



Pilot Study on the Intercalibration of a Categorisation System for FAIRer Digital Objects Related to Sensitive Data in the Life Sciences

Christian Ohmann, Romain David, Mónica Cano Abadia, Florence Bietrix,
Jan-Willem Boiten, Steve Canham, Maria Luisa Chiusano, Walter Dastrù,
Arnaud Laroquette, Dario Longo, et al.

► To cite this version:

Christian Ohmann, Romain David, Mónica Cano Abadia, Florence Bietrix, Jan-Willem Boiten, et al.. Pilot Study on the Intercalibration of a Categorisation System for FAIRer Digital Objects Related to Sensitive Data in the Life Sciences. *Journal of Data Intelligence*, 2022, 4 (2), pp.196 - 211. 10.1162/dint_a_00126 . hal-03663699

HAL Id: hal-03663699

<https://hal.science/hal-03663699>

Submitted on 10 May 2022

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.

Pilot Study on the Intercalibration of a Categorisation System for FAIRer Digital Objects Related to Sensitive Data in the Life Sciences

Christian Ohmann^{1†}, Romain David², Mónica Cano Abadia³, Florence Bietrix⁴, Jan-Willem Boiten⁵, Steve Canham¹, Maria Luisa Chiusano⁶, Walter Dastrù⁷, Arnaud Laroquette⁸, Dario Longo⁹, Michaela Theresia Mayrhofer³, Maria Panagiotopoulou¹, Audrey Richard¹⁰ & Pablo Emilio Verde¹¹

¹European Clinical Research Infrastructure Network (ECRIN), Paris 75013, France

²European Research Infrastructure on Highly Pathogenic Agents (ERINHA AISBL), Brussels 1000, Belgium

³Biobanking and Biomolecular Resources Research Infrastructure (BBMRI), Graz 8010, Austria

⁴European Infrastructure for Translational Medicine (EATRIS), Amsterdam 1081 HZ, The Netherlands

⁵European Advanced Translational Research Infrastructure (EATRIS)/ Lygature, Utrecht 3521 AL, The Netherlands

⁶European Marine Biological Resource Centre (EMBRC)—Department of Agricultural Sciences, University Federico II of Naples via Università, Naples 80138, Italy

⁷Department of Molecular Biotechnology and Health Sciences, Molecular Imaging Center, University of Torino, Torino I-10125, Italy

⁸European Marine Biological Resource Centre (EMBRC), Paris 75252, France

⁹European Research Infrastructure for Biological and Biomedical Imaging (Euro-BioImaging), Torino 10126, Italy

¹⁰European Research Infrastructure on Highly Pathogenic Agents (ERINHA AISBL), Brussels 1000, Belgium

¹¹Coordination Centre for Clinical Trials, Heinrich Heine University Düsseldorf Ringgold standard institution, Nordrhein-Westfalen 40225, Germany

Keywords: Sensitive data; CWFR; EOSC-Life; Life sciences; Toolbox; Categorisation system; Tagging; FAIR

Citation: Ohmann, C., et al.: Pilot study on the intercalibration of a categorisation system for FAIRer digital objects related to sensitive data in the life sciences. *Data Intelligence* 4(2), 196-211 (2022). doi: 10.1162/dint_a_00126

Received: July 6, 2021; Revised: November 24, 2021; Accepted: February 5, 2022

ABSTRACT

Sharing sensitive data is a specific challenge for research infrastructures in the field of life sciences. For that reason a toolbox has been developed, providing resources for researchers who wish to share and use sensitive data, to support the workflows for handling these kinds of digital objects. Common and community approved annotations are required to be compliant with FAIR principles (Findability, Accessibility, Interoperability, Reusability). The toolbox makes use of a tagging (categorisation) system, allowing consistent

[†] Corresponding author: Christian Ohmann (Email: christianohmann@outlook.de; ORCID: 0000-0002-5919-1003).

labelling and categorisation of digital objects, in terms relevant to data sharing tasks and activities. A pilot study was performed within the Horizon 2020 project EOSC-Life, in which 2 experts from 6 life sciences research infrastructures were recruited to independently assign tags to the same set of 10 to 25 resources related to sensitive data management and data sharing (in total 110). Summary statistics of agreement and observer variation per research infrastructure are provided. The pilot study has shown that experts were able to attribute tags but in most cases with a considerable observer variation between experts. In the context of CWFR (Canonical Workflow Frameworks for Research), this indicates the necessity for careful definition, evaluation and validation of parameters and processes related to workflow descriptions. The results from this pilot study were used to tackle this issue by revising the categorisation system and providing an updated version.

