# EOSC-LIFE: BUILDING A DIGITAL SPACE FOR THE LIFE SCIENCES

## EOSC-LIFE WP4 TOOLBOX:
## Categorisation system for resources to be referenced in the toolbox for sharing of sensitive data

WP4 – **Policies, specifications and tools for secure management of sensitive data for research purposes**
Lead Beneficiary: **ECRIN, BBMRI**
Date of publication: **30 November 2020**

Grant agreement no. 824087
Horizon 2020
H2020-INFRAEOSC-2018-2
Type of action: RIA

Authors: **C. Ohmann (ECRIN), S. Canham (ECRIN), J.-W. Boiten (EATRIS/Lygature), M. Cano Abadia (BBMRI), G. Chassang (INSERM), M.L. Chiusano (EMBRC), R. David (ERINHA), M.Th. Mayrhofer (BBMRI), L. Pireddu (CRS4-BBMRI-IT)**

# Table of Contents





This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

# Categorisation system for resources to be referenced in the toolbox for sharing of sensitive data

Authors: C. Ohmann (ECRIN), S. Canham (ECRIN), J.-W. Boiten (EATRIS/Lygature), M. Cano Abadia (BBMRI), G. Chassang (INSERM), M.L. Chiusano (EMBRC), R. David (ERINHA), M.Th. Mayrhofer (BBMRI), L. Pireddu (CRS4-BBMRI-IT)

Status: Version 1, final
Date: 30 November 2020

## ABSTRACT

The Horizon 2020 project EOSC-Life brings together the 13 Life Science 'ESFRI' research infrastructures to create an open, digital and collaborative space for biological and medical research. Sharing sensitive data is a specific challenge within EOSC-Life. For that reason, a toolbox is developed, providing information to researchers who wish to share sensitive data or to use sensitive data. The sensitivity of the data may arise from its personal nature but can also be caused by intellectual property considerations, biohazard concerns, or the Nagoya protocol. The toolbox will not create new content, instead, it will allow researchers to find existing resources that are relevant for sharing sensitive data across all participating research infrastructures (F in FAIR). The toolbox will provide links to recommendations, procedures, and best practices, as well as to software (tools) to support data sharing and reuse. It will be based upon a tagging (categorisation) system, allowing consistent labelling and categorisation of resources, in terms most relevant to data sharing tasks and activities and referring to 8 dimensions (resource type, research field, research design, data type, stage in data sharing life cycle, geographical scope, specific topics or keywords and targeted group). In this document the categorisation system is described and applied to examples. It is planned to evaluate the categorisation system within EOSC-Life in a larger pilot study.

**Key words:**

Sensitive data, tags, categorisation, life sciences, toolbox, data sharing, EOSC-Life

# 1. BACKGROUND

The EOSC-Life Toolbox (1) aims to support identification of existing and relevant information for:

- researchers or other data providers (e.g. sponsors, institutions, private organisations, ...) wishing to make their sensitive data available and accessible for future reuse (enabling future sharing of data);
- researchers or other data consumers wishing to make use of sensitive data made available by a data provider (enabling actual sharing of data).

The focus of the toolbox is on issues pertaining to sharing sensitive data in the life sciences. The sensitivity of the data may arise from its personal nature (in particular health data), but can also be caused by intellectual property considerations, biohazard concerns, or the Nagoya protocol. The concept of sensitive data other than personal data is closely linked to Dual Use Research of Concern (DURC), as defined by the United States Government Policy: DURC is life sciences research that, based on current understanding, can be reasonably anticipated to provide knowledge, information, products, or technologies that could be directly misapplied to pose a significant threat with broad potential consequences to public health and safety, agricultural crops and other plants, animals, the environment, material, or national security (2).

It is not envisaged that new content will be created. Instead, the toolbox will allow people to find existing resources that are relevant for sharing sensitive data. It will act as a 'one stop shop' across all participating life-science infrastructures. The toolbox will provide links to recommendations, procedures, and best practices, as well as to software (tools) to support data sharing and reuse operations. The toolbox will therefore help scientists to find previously collected relevant and high-quality content available throughout our collective infrastructure landscape.

