



HAL
open science

EOSC-Life D4.3 - Guidance and policy on standards and tools to facilitate sharing and reuse of multimodal data (including imaging), cohort integration, and biosamples

Jan-Willem Boiten, Christian Ohmann, Ayodeji Adeniran, Steve Canham, Monica Cano-Abadia, Gauthier Chassang, Maria-Luisa Chiusano, Romain David, Maddalena Fratelli, Phil Gribbon, et al.

► **To cite this version:**

Jan-Willem Boiten, Christian Ohmann, Ayodeji Adeniran, Steve Canham, Monica Cano-Abadia, et al.. EOSC-Life D4.3 - Guidance and policy on standards and tools to facilitate sharing and reuse of multimodal data (including imaging), cohort integration, and biosamples. 4.3, ECRIN; BBMRI; IRFMN; EU-OPENSOURCE; ELIXIR; VHIR; Charité; UNITO; Lygature; INSERM; EMBRC; ER-INHA; EATRIS. 2021. hal-04161974

HAL Id: hal-04161974

<https://hal.science/hal-04161974>

Submitted on 13 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



EOSC-Life: Building a digital space for the life sciences

D4.3 – Guidance and policy on standards and tools to facilitate sharing and reuse of multimodal data (including imaging), cohort integration, and biosamples

WP4 – Policies, specifications and tools for the management of data for biological and medical research

Lead Beneficiary: ECRIN-ERIC and BBMRI-ERIC

WP leader: Jacques Demotes and Michaela T. Mayrhofer

Contributing partner(s): Lygature, INSERM, EMBRC, ERINHA, EATRIS, ECRIN, IRFMN, EU-OPENSOURCE, ELIXIR, VHIR, Charité, UNITO

Authors of this deliverable: **Jan-Willem Boiten, Christian Ohmann**

Contributors: **Ayodeji Adeniran, Steve Canham, Monica Cano Abadia, Gauthier Chassang, Maria Luisa Chiusano, Romain David, Maddalena Fratelli, Phil Gribbon, Petr Holub, Rebecca Ludwig, Michaela Th. Mayrhofer, Mihaela Matei, Arshiya Merchant, Maria Panagiotopoulou, Luca Pireddu, Audrey Richard, Alex Sanchez Pla, Irene Schlünder, George Tsamis, Harald Wagener, Dario Longo, Sergei Gorianin**

Contractual delivery date: **28 February 2021**

Actual delivery date: **9 March 2021**

H2020-INFRA-EOSC-2018-2

Grant agreement no. 824087

Horizon 2020

Type of action: RIA

Table of Contents

Executive Summary.....	3
Project Objectives	3
Detailed Report on the Deliverable.....	4
1. Description of Work	4
2. Next Steps.....	10
Delivery and Schedule.....	11
Adjustments	11



Executive Summary

Sharing sensitive data is a specific challenge within EOSC-Life. This deliverable reports on the preliminary results of the design of a toolbox to provide information to researchers who wish to share and/or use sensitive data in a public cloud environment in general, and in e-infrastructures adhering to the European Open Science Cloud in particular. The sensitivity of the data may arise from its personal nature, but can also be caused by intellectual property considerations, biohazard concerns, or compliance with the Nagoya protocol. The toolbox will not be based on newly generated content, instead, it will guide 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. The toolbox design is steered by a large working group meeting on a monthly basis with participation of a broad range of research infrastructures (the authors of this report).

The current deliverable report provides a high-level summary of the results that already have been detailed further in documents published in Zenodo over the previous year, and which are referenced in this report. Finally, the document will also provide an outlook for the next steps towards the actual implementation of the toolbox and, ultimately, its sustainability beyond the EOSC-Life project's lifetime.