1. INTRODUCTION

The Horizon 2020 cluster project EOSC-Life brings together the 13 Life Science ‘ESFRI’ (European Strategy Forum on Research Infrastructures) research infrastructures (RIs) to create an open, digital and collaborative space for biological and medical research (<https://www.eosc-life.eu/>). Sharing sensitive data is a specific challenge within EOSC-Life. For that reason a toolbox has been developed, providing resources for researchers who wish to share and use sensitive data, to support the workflows for handling these kinds of digital objects (1). The sensitivity of the data normally stems from it being personal data, but can also be caused by intellectual property considerations, biohazard concerns, or because it falls within rules or protocols such as the Nagoya protocol (2). The toolbox does not create new content. Instead, it allows researchers to find existing resources that are relevant for sharing sensitive data across all participating research infrastructures (F in the **FAIR** Guiding Principles for scientific data management and stewardship, providing guidelines to improve the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of digital assets (3)). The toolbox provides links to recommendations, procedures, and best practices, as well as to software (e.g., tools and scripts supporting workflow) to support data sharing and reuse. It makes use of a tagging system, allowing consistent labelling and categorisation of digital objects, in terms most relevant to data sharing tasks and activities. The first version of the categorisation system was developed and its description published before executing the study (4). Components of the system were based upon existing classification systems (5-11). As a result of this pilot study, the categorisation system has been improved and simplified.

The work is relevant i) for developing ideas about tagging workflows within CWFR (Canonical Workflow Framework for Research) and ii) to the characterisation and use of FAIR digital objects (12). Handling of sensitive data and digital objects is a major issue and a challenging topic for many research infrastructures in the life sciences. The workflows around sensitive data are extremely complex and highly dependent on the regulatory environment. The toolbox is targeted at the discovery of resources dealing with sensitive data through a tagging system, which characterises the resources using cross-community approved categories. One of these categories provides the stage of digital objects in the data sharing life cycle. The toolbox also characterises, in broad terms, the data type investigated, contributing to the description of data collections as required in the context of CWFR. Because the work is spanning all life-science infrastructures, it also offers valuable input into the discussion of the applicability of generic workflow fragments across

infrastructures and an inter RIs approved tagging workflow system. Thus, the toolbox and the underlying categorisation system will contribute to the further development of CWFR in the area of digital objects with sensitive data.

2. RELATED WORK

Various resources have been developed to support the management of sensitive data. These include codes of conduct, guidelines, recommendations, policies, and descriptions of best practice, but also computer tools and services (e.g., for de-identification of data). For identifying these resources, various databases, catalogues and repositories are available and can be used, often with their own internal tagging system. For specific areas dedicated portals have been developed and implemented, such as the ELSI (Ethical, Legal and Social Implications) Knowledge Base from BBMRI (Biobanking and Biomolecular Resources Research Infrastructure) (<https://www.bbmri-eric.eu/elsi/knowledge-base/>) or the RDA (Research Data Alliance) COVID-19 Recommendations and Guidelines for Data Sharing (<https://www.rd-alliance.org/group/rda-covid19-rda-covid19-omics-rda-covid-19-epidemiology-rda-covid19-clinical-rda-covid19-0>). The latter incorporates a tagging system developed to characterise documents, allowing better support for searching and filtering. What is still missing is a system providing guided access to resources dealing with sensitive data in general and spanning the full area of life sciences.

3. PROBLEM FORMULATION

A large part of the compliance with FAIR principles, and the creation of minimal metadata sets required for FAIR data, are driven by community level processes (17, 18). Any toolbox being developed should be based on common and community approved metadata, and it should refer to standardised workflows and an integrated approach to arrive at community approval for FDO and CWFR concept requirements. The categorisation system under study is a crucial part of these common, appropriate and sufficient metadata (19). The objective of the pilot study was to evaluate the first version of the categorisation system by human experts from different life science RIs with respect to usefulness, reliability and consistency. From the results of the study an improved version of the categorisation system was provided and this will be used as a central component within the toolbox to support rapid and intuitive retrieval of resources by future users. Additionally, the study was intended to provide insights into handling discrepancies in developing and applying metadata, through a defined testing and community approval process.