In the context of the RDA COVID-19 Working Group activities, a tagging system was developed to characterise documents, allowing better support for searching and filtering (3). This system covers several dimensions, such as actors/stakeholders, research domain (Frascati), themes, stage in research cycle, type of resource, type of file/data, jurisdiction, etc. This approach was used as input to the RDA categorisation system.

We propose a similar system for the EOSC-Life WP4 toolbox. Such a tagging or categorisation system would be used in conjunction with both basic metadata (title, authors, year of publication, resource type etc.), and a brief summary of each resource. The tags are designed to support consistent labelling and categorisation for the stored resources, in terms most relevant to data sharing tasks and activities, so that they are available to users (e.g. as on screen filters) when searching in the information system. They are designed to be used in conjunction with traditional text-based searching methods, e.g. of the resources' titles.

# 2. CATEGORISATION SYSTEM

The proposed system has 8 categories:



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- Resource type
- Research field
- Research design
- Data type
- Stage in data sharing life cycle
- Geographical scope
- Specific topics
- Targeted group

The intention is to provide a pre-specified list of available tags within each category, each acting as a controlled vocabulary for the 'key words' assigned to resources. A resource may have more than one tag applied in each category, but the tag(s) must be selected from that pre-set list. These lists could be expanded, following discussion within the group, but not on an *ad hoc* basis by individuals carrying out tagging.

Each of these dimensions is considered in more detail below.

## 2.1 RESOURCE TYPE

This category indicates in broad terms the main purpose, aims or area covered by the resource rather than the details of its content. The categories have been selected to cover the main types of resources to be included in the toolbox and are purposely broad, to try and ensure all resources can be categorised in this dimension relatively easily. At least one resource type should be identified for a resource.

- Legislation and regulations *(including case-law)*
- Position papers, policies, and principles
- Background and explanatory material *(including comments)*
- Best practice *(use cases with very well organised processes)*
- Guidances and recommendations
- Systems, tools and services
- Repositories or other infrastructures
- Other resource type *(including protocols, modelling)*
- Not applicable

## 2.2 RESEARCH FIELD

Usually, a resource is linked to one research field. There may be specific cases, where a mixture of research fields can be allocated (e.g. corals = animals with zoxantelles, lichens mix mushrooms & algaes). In that case multiple categories should be ticked (e.g. botany and zoology) or if not adequate, the category "other" should be chosen.

- Health research *(medical / health research involving human subjects, including medical technology research, translational research, nutrition)*

- Pre-clinical research *(molecular / physiological / pharmacological research not involving human subjects but ultimately intended to impact medicine or health, including translational research).*
- General life sciences *(Cell/molecular biology, including microbiology, bacteriology, virology, archeological research)*
- Plant sciences, botany *(research on plant species, classification, phenotyping, physiology, genetics/molecular biology, behaviour etc., excluding ecology)*
- Zoology *(including ornithology, entomology, behavioural biology)*
- Marine and water (or aquatic) biology
- Microbiology
- Ecology *(research into ecosystems, interactions between species, populations, interactions of life with environmental factors, and/or impact of humans including agriculture-and climate change)*
- Other *(anything that does not fit into the categories above)*
- not specified/not clear

We started the discussion about research fields to be covered under this category with a look at the OECD Frascati manual 2015 (4). This manual is generic, spanning all major research areas, including "Medical and health sciences". It lists the following research: basic medicine, clinical medicine, health sciences, medical biotechnology and other medical science. In order to better take the transition phases from basic via translational to clinical research into consideration (see 5 for definition) and to also include the research fields of EOSC-Life partners, which are not directly linked to clinical or pre-clinical research, we looked at the research infrastructures participating in EOSC-Life. The infrastructures were grouped in terms of the type of research they support (or might support). Some of the infrastructures are involved in more than one field.

- BBMRI, EATRIS, ECRIN, ELIXIR, ERINHA and EuroBioImaging are involved in health research involving human subjects. This includes medical technology research and translational research at the 'clinical end'.