Project Objectives

Sharing and reusing multimodal data, images, and biosamples is essential to optimize the use of research data and the corresponding effort and funding. Some institutions and biomedical research infrastructures have already developed data sharing instruments and support services (repositories, data warehouses etc.) for that purpose. However, these activities are still fragmented, and often institution or country specific. Moreover, policies for data sharing and reuse are still available to a limited extent, which hampers the reuse of sensitive data at a larger scale across national borders. Task 4.2 in EOSC-Life ("Policy and guidance on sharing and reusing biomedical research data, images, and samples") is designed to address this situation, guided by specific use cases from within EOSC-Life (WP1, WP3) as well as from outside the project. The objective of the task is to create a comprehensive set of recommendations and policies about:

- a. sharing and reuse of observational study data and cohort integration with guidance for the reuse of cohort data collected across multiple countries;
- b. multinational observational biosample sharing and reuse;
- c. sharing and reuse of sensitive images from multinational studies;
- d. sharing other forms of sensitive data, e.g. data that need protection or could potentially give rise to a biohazard, IP-sensitive data, or data subject to the Nagoya protocol.

The resulting recommendations should inform the further use cases within EOSC-Life involving sensitive data.



Detailed Report on the Deliverable

1. Description of Work

The scope of the recommendations to be developed within this task is potentially near infinite, while there are already many activities ongoing within and outside the biomedical research infrastructures. It was therefore decided at the last EOSC-Life Annual Meeting (Brussels, March 2020) to develop a toolbox dedicated to those resources already available for handling sensitive data, instead of developing a new set of recommendations from scratch. This EOSC-Life Toolbox will aim to provide specific guidance to:

1. Researchers or other data providers (e.g., sponsors, institutions or private organisations) wishing to prepare their sensitive data for future reuse, i.e., enabling future sharing of data;
2. Researchers and data providers in the case they wish to make use of sensitive data made available by the data provider, i.e., enabling actual sharing of data.

The content of the toolbox will be derived from two main sources:

1. existing data, recommendations, procedures, best practices, and links to software (tools) to support data sharing and reuse relevant for sensitive data management in a cloud environment;
2. guidelines and other useful resources drafted in the context of EOSC-Life (e.g., WP7 guidelines on the required capabilities of cloud providers¹ and WP7: Sensitive data concerning health and trusted cloud²).

The toolbox will help scientists to navigate to previously collected high-quality content available throughout our collective infrastructure landscape. Also, the sustainability of the toolbox will be a design consideration from the start: EOSC-Life is a project with a clear end date. Any toolbox can only be useful and successful if it is maintained and updated beyond that end date.

The creation of the toolbox and its enrichment with content will be driven by use cases, which may stem from within EOSC-Life (e.g., WP1, WP2, WP3, WP6 (Nagoya protocol)), WP7, WP9 or from the Research Infrastructures' network/other EU-projects. Data sharing issues highlighted by those use cases will prioritize the content collection for the toolbox and will provide test cases for the usability of the toolbox. More details on the scope and anticipated content of the toolbox can be found in the design document for the toolbox "*Toolbox for data sharing of sensitive data - a concept description*" (1). This design document was prepared by a dedicated working group composed of members from across the life-sciences research infrastructure landscape (the authors of this report). This working group meets on a monthly basis and is planning to continue to do so over the next year. In addition, specific subgroups have been formed dedicated to specific tasks (categorization system and technical working group) as will be outlined below.

¹ https://docs.google.com/document/d/1BJG_Df6wM-d3x8fV0v3hfCBWGM3r9bGZIMFP8EWObZA/edit

² https://docs.google.com/document/d/1ZMyv0VqQcIBN_CxPGE6THOtsK9gwdQdNuGBzF5X_33U/edit



1.1. Scope of the toolbox

Control of the scope of the toolbox is essential in order to be able to create a useful product within the time and budget available. This observation has led to the following scope constraints.

