased comparison
treated vs. untreated comparison
time course
dose response
effect of gene knock-out
effect of gene knock-in (transgenics)
shock
(multiple types possible)

c) experimental variables, i.e. parameters or conditions tested (e.g., time, dose, genetic variation, response to a treatment or compound)
d) single or multiple hybridisations.
For multiple hybridisations:

serial (yes/no)
    o type (e.g., time course, dose response)
grouping (yes/no)
    o type (e.g., normal vs. diseased, multiple tissue comparison)
Relationships between all the samples, arrays and hybridisations in the experiment. Each sample, each array, and each hybridisation should be given a unique ID, and all the relationships should be listed (with appropriate comments where necessary). For instance:
Samples: S1, S2, S3

Interoperable

1. Use OPEN FILE FORMATS as much as possible

= Formats that can be used and implemented by anyone

2. Use STANDARDS common in your discipline

For example
- **Minimum information:** metadata that as a minimum should be included to describe a data set
- **Vocabularies:** predefined and authorized terms for metadata

DDI Controlled Vocabulary for Time Zone

Description

Time zone specification as an offset from UTC (Coordinated Universal Time) in terms of hours and minutes.

Code List

Value of the Code	Descriptive Term of the Code	Definition of the Code
-12.00	UTC- 12:00	Baker Island, Howland Island (both uninhabited)
-11.00	UTC- 11:00	Samoa Standard Time (SST)
-10.00	UTC- 10:00	Hawaii Standard Time (HST)
-09.30	UTC- 09:30	Marquesas Time (MART)
-09.00	UTC- 09:00	Alaska Standard Time (AKST)
-08.00	UTC- 08:00	Pacific Standard Time (PST)
-07.00	UTC- 07:00	Mountain Standard Time (North America) (MST)
-06.00	UTC- 06:00	Central Standard Time (North and Central Americas) (CST)
-05.00	UTC- 05:00	Eastern Standard Time (North and Central Americas, Caribbean) (EST)
-04.30	UTC- 04:30	Venezuelan Standard Time (VET)
-04.00	UTC- 04:00	Atlantic Standard Time (North America, Caribbean) (AST)
-03.30	UTC- 03:30	Newfoundland Standard Time (NST)
-03.00	UTC- 03:00	West Greenland Time (WGT)
-02.00	UTC- 02:00	Fernando de Noronha Time (FNT)
-01.00	UTC- 01:00	Central African Time (CAT)
00.00	UTC	Universal Coordinated Time (UTC) / Greenwich Mean Time (GMT)
+01:00	UTC + 01:00	European Central Time (ECT)
+02:00	UTC + 02:00	Eastern European Time (EET), (Arabic) Egypt Standard Time (ART)

Taxonomy of Living Things

Domain (Eukarya)
Kingdom (Animalia)
Phylum (Chordata)
Class (Mammalia)
Order (Primates)
Family (Hominidae)
Genus (homo)
Species (sapiens)

human (homo sapien)

YOURDICTIONARY

Interoperable

1. Use OPEN FILE FORMATS as much as possible

= Formats that can be used and implemented by anyone

2. Use STANDARDS common in your discipline
For example
- **Minimum information:** metadata that as a minimum should be included to describe a data set
- **Vocabularies:** predefined and authorized terms for metadata
- **Taxonomies:** organisation of content into hierarchical relationships

Interoperable

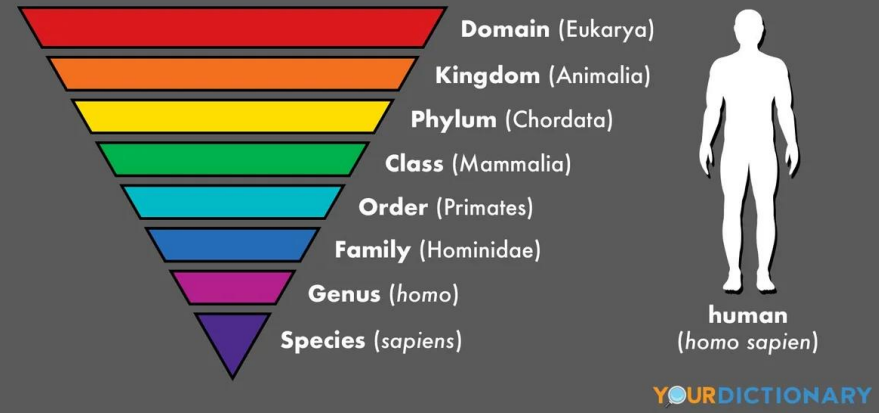**1. Use OPEN FILE FORMATS as much as possible**