4. APPROACH AND METHODOLOGY

The pilot study was performed within the EOSC-Life Work Package (WP) dedicated to sensitive data, supported by the partners of this WP and coordinated by the European Clinical Research Infrastructure Network (ECRIN) (<https://ecrin.org/>). The study protocol was registered before the study started on 8 December 2020 (16). The pilot study was organised using a strict methodology. 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 a given number of resources—25 for each of ECRIN, BBMRI, and EATRIS (European Advanced Translational Research Infrastructure in Medicine), and at least 10 for each of EMBRC (European Marine Biological Resource Centre), ERINHA (European Research Infrastructure on Highly Pathogenic Agents), and Euro-Bioimaging. In all cases they spanned 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 were relevant for their research infrastructure and then assessed these resources using the categorisation system independently of each other.

As a technologically agnostic first step, the evaluation was supported through the bibliographic tool Zotero Tags (<https://www.zotero.org/>). The categorisation system used in the pilot study consisted of eight dimensions: resource type, research field, research design, data type, stage in data sharing life cycle, geographical scope, specific topics or keywords and targeted group (see the short form of categorisation system in supporting material S1 and (4)). It was a requirement that all dimensions of the categorisation system were applied to each resource. Multiple tags per dimension were possible.

Summary statistics of agreements/disagreements between the two experts of each research infrastructure were generated. Separately for each research infrastructure and per category, cases where both experts agreed to assign a tag to a resource (yes-yes) and cases where one expert used a tag and the other not (yes-no or no-yes) were counted. The agreement rate was calculated as the percentage of “yes-yes” on all tags that were used at least once. For measuring interrater reliability between the assessment of experts of the same infrastructure, the kappa coefficient developed by Fleiss was applied (17). The statistical analysis was performed using the statistical software R version 4.0.2.

5. EXPERIMENTS AND ANALYSIS

The agreement rates between experts per infrastructure and category are presented in the supporting material (see supplementary material S2: Final study report). The rate was very high for BBMRI with 70.2%. For the remaining research infrastructures, the agreement rates were considerably lower with a range between 33.2% (ECRIN) and 20.2% (EMBRC). The agreement rates differed between the categories. Rates higher than 40% were achieved for category 2—research field (49.3%), category 6—geographical scope (44.5%) and category 1—resource type (41.8%). The lowest rates were measured for category 3—research design (18.3%) and category 5—stage in data sharing life cycle (25.1%). It is important to note here that some experts assigned many more tags from specific categories than other experts, contributing significantly to the disagreement.

The findings were confirmed by the interobserver reliability analysis (Table 1). A high kappa value was found for BBMRI (median kappa for 25 resources: 0.84). The kappa values were much lower for the other research infrastructures with a range between 0.44 for ECRIN and 0.22 for EMBRC (median kappa value). For BBMRI only one assessment was lower than 0.5 and only 9 assessments lower than 0.8. For ECRIN, assessment of 6 out 25 resources resulted in a kappa ≥ 0.50 . This was the case only for 2 out of 12 resources for EMBRC, for 2 out of 25 for EATRIS and for 1 out of 10 for ERINHA. All the other assessments resulted in kappa's less than 0.50. The results are summarised in Table 1.

Table 1. Interobserver reliability (Kappa values).