- Only secondary use of data is considered; in case of primary use only those aspects related to future reuse of the data will be considered.
- The content of the toolbox should be of relevance for cloud management of sensitive data in general, and the European Open Science Cloud in particular. The priority for filling the toolbox will be set by either:
 - representative use cases, or
 - specific topics raised by other EOSC-Life work packages, or
 - specific input from the WP4 EOSC-Life toolbox team.
- The toolbox can consist of recommendations, procedures, links to tools, and examples/best practices. Preferably the best practices are derived from use cases.
- The content of the toolbox will not be developed de novo, but assembled from existing resources, preferably by providing pointers to resources rather than copying content. However, the toolbox will provide sufficient context to those pointers to allow users to understand their scope and applicability.

It would be a pitfall to aim for the perfect product from the start. The initial focus should be the “minimum lovable product”.

1.2. Input for the toolbox design

Although it is not the intention of the EOSC-Life project to develop many new guidelines, best practices, tools/services, still some of these may be needed to cover certain gaps. In those cases, EOSC-Life will organize expert workshops to generate specific content for the toolbox.

In this context, two expert workshops have already been organised. First, the workshop “Fair Enough?” (4-5 September 2019, Brussels) allowed WP4 participants to exchange on data protection, gaps and identify synergies to develop guidance and services to handle sensitive research data. Second, the workshop “Towards pseudonymised/anonymised health data: methodologies, tools and practical solutions” (22-23 January 2020, Paris) aimed at assessing technical solutions for anonymization and pseudonymization of data and their GDPR (General Data Protection Regulation) compliance.

In a series of prioritization meetings, the topic of federation vs centralization of sensitive data collections was selected as the highest-priority issue for a future workshop. The workshop will focus on the criteria required to make this decision, and on the specific challenges to deal with sensitive data in either approach, in particular in a cloud context. This workshop will be the next one to be organized in the context of EOSC-Life WP4, followed by a workshop focusing on the Nagoya protocol on Access and Benefit Sharing (ABS). Both workshops will result in specific content to be included in the EOSC-Life toolbox.



The BBMRI-ERIC ELSI Knowledge Base³ (2) may also provide an important starting point for the design, relevant content, and sustainability aspects of the toolbox. A demonstration of that Knowledge Base for this working group was organized July 27th, 2020. Please see also Screenshot 1 for an impression of the look-and-feel of that toolbox.

Another example that could inform the further development of the toolbox is the Data Stewardship Wizard created for the RDA COVID-19 initiative. A demo of this toolbox was organized for the larger EOSC-Life Toolbox working group on August 24th, 2020. Screenshot 2 provides an impression of the RDA DS Wizard.

Effective navigation is essential to the use of the toolbox and should be based upon the needs of a user related to a specific problem or use case. EOSC-Life WP9 is supporting the usability aspects of the toolbox implementation (“user journey”). A usability webinar (September 28th, 2020) was organized to inform the toolbox team about tips and tricks in this area. It was based on the EBI Training on user experience (3). See also Screenshot 3.

³ <https://www.bbmri-eric.eu/elsi/knowledge-base/>





Screenshot 1: Demonstration of the BBMRI ELSI Knowledgebase

ELSI KNOWLEDGE BASE

The ELSI Knowledge Base is our open-access resource platform containing information on ELSI-related matters relevant in biobanking. Here you will find practical guidance on specific ELSI topics.

In need of ethical legal, or societal guidance? Browse our library below!



Search Base 

Hot Topics

- ELSI & COVID-19
- Incidental findings
- Informed consent
- Paediatric Biobanking