- Pre-clinical medical research, covering molecular / physiological / pharmacological research not involving human subjects but ultimately intended to impact medicine or health, including translational research at the 'bench end', is performed by, EATRIS, ELIXIR, ERINHA (animal experimentation), EuroBioImaging, EU Openscreen and Infrafrontier.

- General life sciences research, covering research into life systems in general with no direct application to medicine or health, is performed by ELIXIR, ERINHA, EuroBioImaging, INSTRUCT, ISBE.

- Further research fields that are covered in EOSC-Life are plant sciences/botany (Emphasis, EMBRC), zoology/marine biology (EMBRC, ERINHA), microbiology (MIRRI, ERINHA, EMBRC) and
  ecology (Emphasis, EMBRC, MIRRI).



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

## 2.3 RESEARCH DESIGN

There are many research designs in use, often adapted to the research field under consideration. An example are the many different types of observational studies used in social and medical sciences. In order to be able to involve as many research infrastructures as possible, a decision was taken to use a few very broad categories of research design. For a given resource, at least one research design should be selected; there may be resources where a mix of research designs has been applied.

- Experimental or interventional research *(in any domain, and including clinical trials, biological / biochemical or ecological experiments etc.)*
- Observational research *(including survey and population data, observation of behaviour, treatment cohort data etc.)*
- Secondary *research (including meta-analyses and systematic reviews using existing primary data, research on research, etc.)*
- Modelling research *(to be discussed after the pilot and if validated, to be better described)*
- Other research designs
- Not specified / not *clear (it is not clear if the resource applies to any particular type of research or to all of them)*
- Not applicable

We started with different typologies for research and study types (6, 7, 8, 9) and had a look at the major categories, which could be applied to as many EOSC-Life infrastructures as possible. Of importance is the differentiation between observational and experimental studies, which can be applied widely (10). Other major research approaches to be covered could be modelling research and secondary use of data from primary research for re-analysis, meta-analysis or further analysis. More detailed listings of study types (e.g. case-control, cross-sectional, cohort) were perceived as often being restricted to particular research fields, which could make the categorisation of a resource in this category difficult.

## 2.4 DATA TYPE

This category indicates in broad terms the data sources being considered within EOSC-Life and to be referenced in the toolbox. The toolbox is aimed at resources for data sharing in general – certainly with the focus on sensitive data but not exclusively so. If the resources are limited to only those explicitly dealing with sensitive material useful resources may not be included. So, the classification of data types is primarily structured according to whether the data are related to living human beings (A) or not (B). For the B an extra category has been introduced (B6), which can be applied in addition to a selection from B1 to B5. At least one type of data source should be selected, unless no data source is specified or it is not clear. Multiple data sources are possible.

- A) Data with data from / about identifiable living human beings (even if de-identified, or

claimed to be anonymised)
A1. Real world or routine health data (*including lifestyle and wearables data*)
A2. Clinical research data (*from both interventional and observational studies*)
A3. Biobank and registry data
A4. Human population level health or socio-economic data
A5. Data including images of humans
A6. Genetics and molecular biology data

- B) Data not from / about identifiable living human beings
  B1. Omics data generated by basic research
  B2. Pre-clinical research data (*all types*)
  B3. Organism or species specific data (all types)
  B4. Ecological / environmental data
  B5. Other biological data
  B6. Non-personal sensitive data *(additional tag under B)*

- C) Other data
  C1. Other type of data, not listed above
  C2. Not specified or not clear
  C3. Non applicable

A large segment of the data to be handled and processed by life sciences research infrastructures is sensitive data (see 2, 11 for definition). The concept of "data type" in the categorisation system is intended to cover sensitive data from all research fields involved, not only from health research.