Resource*	Research infrastructure					
	BBMRI	EATRIS	ECRIN	EMBRC	ERINHA	Euro-Bio Imaging
1	0.79	0.47	0.39	0.09	0.33	0.18
2	0.81	0.32	0.31	0.11	0.29	0.20
3	0.84	0.36	0.13	0.24	0.55	0.33
4	1	0.24	0.73	0.08	0.12	0.33
5	0.54	0.21	0.26	0.22	0.03	0.24
6	0.10	0.36	0.34	0.16	0.42	0.28
7	0.86	0.18	0.47	0.55	0.28	0.30
8	0.64	0.42	0.45	0.42	0.41	0.26
9	0.75	0.14	0.20	0.51	0.12	0.36
10	0.94	0.20	0.47	0.22	0.31	0.33
11	0.81	0.44	0.79	-0.04	-	0.38
12	0.78	0.45	0.45	0.33	-	0.26
13	0.86	0.32	0.44	-	-	0.30
14	1	0.20	0.52	-	-	-
15	0.86	0.24	0.24	-	-	-
16	0.95	0.24	0.55	-	-	-
17	0.64	0.62	0.44	-	-	-
18	1	0.20	0.22	-	-	-
19	0.84	0.33	0.28	-	-	-
20	0.90	0.20	0.36	-	-	-
21	0.64	0.12	0.50	-	-	-
22	0.81	0.36	0.33	-	-	-
23	0.86	0.33	0.60	-	-	-
24	0.84	0.24	0.64	-	-	-
25	0.88	0.55	0.16	-	-	-
Median	0.84	0.32	0.44	0.22	0.3	0.3

Note: *Digital objects assessed were different between the research infrastructures <0 poor agreement, 0.0-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81-1.0 almost perfect agreement (>= 0.5: bold).

Several experts suggested new tags. This covered additional “resource types” that were not listed (e.g., webinar, template, Q&A), new “research fields” that were not included (e.g., law, ethics, sociology, genomic research), the answer “any” to “research design” and “data type”, more generic terms additional to specific “stages of data sharing” (e.g., data transfer, data sharing, secondary processing of data), additional “specific topics” (e.g., code of conduct, ELSI, COVID-19, material transfer agreement and additional “perspectives” (e.g., participant, resource provider).

The results from this pilot study were used to revise the categorisation system, generating an improved and simplified updated version 2 (Figure 1 and supplementary material S3: Short form of the categorisation system (Version 2)).