Areas of Interest


- Data Protection
- Ethics
- Social engagement


Project Results


- ADOPT BBMRI-ERIC
- B3Africa
- BBMRI-LPC
- CINECA
- CORBEL
- RD-Connect
- SPIDIA4P


Countries


- Europe
- International

ELSI TOOLKIT 

LEGAL FRAMEWORK 

GUIDELINES, RECOMMENDATIONS & ETHICAL STANDARDS 

PUBLIC CONSULTATIONS 

REPORTS 



Screenshot 2: Demonstration of the RDA DS Wizard

The screenshot shows the 'DS Wizard' interface. On the left is a teal sidebar with navigation options: Users, Knowledge Model Editor, Knowledge Models, Questionnaires, Documents, and Templates. The main content area is divided into sections: 'Current Phase' (Before Submitting the Proposal), 'Chapters' (a list of five chapters with checkmarks), and 'I. Preparation for data sharing in general' (Chapter text). Under 'Chapters', the first chapter is highlighted. The main content area contains two numbered sections: '1 Learn about data sharing in general' (with sub-points: Web pages, Courses, Videos, Texts) and '2 Develop local SOPs and related quality documents' (with sub-point: Example SOPs and proformas).

Screenshot 3: UX Design webinar

The slide is titled 'How can we show the data?'. It features a central circular diagram with three nodes: 'Research' (top), 'Design' (bottom right, highlighted in orange), and 'Feedback' (bottom left). Arrows indicate a clockwise cycle: Research to Design, Design to Feedback, and Feedback to Research. Three photographs illustrate the process: a group of people at laptops, two people at a computer, and three people around a table with documents. The slide includes a 'UX' logo in the bottom left and the 'EMBL-EBI' logo in the bottom right.



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

1.4. User interface & categorization system

Multiple navigation approaches need to be supported. Direct search may be supported by tagging the content. This basic approach to navigation could be supported by a form/questionnaire to be filled in, which characterizes the main dimensions of data sharing. The content (e.g., guidelines, best practice, tools) could be tagged with these characteristics and thus made identifiable via search.

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 (4). This system covered several dimensions, such as actors/stakeholders, research domain, themes, stage in research cycle, type of resource, type of file/data, jurisdiction, etc., and was used as input to the RDA categorisation system. We propose a similar system for the EOSC-Life WP4 toolbox. Such a tagging or categorization 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 categorization 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. A dedicated EOSC-Life working group was instantiated to create a detailed proposal for such a tagging system, which has been published in a report (5).

In a pilot, the categorization system was applied to a limited corpus of resources. Two experts from each involved infrastructure willing to perform the assessment were recruited. The experts from the involved infrastructures selected a given number of resources by themselves, spanning the range of resource types defined in the categorization system (e.g., legislation & regulations, position papers, policies and principles, background & explanatory material, recommended practice, systems, tools & services, repositories or other resources). Infrastructures with more person-months in EOSC-Life WP4 (BBMRI, EATRIS, ECRIN) selected 25 resources, while the others (EMBRC, ERINHA, Euro-BioImaging) identified 10 resources. To support this process, the bibliographic system Zotero was used. The protocol for the pilot study has been formally published (6) in order to support the future scientific publication of the results. At time of writing this deliverable report the pilot study results are being summarized and translated into recommendations for further implementation and publication. As a preliminary result it can be concluded that simplification of the original proposed tagging system is required to arrive at consistent results between independent experts. The aim is to make the tagging system intuitive and transversal.

1.5. Technical working group

The toolbox will be implemented as a portal, where different navigation options should guide users to the right content (e.g. direct search, classical alphabetical indices, infographics, decision trees, filters). Maintenance and sustainability of that website will be an essential design consideration: it should be easy to keep the content up-to-date and the responsibility beyond the EOSC-Life project lifetime should be clear from the start. The value of the toolbox will depend to a



great extent on the reliability and currency of the content, not only in the toolbox itself, but also with respect to the content from other sources that are referenced.

A technical working group was instantiated in December 2020 with the aim of exploring technical options for the implementation of the toolbox, and to arrive at a first set of recommendations for a limited number of preferred solutions. The working group started by refining the minimum desired functionalities of the toolbox and analysing implementation options based on existing approaches relevant for its design (e.g. the BBMRI ELSI Knowledge Base). It is not intended to do any programming work at this stage but rather shortlist 2-3 alternative implementation scenarios, including cost and resources estimations, to feed the sustainability discussions.