## 2.5 STAGE IN DATA SHARING LIFE CYCLE

This category defines the stage in the data sharing life cycle, under consideration in the resource to be referenced. The data sharing life cycle covers the full workflow from planning of data sharing via actual sharing via use of shared data till the impact of data sharing (e.g. research output from sharing). A resource can cover one specific step in the data sharing life cycle, several steps or even the full data sharing life cycle:

- Preparation for data sharing (*e.g. processes to prepare FAIR implementation*)
- Planning for data sharing (specific study) (*including data sharing plan*)
- Data preparation at the end of the study (*after data collected*)
- Transfer of data to a repository *(when data are ready for sharing)*
- Managing data access (*for controllers/repositories*)
- Access to data for re-use by researchers
- Publication of results from re-using data *(e.g. made available by repositories)*
- Monitoring data sharing/access
- Discovering datasets for secondary use



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- Other
- Any (*covering the full data sharing life cycle*)
- Not applicable *(cannot be linked to data sharing life cycle)*

To allow compatibility with general systems of data/research life cycle, the answer categories are mappable to the vocabulary of the UK Data Archive (12). For the definition of this category it may also be useful to have a look at the EDAM Topics Subset on data management (13) and the classification of processes involved in sharing individual participant data from clinical trials (14). Preparation for data sharing covers identifying the processes required to prepare FAIR implementation and deployment of procedures and training adopted to each practice and level of understanding (15). The classification of stages in the data sharing life cycle aims to cover both health and non-health life science data sharing activities. If the classification does not fit adequately for a specific resource, the category "other" should be used.

## 2.6 GEOGRAPHICAL SCOPE

The geographic area being considered or included by the resource, if any specific such scope is present, explicitly or implicitly. The names used (for countries etc.) would need to be taken from pre-prepared lists to ensure consistency. If the resource refers to several countries, the involved countries should be listed.

- Local *(limited to outstanding references)*
- Global
- Continental *(named*) – e.g. 'Europe', 'North America'
- Region in the world *(named)* – e.g. 'Middle East', 'Caribbean', 'Sub-Saharan Africa'
- National *(named)*
- Sub-National *(region in a country, named)* – e.g. 'California', 'Ile de France'
- Not applicable

"Geographic regions and countries" are taken from the United Nations publication "Standard Country or Area Codes for Statistical Use" originally published as Series M, No. 49 and now commonly referred to as the M49 standard (16).

## 2.7 SPECIFIC TOPICS

These are specific named topics or aspects of data sharing – often corresponding to points of particular interest, concern or confusion – so likely to occur frequently in users' searches. The category list is being constructed pragmatically – as a response to both the expressed interest in different topics and to the numbers of resources available that deal with those topics. It will therefore be partly content-driven.

The initial list was derived from clinical research and may need to be extended to cover topics in other research areas. We would suggest, however, that there should not be more than about 25

topics listed in total. It is expected that usually several “specific topics” are allocated to a resource.

- Legal aspects
- GDPR
- Data transfer agreement
- Data use agreement
- Data storage agreement
- Broad consent
- Informed consent
- Alternatives to consent
- Ethics of data sharing
- Planning for data re-use
- Data governance
- Data access committee
- Metadata for data sharing
- Attribution and credit for data sharing
- Anonymisation
- Pseudonymisation
- De-identification
- Repository quality *(assessing quality of repositories)*
- Managing data access
- Technical and organisation control measures (*data security*)
- Other topics *(to be named and included in answer categories after cleaning)*

Basic definitions of terms used in research data management and applicable to the criterion “specific topics” should be used from the RDY-CASRAI research data management glossary (17) and from the CODATA research management glossary (18).

## 2.8 TARGETED GROUP

This category allows the identification of a user or user group that is the specific focus of the resource. It should only be used if a resource is specifically written for or about a particular group and their interaction with one or more data sharing tasks. If applicable, it is suggested to both document the developer of the resource (e.g. resource funder) and the receiver of the resource (e.g. data provider).

- Resource funder
- Policy maker
- Coordination forum *(e.g. RDA)*
- Standardisation body
- Research communities
- Data service providers



This project has received funding from the *European Union’s Horizon 2020 research and innovation programme* under grant agreement No 824087.

- Data stewards
- Data provider
- Data consumer
- Other group
- Cannot be specified clearly
- Not applicable

There are many sources referring to the category "targeted group". We applied the EU Commission classification used in FAIR (DMP) (19).