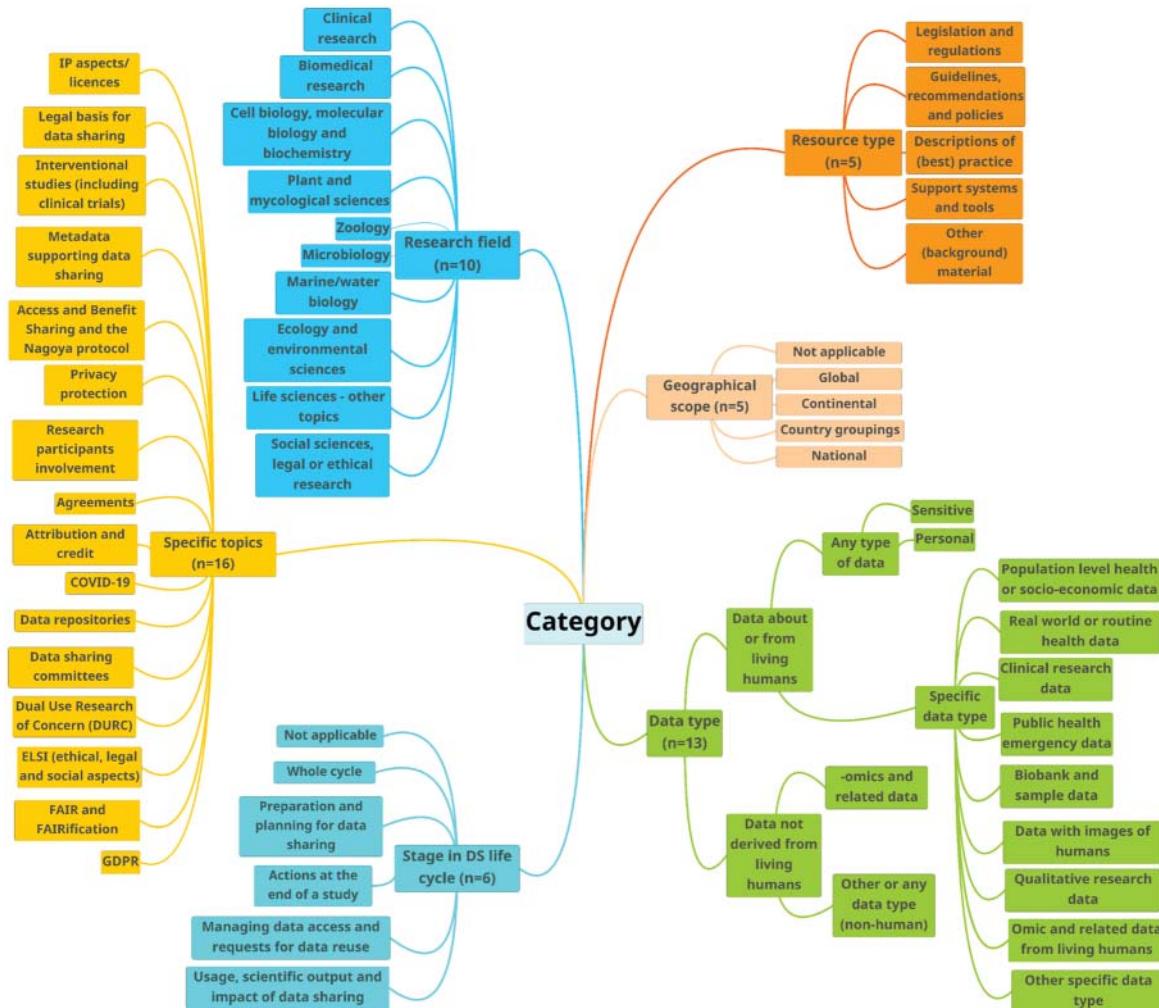


Figure 1. Mindmap of short version of categorisation system (version 2) (See also: Supplementary material S3 for presentation as a table).

6. DISCUSSION

BBMRI had already performed major work related to the handling of sensitive data in the field of health research, so the high interobserver correspondence in the pilot study is not surprising. The BBMRI ELSI Knowledge Base had already been structured using resource categories that included topic, areas of interest and geographical scope. The resources assessed in the pilot study had been selected from this knowledge base and existing tags were used in the assessment. The resources selected by BBMRI are closely related to the BBMRI work and concentrate on ELSI aspects, covering biobanks, de-identification, ethical handling, GDPR, consultation, transfer agreements, informed consent, etc. In contrast, some of the research

infrastructures deliberately tried to select a wide variety of resources, to create a more demanding test of the categorisation system (e.g., EATRIS).

With respect to the diverging results of the remaining five research infrastructures, the following mechanisms were identified as possible ways of improving the categorisation system, to achieve the goal of a canonical workflow for cross-disciplinary tagging in the life sciences:

- Training: The categorisation was discussed during several telephone conferences by the group and most experts involved in the assessment participated in the development of the system. Nevertheless, there was no systematic training on the application of the system before the pilot study.
- Standard definitions: In the development of the categorisation system, existing terminologies and classifications were taken into consideration (5-11). However, standardised definitions and a glossary were not provided. This was expressed as a deficit by several of the experts. The first study helped to showcase where the most misunderstandings and polysemic biases exist.
- Elimination of ambiguity: For some of the categories it was not totally clear how they should be applied to the resources. For example, category 8 (perspective) could be considered from the developer's perspective (the person who developed the resource) or from the user's perspective (the person at whom the resource is targeted). Similarly with category 6, a geographical scope could refer to the developers (e.g., produced in one country) or to the target users (e.g., global).
- Guiding the number of tags assigned for a category: The number of tags assigned to resources for a specific category was not uniform between experts. Ways of reducing these discrepancies (e.g. by putting a maximum on tag numbers) should be explored.
- Filling gaps: For several categories, important tags were seen as missing from the pre-specified list, for example, missing resource types (category 1), research fields (category 2) and specific topics (category 7). These gaps were considered in detail during the creation of version 2 of the tagging categories.
- Elimination of meaningless tags: For a considerable number of resources and most of the infrastructures "not applicable", "not specified" or "not clear" was assigned to category 3 (research design). It seems to be that specific study types are only relevant for some research infrastructures (e.g., ECRIN, EATRIS, EMBRC).
- Simplifying research stages: It proved difficult to allocate specific stages in the data sharing life cycle (category 5) to a number of resources despite using a widely used classification of stages in the data sharing life cycle. As a consequence, more generic tags were suggested that cover broader phases. This finding is relevant for the discussion on CWFR.
- Simplifying data type tags: For category 4 (data type), specific combinations of data types were selected for a considerable number of resources, indicating that more generic terms could improve applicability of the categorisation system. It may be worth exploring the idea of a hierarchy of tags, because in some cases both a generic and a specific term could apply. In addition, some specific data types are only relevant for some research infrastructures, in particular those dealing with human data. Again, this point is of relevance for CWFR.