A brief terms of reference document for this working group was written in January 2021 to define its scope and remit. The technical working group is scheduled to arrive at a conclusion in April 2021.

1.6. Sustainability & governance

The usefulness of the toolbox is not only determined by the usability of the toolbox, but also by the quality, the reliability, and the currency of its content. It is therefore crucial to have a (lightweight) governance process in place, e.g., by reviewing every toolbox topic by one or two independent experts before publishing it. One of the review criteria should cover the predicted sustainability of the content. It should be unlikely that the content will be outdated rapidly; nevertheless, some periodic review will be needed to maintain the content of the toolbox over time. It would be desirable to allow the community to participate in the process through a simple lightweight process. 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.

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. A first meeting with the directors of these infrastructures was organized December 4th, 2020.

2. Next Steps

The further maturation of the EOSC-Life toolbox for sensitive data is still in full swing. A number of short-term next steps have already been scheduled:

- The technical working group is scheduled to deliver its recommendations for proposed solutions for the toolbox implementation March/April 2021.
- The recommendations of the technical working group will also inform the sustainability discussion. The resulting implementation approach will be discussed among the most involved Research Infrastructures before actual implementation in order to ensure future sustainability and maintainability.
- The results of the pilot study with the tagging/categorization system will become available in a report March 2021, which will subsequently also be published in a formal scientific



publication. The recommendations from this report will subsequently be translated in a refined version of this tagging system that can be implemented in practice.

- A tentative program for a workshop on centralizing vs federating sensitive data has been established. It will be organized in the first half of 2021.

Delivery and Schedule

The deliverable has been delivered according to schedule.

Adjustments

The original goal of the deliverable was to create a set of recommendations on how to handle sensitive data. At the last General project meeting (Brussels, March 2020) this objective has been refined and extended into the concept of a toolbox capturing those recommendations and sustaining them. This evolution of the objectives clearly stepped up the ambition of the task, but will also offer the opportunity to sustain and maintain the results in the future.

References

- 1) Boiten, Jan Willem, Ohmann, Christian, Adeniran, Ayodeji, Canham, Steve, Cano Abadia, Monica, Chassang, Gauthier, ... Wagener, Harald. (2021, January 31). EOSC-LIFE WP4 TOOLBOX: Toolbox for sharing of sensitive data - a concept description. Zenodo. <http://doi.org/10.5281/zenodo.4483694>
- 2) Mayrhofer, M.Th., Schlünder, I.: Mind the Gap: From Tool to Knowledgebase; Biopreservation and Biobanking 16; (2018); DOI: [10.1089/bio.2018.0018](https://doi.org/10.1089/bio.2018.0018)
- 3) EBI training on user experience: <https://www.ebi.ac.uk/training-beta/online/courses/user-experience-design/what-is-user-experience-design/>
- 4) RDA COVID-19 Working Group. recommendations and guidelines. Research Data Alliance, 2020. DOI: <https://doi.org/10.15497/rda00046>
- 5) Ohmann, Christian, Canham, Steve, Boiten, Jan-Willem, Cano Abadía, Mónica, Chassang, Gautier, Chiusano, Marie Luisa, ... Pireddu, Luca. (2020, December 8). EOSC-Life WP4 Toolbox: Categorisation system for resources to be referenced in the toolbox for sharing of sensitive data (Version Version 1). Zenodo. <http://doi.org/10.5281/zenodo.4311094>
Ohmann, Christian, Canham, Steve, Boiten, Jan-Willem, Cano Abadía, Mónica, Chassang, Gautier, Chiusano, Maria Luisa, ... Pireddu, Luca. (2020, December 8). EOSC-Life WP4 Toolbox: Pilot study protocol for the evaluation of the categorisation system (Version 1). Zenodo. <http://doi.org/10.5281/zenodo.4311114>