= Formats that can be used and implemented by anyone

2. Use STANDARDS common in your discipline
For example
- **Minimum information:** metadata that as a minimum should be included to describe a data set
- **Vocabularies:** predefined and authorized terms for metadata
- **Taxonomies:** organisation of content into hierarchical relationships
- **Ontologies:** definitions and relations between metadata elements

Foundation URI : http://id.who.int/icd/entity/323148092

7A00 Chronic insomnia

All ancestors up to top
- 07 Sleep-wake disorders
  - Insomnia disorders
    - 7A00 Chronic insomnia

Description

Chronic insomnia is a frequent and persistent difficulty initiating or maintaining sleep that occurs despite adequate opportunity and circumstances results in general sleep dissatisfaction and some form of daytime impairment. Daytime symptoms typically include fatigue, depressed mood or irri malaise, and cognitive impairment. The sleep disturbance and associated daytime symptoms occur at least several times per week for at least 3 m individuals with chronic insomnia may show a more episodic course, with recurrent episodes of sleep/wake difficulties lasting several weeks at a ti Individuals who report sleep related symptoms in the absence of daytime impairment are not regarded as having an insomnia disorder. If the inso sleep-wake disorder, a mental disorder, another medical condition, or a substance or medication, chronic insomnia should only be diagnosed if th independent focus of clinical attention.
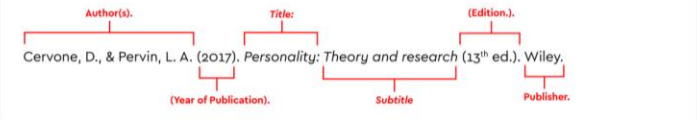
Postcoordination ❓

Add detail to **Chronic insomnia**

Has severity (use additional code, if desired .)

XS5W	Mild
XS0T	Moderate
XS25	Severe

Interoperability – a simple example on standardizing ant research

- Documenting my data in a **language everyone understands**: English

- Referring to the ant collection location using **internationally recognised coordinates**: 9.1165° N, 79.6965° W

- Referring to the ant species according to Linneaus' **taxonomy and binomial nomenclature** using genus & species name: *Atta colombica*

- Referring to the **Barcode Index Number (BIN)** for *Atta colombica*: BOLD:AAZ9697

- Describing my experiment using the MIFlowCyt 1.0, a **standard for outlining the minimum information** required to report the experimental details of flow cytometry experiments

- Saving my spreadsheets in **formats that others can open**: CSV

- Adhering to the APA citation style for the references in my manuscripts:

Cervone, D., & Pervin, L. A. (2017). *Personality: Theory and research* (13th ed.). Wiley.

Author(s). Title: (Edition).
(Year of Publication). Subtitle Publisher.

Interoperable – a question and an answer

> *What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines?*

"Data will be presented in **standard file formats** such as CSV, JSON and FITS, and files will be **logically structured** (e.g. multi-extension FITS files) to make it easier for others to open and use the files. When addressing astronomical concepts and their inter-relationships, we will as much as possible use the definitions and links for these concepts as outlined in the community supported **Unified Astronomy Thesaurus (UAT)**."

Fictional answer based on: https://librarycarpentry.org/Top-10-FAIR/2019/09/06/astronomy/

Reusable

1. Create sufficient DOCUMENTATION that explains the data. Make this documentation available along with the data

Data provenance, research documentation (interview guide, codebook, workplan, protocol, variable list, ReadMe), open source code, etc.

Description of the Climate Change and Political Parties Dataset

Units	64 parties at national elections
Number of Countries	6
Number of Elections	32
Time Period Covered	1993-2015

Basic information

Country	Country name
RL	Right- or left-of-centre
Party.name	Party name
Year	Election year
Crisis	Before or after mid-2008

Measures of climate policy preferences and their components

Wordcount	Number of words in the document
Quasi.sentences	Number of quasi-sentences (QS) in the document
Pro.climate.QS	Number of pro-climate QS
Pro.climate.pct	Pro-climate QS as % of all QS
Anti.climate.QS	Number of anti-climate QS
Anti.climate.pct	Anti-climate QS as % of all QS
Core.pro.N	Number of Core pro-climate QS
Core.pro.pct	As % of all QS
Core.anti.N	Number of Core anti-climate QS
Core.anti.pct	As % of all QS
Position	Pro.climate.pct – Anti.climate.pct
Core.position	Core.pro.pct – Core.anti.pct