Based on the analysis and comments of the experts, the categorisation system was recalibrated and simplified. Suggestions for additional tags were taken into consideration, but there was also a general request to keep things as simple as possible. In order to reduce the considerable interobserver misalignment between the experts, the definitions of the tags were clarified and the distinction between tags was improved. The value of the original 8 tagging dimensions was re-evaluated and it was concluded that two could be dropped. Where possible tags were combined to keep the number of available options as low as possible. The result is a categorisation system with 55 tagging values spread over 6 dimensions, as compared to the original 93 tags spread over 8 dimensions (Figure 1 and supplementary material S3: Short form of the categorisation system (version 2)). Nevertheless, discussion and further evaluation of the categorisation system is still ongoing, and improvements are under discussion for future versions.

Relevance of the pilot study for CWFR and FDO (FAIR Digital Objects) typology

- Data annotation workflows constitute a basic element for improving reuse (R1.3 of Wilkinson criteria) and a first step in building rich community approved metadata (F2 and F4) and FAIR vocabularies linked to other metadata (I2 and I3; (3)).
- The pilot study is relevant to proposals for cross-disciplinary methods in life sciences to define FDO typology that could be reused by the new FDO Semantics WG (FDO-SEM | FAIR Digital Objects Forum (fairdo.org)). These typologies will be necessary to evaluate what should be incorporated within FAIR metadata (F2 and R1).
- CWFR includes a discussion of the role of generic workflows and whether modelling across disciplines is possible (18). Developing the categorisation system for the toolbox has demonstrated an approach to tackling this problem in the context of the life-sciences infrastructures. This tagging system may be considered as a first level of interoperability for vocabulary typology (I2). The pilot study has demonstrated, however, considerable differences between research infrastructures, which has to be taken into consideration in the discussion on future generic workflows.
- Harmonisation around how data collections are being described is required in the context of CWFR (18). On a high level and across all life-sciences infrastructures, this issue has been investigated in the categorisation system by incorporating the dimension “data type”, covering various specific data types about or from living humans. Again, issues were observed that indicated the necessity for clear definitions of parameters and clear rules on how to apply them.
- As the CWFR initiative wants to discuss ways to reduce the large gap between technology and principles on the one hand and data practices in the labs on the other hand, the toolbox and its categorisation system constitute a practical example of how to improve interoperability between a large panel of (life sciences) disciplines. This exercise is based on real resources / digital objects, but as with other FAIRification processes it needs successive iterations and further community approval (19).
- The toolbox improves the discoverability (F in FAIR) of digital objects linked to sensitive data and thus will strengthen the provision of FAIRer digital objects because systematic application of a categorisation methodology will yield in metadata enrichment (for F4, and R1.3 compliance). This is a critical aspect for the handling of sensitive data in life sciences and to arrive at the necessary

standard solutions and workflows to comply with rules and regulations. The discussion around CWFR and FDO should be informed that this type of data is likely to need specific workflows to handle it. The toolbox will become a major resource to support this work by referencing relevant resources (F4).

- Consideration should also be given to the tagging workflow itself, and how it might be better standardised. Elements of this tagging process are: Selection of resources to be tagged, assigning taggers to resources, performing the tagging (under standardised conditions), quality control and approval of the tagging, and monitoring and supervision of the process. The pilot study demonstrated major variation in tagging of resources if independent taggers are assessing the same resource (inter-observer variation). The example of BBMRI has shown that with adequate measures this inter-observer variation can be considerably reduced. Another source of variance may occur if a tagger assesses the same resource at different time points. In order to come to a valid and reliable tagging workflow, both inter- and intra-observer variation have to be taken into account and be reduced. For example, the workflow of the tagging process should include an independent assessment by a second reviewer and a final consensus by both experts. An intercalibration tool for future taggers (with a training set) could be provided based on the first set of digital objects. Recurrent training could help to make the tagging workflow more replicable, although the tagging system itself should be mostly self-explanatory, as end users of the toolbox have to use the same system for querying purposes.
- In summary, the work is of major relevance for CWFR and FDO. An FDO-enabled application is able to identify the type of object (through reference to the object's type), directly operate on the object (through the object's location reference) or get more information about the object (through reference to the objects metadata records). A prerequisite is good, appropriate and sufficient metadata for an FDO, in our case characterising digital objects related to managing sensitive data. The categorisation system developed in the pilot study could evolve into a "methodological resources in life sciences" FDO type and could thus be of major help by linking services to digital objects. So far, FDO types are not well described in the life sciences and work has only been started in the FDO forum.

