Pilot study on the intercalibration of a categorisation system for FAIRer digital objects related to sensitive data in the life sciences

Romain David, Christian Ohmann,, Jan-Willem Boiten,, Steve Canham,, Maria Panagiotopoulou,, Theresia Mayrhofer, Michaela, Dario Longo,, Florence Bietrix,, Luisa Chiusano, Maria, Arnaud Laroquette,, et al.

Pilot study on the intercalibration of a categorisation system for FAIRer digital objects related to sensitive data in the life sciences

R. David (ERINHA), C. Ohmann (ECRIN, pilot study coordinator), M. Cano Abadia (BBMRI), F. Bietrix (EATRIS), J.-W. Boiten (EATRIS/Lygature), S. Canham (ECRIN), M.L. Chiusano (EMBRC), W. Dastru (Euro-BioImaging), A. Laroquette (EMBRC), D. Longo (Euro-BioImaging), M.Th. Mayrhofer (BBMRI), M. Panagiotopoulou (ECRIN), A. Richard (ERINHA), P. Verde Contact: christianohmann@outlook.de

18th Research Data Alliance Virtual Plenary Meeting
3-11 November 2021

Background: Sharing sensitive data is a specific challenge within EOSC-Life. A toolbox is developed, providing information to researchers who wish to share and use sensitive data and to support the workflow of handling these kinds of data objects.

The **objective** of the pilot study (Ohmann *et al.*, 2021) was to evaluate the first version of a categorisation system for digital objects from life sciences by human experts from different research infrastructures with respect to **applicability, reliability, sustainability and consistency**. The primary objective is to identify the main improvements that are necessary to arrive at an inter-community, well understood and **commonly approved categorisation system**. From the results, an improved version of the categorisation system has been provided and will be used as a basic component for the toolbox under development.

Pilot study method: Each involved infrastructure nominated two experts, willing to perform the individual assessment of a number of typical resources around sensitive data to be included in the EOSC-Life toolbox. The experts from the involved infrastructures **selected up to 25 resources**, spanning a wide range of resource types (e.g. legislation & regulations, position papers, policies & principles, background & explanatory material, tools).

The experts selected the resource types that are relevant for their research infrastructure and then assessed these resources independently of each other using the categorisation system that we developed.