'Document attributes' related to climate change
See Appendix D for full questionnaire

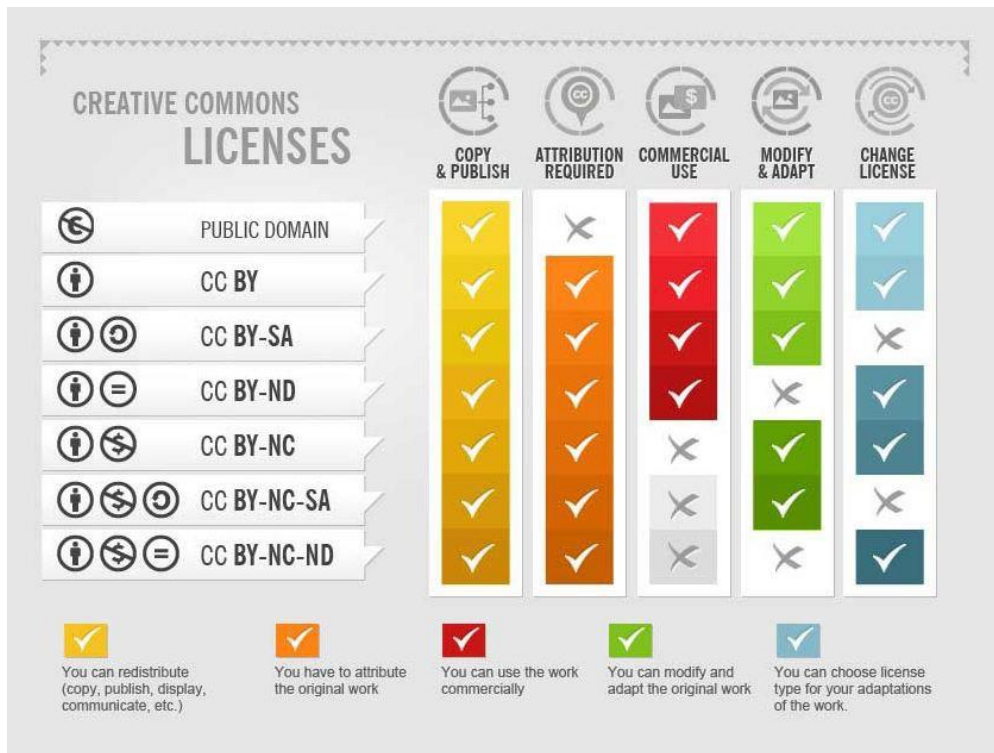|	
Acknowledges	Acknowledges climate change as a problem
Goals	Commits to general national climate goals
Prominence.pages	After what proportion of the document does the section dealing with climate change appear?
Front.matter	Is climate change mentioned in the 'front matter' of the document?
Climate.mentions	Number of mentions of climate change and cognate terms
Mention.pct	(Number of mentions of climate change / wordcount)*100

https://reshare.ukdataservice.ac.uk/852669/

Reusable

1. Create sufficient DOCUMENTATION that explains the data. Make this documentation available along with the data

Data provenance, research documentation (interview guide, codebook, workplan, protocol, variable list, ReadMe), open source code, etc.

2. Use REUSE LICENSES to communicate how others can reuse your data files

Overview of open licenses
https://opendefinition.org/licenses/

CREATIVE COMMONS LICENSES

		COPY & PUBLISH	ATTRIBUTION REQUIRED	COMMERCIAL USE	MODIFY & ADAPT	CHANGE LICENSE
	PUBLIC DOMAIN	✓	✗	✓	✓	✓
	CC BY	✓	✓	✓	✓	✓
	CC BY-SA	✓	✓	✓	✓	✗
	CC BY-ND	✓	✓	✓	✗	✓
	CC BY-NC	✓	✓	✗	✓	✓
	CC BY-NC-SA	✓	✓	✗	✓	✗
	CC BY-NC-ND	✓	✓	✗	✗	✓

You can redistribute (copy, publish, display, communicate, etc.)

You have to attribute the original work

You can use the work commercially

You can modify and adapt the original work

You can choose license type for your adaptations of the work.

Reusable – a question and an answer

> *Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?*

> *Will the data produced in the project be useable by third parties, in particular after the end of the project?*

"By default, all data produced in the project are usable by third parties without restrictions after the end of the project and will be made available in the abovementioned repositories under a Creative Commons Attribution 4.0 license (CC-BY). Any personal data collected in WP3 that cannot be anonymized upon project end may only be reused, IF study participants have provided informed consent for this AND legal agreements are in place that outline conditions for the secure transfer to, and subsequent management of, data by third parties (e.g. data disclosure agreements)."