## 3. EXAMPLES

The categorisation system has been applied to 5 diverse examples, as an initial test of its applicability:

| **Source** | **1 Resource type** | **2 Research field** | **3 Research design** | **4 Data type** | **5 Stage in DS life cycle** | **6 Geo-graphical scope** | **7 Specific topics** | **8 Targeted group** |
|---|---|---|---|---|---|---|---|---|
| Hallinan, Broad consent under the GDPR (20) | Publication | Health research, Pre-clinical research | Experimental/interventional, observational research | Biobank/ registry data, genetic and molecular biology data | Planning for DS | EU | Broad consent, GDPR | Researcher |
| Article 29 Data Protection WP: Opinion (21) | Guidance/ recommendations | not specified/ not clear | not specified/ not clear | Not specified/ not clear | Data preparation at the end of study, transfer of data to repository, managing data access, access to data for re-use | EU | De-identification, anonymisation, pseudonymisation | Researcher |
| Lin et al. TRUST principles, 2020 (22) | Position papers/policies/principles | not specified/ not clear | Not specified/ not clear | Not specified/ not clear | Transfer of data to repository, managing data access | Global | Repository quality | Repository manager, researcher, funder |
| Anamnesia tool (23) | Systems/ tool/ services | Sensitive data | Not specified/ not clear | Real world or routine health | Data preparation at the end of | Global | Anonymisation, | Researcher |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | data, clinical research data | study, transfer of data to repository, managing data access, access to data for re-use | | de-identification | |
| Gahl et al., Swiss CTU Network, 2020 (24) | Guidance/ /recommendations | Health research | Experimental/Interventional observational research | Clinical research data | Planning for DS, data preparation at the end of study, transfer of data to repository, managing data access, access to data for re-use | Switzerland | Legal aspects, informed consent, planning for data re-use, anonymisation. metadata for data sharing, repository quality, data governance, managing data access | Sponsor investigator statistician, data manager |

# 4. NEXT STEPS

## 4.1 DEVELOPMENT OF A DEMONSTRATOR

In a first step a demonstrator will be developed. This demonstrator will contain a pre-specified number of resources (around 100), independently assessed by experts from the individual life-science infrastructures. It will provide a first version of the portal, allowing findability of the assessed resources to support sharing of sensitive data. This demonstrator will be tested for feasibility, user-friendliness and potential benefit. The information collected will be used for an update of the portal. In parallel, a maintenance and sustainability plan will be developed within EOSC-Life WP4, taking into consideration
necessary resources and regular involvement of experts.

Technically, the categorisation system could be realised via an electronic form, where the user



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

ticks multiple boxes to filter according to the available categories, and the references to the resources are displayed. In order to achieve this, a portal with a filtering function has to be developed. This approach could be combined with a search engine, allowing free text search in the title, abstract or keywords (similar to the metadata repository (MDR) developed by ECRIN (25)). The approach may be compatible with the BBMRI ELSI knowledge base, which also uses a filtering system, however, with different
categories (26). Whether it can be made compatible has to be checked (27).

Whether a navigation approach via a decision-tree would be useful could be explored in parallel. Here the categories are transferred into questions and are sequentially implemented in the DS-wizard (28), using
the available categories to drive the dialogue. For example:

*What type of resource are you interested in?*

- *For which stage in the data sharing life cycle you are looking for?*
  - *Are you interested in information for specific data sources?*
    - *Are you looking for resources from a specific geographical area?*
      - *Are you interested in specific topics?*
        - *Are you interested in a specific user group?*

The decision-tree query approach could be of great interest provided it could be implemented adequately based on the categories defined. In the decision-tree approach (and perhaps also in the categorisation) two target users could be separated: the one who wants to share data and the one who wants to reuse, or is re-using, other people's data. It was proposed to extend the work on this before finalising the categorisation system, to try to ensure that the categorisation scheme supports the decision tree(s) needed. For lack of time it is suggested to perform the pilot with assessment of a limited number of resources and to do this exploration in parallel.