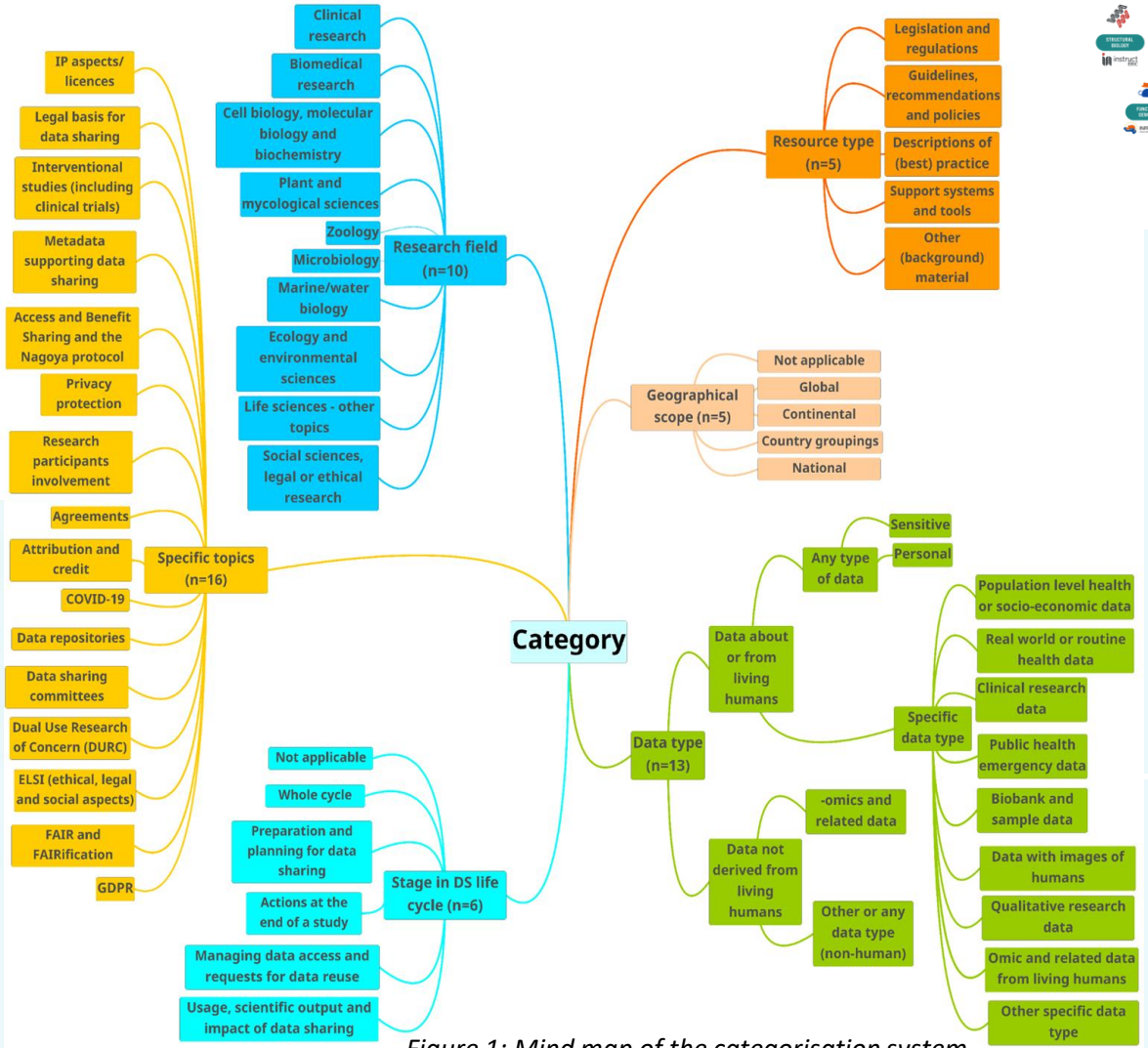

Figure 1: Mind map of the categorisation system

What is EOSC-Life? EOSC-Life is an H2020-funded project bringing together the 13 Life Science 'ESFRI' Research Infrastructures to create an open, digital and collaborative space for life science research.

www.eosc-life.eu

Main results: The aspects to be improved to reach the goal of a consistent, sustainable, cross-disciplinary tagging in life sciences are:

- Standard definitions of tags required
- Elimination of ambiguity between tags
- Guiding the number of tags assigned for a category
- Filling gaps particularly in missing "resource types" and "research fields"
- Elimination of meaningless tags
- Simplifying research stage
- Simplifying data type tags
- Guidance note and training needed for taggers

The result is a categorisation system with 55 tagging values spread over 6 dimensions (Figure 1), as compared to the original 93 tags spread over 8 dimensions.

Next step / take-home message

Avoiding polysemic bias is one of the most difficult issues to resolve and is necessary across disciplines.
This categorisation system will **need iterative improvement**. The pilot study on tagging intercalibration could serve as a model for a future cross-disciplinary workflow tagging system in Canonical Workflow Framework of Research (CWFR).

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

Reference: C. Ohmann, R. David, M. Cano Abadia, F. Bietrix, J.-W. Boiten, S. Canham, M.L. Chiusano, W. Dastru, A. Laroquette, D. Longo, M.Th. Mayrhofer, M. Panagiotopoulou, A. Richard, P. Verde, 2021, "Pilot study on the intercalibration of a categorisation system for FAIRer digital objects related to sensitive data in the life sciences", Data Intelligence Journal, *under review*.