Fictional answer

Allocation of resources

Here you show that there is a solid plan:
1) You are aware of possible costs for RDM and can cover them
2) There is alignment of expectations regarding division of responsibilities

> What will the costs be for making data or other research outputs FAIR in your project?
 "The main task to be undertaken to ensure data is FAIR, is the deposition of the final dataset with the Archaeology Data Service, estimated at 1200 euros. These archiving costs are one-off, and cover the management and preservation of the dataset for eternity."

> Who will be responsible for data management in your project?
" **The Project Coordinator** is responsible for overall data management of PROJECT Z. The PC coordinates with Work Package Leaders to determine how the data generated by the project become available for re-use
The Work Package Leaders are responsible for coordinating the implementation of the data processing activities performed under the WPs they are leading, and are responsible for assuring the quality of the data.
The Data managers are tasked to collect, digitise, anonymise, store, destroy and/or otherwise process data for the specific purpose of the WP they have been assigned to."

Fictional answers

Data security

Here you show that:
1) You can identify which data need extra protection (data classification)
2) You understand that your data storage and security set-up should match the data classification
3) Data security entails a focus on digital (IT) measure AND physical measures AND project governance

> What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?
 "Digital data sets, as well as the associated documentation that identifies the participants in the study (ID key), are stored on a secure university drive (S drive) where all activities are logged. Only project members will have access to the digital data files, and only the PI of the project has access to the ID key. A subset of the data will be sent to German university X for analysis Y, using the Sensitive Information Facility (SIF), a secure platform for the sharing of sensitive data at the University of Copenhagen. A data processing agreement will be set up, specifying the security requirements that the German university will have to have in place in order to manage our data."

Fictional answer

Susanne P.A. den Boer

3 Take home message

Tips

- Look at example DMPs, there are many online at OpenAIRE, Zenodo, DCC, RIOjournal, LIBER
- Look at existing guidance to write a DMP
   !!! Your university may offer guidance, e.g. in DMPonline: https://dmponline.deic.dk/
- Ask others: colleagues who have already written a DMP, data management support person

Useful links:
- Browse data repositories: www.re3data.org
- Browse disciplinary metadata standards:
  https://www.dcc.ac.uk/guidance/standards/metadata
- 20-minute eLearning videos RDM, FAIR and DMPs: https://deic.dk/en/RDMELearn
- FAIR website with examples from NAT/TEK, SUND, HUM, SAMF: https://howtofair.dk/

Take home messages

- DMPs are planning tools/manuals for the project, and not reports written to your project officer

- Main focus in Horizon Europe is on data sharing by default and the FAIR principles

- Data repositories are key in addressing FAIR! They
  - Can issue persistent identifiers (F)
  - Make your metadata (e.g. keywords) searchable in search engines (F)
  - Can regulate access to your data (A)
  - Can ensure your metadata remains available, even if the data is not (A)
  - Allow you to upload documentation along with your data (R)
  - Can issue data usage licenses to communicate how your data are to be reused (R)

  > Review what your repository of choice can do and use this info in your answers

Trusted repositories support data FAIRification

Findable:

✓ F1. (meta)data are assigned a globally unique and persistent identifier;

✓ F2. data are described with rich metadata;

✓ F3. metadata clearly and explicitly include the identifier of the data it describes;

✓ F4. (meta)data are registered or indexed in a searchable resource;

Interoperable:

✓ I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

✓ I2. (meta)data use vocabularies that follow FAIR principles;

✓ I3. (meta)data include qualified references to other (meta)data;

Accessible:

✓ A1. (meta)data are retrievable by their identifier using a standardized communications protocol;

✓ A1.1 the protocol is open, free, and universally implementable;

✓ A1.2. the protocol allows for an authentication and authorization procedure, where necessary;

✓ A2. metadata are accessible, even when the data are no longer available;

Reusable:

✓ R1. (meta)data are richly described with a plurality of accurate and relevant attributes;

✓ R1.1. (meta)data are released with a clear and accessible data usage license;

✓ R1.2. (meta)data are associated with detailed provenance;

✓ R1.3. (meta)data meet domain-relevant community standards;

Slide by Luiz Bonino/ Phortos Consultants, *FAIR Data Stewardship Course*, published under CC BY-NC-SA 4.0