The accuracy of the toolbox will be a key aspect of its implementation. The categorisation system has to be updated periodically if needed. In addition, experts need to be involved regularly to check whether adaptations in the categorisation system need re-assessment of resources already included in the toolbox and whether the assessment of a resource has to be changed after a while. Documentation might be outdated fast without the proper intervention of experts that make sure that the content of the toolbox is accurate and up to date. In order to implement the tool, a first intervention of experts is fundamental to assess the referenced documentation, and this assessment needs to be done again regularly to ensure the correct implementation of the tool.

## 4.2 PILOT STUDY

As a pilot, the categorisation system should be applied to a limited corpus of resources (e.g. 25 respectively 10 resources per participating infrastructure). Two reviewers should assess a resource independently of each other. In case of disagreement consensus should be sought. To

support this process a bibliographic system will be used. It is proposed to use Zotero. The protocol for the pilot study will be registered in OSF before the start of the study (https://osf.io/).

After the pilot phase, the categorisation system will be reviewed and updated. If necessary, the categorisation of the resources for the pilot study will be updated. The resulting database of tagged resources will form the basis for the demonstrator of the toolbox.

## 4.3 SUSTAINABILITY

The issues about the sustainability of the toolbox, including the underlying categorisation system should be addressed since the beginning of the design phase. The experience in BBMRI is that it is hard to realise in practice and that resources should be allocated for this task. Somebody has to be responsible for potential changes in the tagging system when a new document is introduced, or an old document has become obsolete and has to be removed. Experts need to be consulted for this task, to make sure that the content of the tool is up to date. Adequate resources and reliable agreements are necessary to ensure the involvement of experts (collaboration or consultancy agreements).

For the content sustainability someone needs to be ultimately responsible for curating the repository. Also, curation of larger applicability areas could be split among different research infrastructures (e.g. health research, microbiology, zoology, etc.), each contributing to the part most relevant to them. It should also be assessed whether the categories defined are used and based on what definitions.

In addition, we should strive to find ways to allow the community to participate in the process. Once the toolbox becomes useful to the community, one could hope that users would be inclined to contribute (e.g., in simple ways, such as reporting problems, or in more advanced ways, such as proposing new content with suggested categorisation). We should try to establish a simple light process through which this can be done; nonetheless, even a simple process will require experts to assess the contributions periodically. For this, it might also be possible to borrow from what's done for the maintenance of open source software (e.g., open repositories, pull requests, etc.).

Work on the maintenance and sustainability of the toolbox has been initiated and discussion among the main involved research infrastructures (e.g. BBMRI, EATRIS, ECRIN, ELIXIR) has been started.



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

# 5. REFERENCES


1. Boiten JW (EATRIS), Ohmann, C (ECRIN), Ludwig R (EATRIS) Pla AS (EATRIS), Fratelli M (EATRIS), Matei M (ECRIN), Panagiotopoulou M (ECRIN), Mayrhofer M (BBMRI), Holub P (BBMRI), Schlünder I (BBMRI), Adeniran A (BBMRI), Abadia MC (BBMRI), Chassang G (BBMRI), Merchant A (ELIXIR), David R (ERINHA), Chiusano ML (EMBRC), Tsamis G (EMBRC), Gribbon P (WP1), Pireddu L (WP2), David R (WP6), Wagener (WP7), Ludwig R (WP9): Toolbox for data sharing of sensitive data - a concept. https://docs.google.com/document/d/1ek-zluZjg7EFn-DJ9AmQvTeB2OvTGuWMmYTNs1GWRAU/edit#

2. RDA COVID-19 Working Group. recommendations and guidelines. *Research Data Alliance, 2020*. DOI: https://doi.org/10.15497/rda00046. https://www.rd-alliance.org/group/rda-covid19-rda-covid19-omics-rda-covid19-epidemiology-rda-covid19-clinical-rda-covid19-0

3. OECD (2015), Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264239012-en https://www.oecd.org/publications/frascati-manual-2015-9789264239012-en.htm)

4. Institute of Translational Health Sciences: T-Phases of Translational Health. Research. https://www.iths.org/investigators/definitions/translational-research/

5. National Centre for Research Methods Report Review of the Typology of Research Methods within the Social Sciences, January 2015. http://eprints.ncrm.ac.uk/3721/1/research_methods_typology_2015.pdf

6. Thompson CB, Panacek EA: Research Study Designs: Non-experimental, 2007. https://www.airmedicaljournal.com/article/S1067-991X(06)00309-9/pdf

7. EDDA study designs taxonomy, 2016. https://bioportal.bioontology.org/ontologies/EDDA

8. Clinical Trials Ontology (CTO), 2014. https://bioportal.bioontology.org/ontologies/CTO

9. Cath: What are interventional studies? https://www.drcath.net/toolkit/intervention-studies#:~:text=What%20are%20intervention%20studies%3F,of%20a%20health%20service%20delivery

10. Canham S, Ohmann C, Matei M, Battaglia S: White Paper 4: Ethics Supporting Document to D3.3 Draft Policy Recommendations. https://eoscpilot.eu/sites/default/files/eoscpilot_d3.3_whitepaper_4_ethics.pdf

11. National Institutes of Health. Office of Science Policy: Dual Use of Research Concern (DURC) https://osp.od.nih.gov/biotechnology/dual-use-research-of-concern/

12. UK Data Archive: Data life cycle & data management planning. https://www.ukdataservice.ac.uk/media/187718/dmplanningdm24apr2013.pdf

13. EDAM Ontology. http://edamontology.org/page

14. Ohmann C, Canham S, Banzi R *et al.* Classification of processes involved in sharing individual participant data from clinical trials [version 2; peer review: 3 approved]. *F1000Research* 2018, 7:138 https://doi.org/10.12688/f1000research.13789.2

15. David R, Mabile L, Specht A, et al.; FAIRness Literacy: The Achilles heel of applying FAIR principles. CODATA. https://datascience.codata.org/articles/10.5334/dsj-2020-032/

16. United Nations: Standard country or area codes for statistical use (M49). https://unstats.un.org/unsd/methodology/m49/

17. CASRAI: Research Data Management Glossary. https://casrai.org/rdm-glossary/

18. CODATA: Research data management glossary https://codata.org/rdm-glossary/

19. David R, Mabile L, Specht A, Stryeck S, Thomsen M, Yahia M, SHAring Reward & Credit (SHARC) Interest Group. (2020). Implementing FAIR in data sharing: who are the actors and what are their responsibilities? (Version V1.0). Presented at the RDA Plenary 15 (RDA P15), Melbourne: Zenodo. http://doi.org/10.5281/zenodo.3743946

20. Hallinan, D. Broad consent under the GDPR: an optimistic perspective on a bright future. *Life Sci Soc Policy* 16, 1 (2020). https://doi.org/10.1186/s40504-019-0096-3



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

21. Article 29 Data Protection Working Party. Opinion 05/2014 on Anonymisation Techniques Adopted on 10 April 2014

22. Lin, D., Crabtree, J., Dillo, I. *et al.* The TRUST Principles for digital repositories. *Sci Data* 7, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7

23. Anamnesia tool. https://amnesia.openaire.eu/

24. Gahl, B.; Haynes, A.; Sluka, C.; Dupuis-Lozeron, E.; Jörger, F.; Schur, R.; Christen, A.; Trelle,
S. Sharing of Data from Clinical Research Projects – Guidance from the Swiss CTU Network.
*Preprints* **2020**, 2020060344.
https://www.preprints.org/manuscript/202006.0344/v1

25. ECRIN metadata repository. http://crmdr.org/

26. Mayrhofer MT, Schlünder I. Mind the Gap: From Tool to Knowledge Base. Biopreserv Biobank. 2018 Dec;16(6):458-462. doi: 10.1089/bio.2018.0018. Epub 2018 Aug 13. PMID:
30102551; PMCID: PMC6308279

27. BBMRI ELSI knowledge base. https://www.bbmri-eric.eu/elsi/knowledge-base/

28. Data sharing wizard https://ds-wizard.org/resources.html