7. CONCLUSIONS AND FUTURE WORK

The first thing we demonstrated was that to have a relatively stable and generally applicable categorisation system across life sciences RIs, an iterative process was necessary to align the variation between experts from different RIs. As the categorisation system is part of an essential minimal metadata in this interdisciplinary study, this is of relevance for a "resources in life sciences" FDO and may also be significant for other FDO types (in particular more generic "FDO resources" type), such as FDOs describing workflows at a canonical level, particularly in the CWFR. In this context an important way to promote FAIRness is to request that all canonical components should support the concept of FDO (<https://fairdo.org/wg/fdo-cwfr/>). Secondly, the results from our study show that it will not be simple to define widely agreed data types across research domains and that pragmatic approaches are needed, which may carry the risk of a proliferation of data types. Thirdly, the pilot study demonstrated that highly sophisticated algorithms will be required to develop reliable automatic methods to interpret the categories relevant for sensitive data workflows (and indeed in the CWFR). It may be difficult, even impossible, to find purely machine-actionable algorithms for categorising digital objects dealing with managing sensitive data.

Avoiding polysemic bias is one of the more difficult issues to resolve and is necessary across disciplines. The results from this pilot study have been used to tackle this issue by revising the categorisation system and providing an updated version 2 (Figure 1, supplementary material S3). This version contains definitions for each term of the categorisation system, though these need to be further evaluated and approved by all life-sciences infrastructures. Currently, the 110 resources from the pilot study are being re-tagged with the new version of the categorisation system. This initial set of digital objects will be used as the content for the first demonstrator of the toolbox. This is a starting point, but the categorisation system will need iterative improvement. The pilot study on tagging inter-calibration could serve as a model for a future CWFR cross-disciplinary workflow tagging system.

ACKNOWLEDGEMENTS

This project has received funding from the European Union’s Horizon 2020 research and innovation programme “EOSC-Life—Providing an open collaborative space for digital biology in Europe” under grant agreement No. 824087.

AUTHOR CONTRIBUTIONS

C. Ohmann (christianohmann@outlook.de) was involved in conceptualisation, methodology and administration of the study, took part in the formal analysis and did the writing, partially based on R. David (romain.david@erinha.eu) inputs (original draft, review and editing, mind map visualization, and final manuscript). P.E. Verde (Pablo.Verde@med.uni-duesseldorf.de) was involved in the formal statistical analysis. M. Panagiotopoulou (Maria.Panagiotopoulou@ecrin.org) was involved in the mind map visualization. J.-W. Boiten (janwillem.boiten@lygature.org) provided project management, wrote first drafts of toolbox design documents, and organised video conferences. All authors were involved in the conceptualisation of the study, performed the investigation (tagging of the resources) and were involved in writing (review and editing). Regular video conferences were performed with all authors to discuss study progress.

DATA AND SOFTWARE AVAILABILITY

The detailed results of the pilot study are available at Data Intelligence’s data repository at the Science Data Bank, <http://www.doi.org/10.11922/scientedb.01529>, under an Attribution 4.0 International (CC BY 4.0).

REFERENCES

- [1] Boiten, J.W., et al.: EOSC-LIFE WP4 toolbox. Toolbox for sharing of sensitive data—A concept description. Available at: <http://doi.org/10.5281/zenodo.4483694> (2021). Accessed 22 November 2021
- [2] Secretariat of the Convention on Biological Diversity: Nagoya protocol on access to genetic resources and the Fair and equitable sharing of benefits arising from their utilization to the convention on biological diversity. Third World Network, Penang (2013)

- [3] Wilkinson, M.D., et al.: The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3(1), Article No. 160018 (2016)
- [4] Ohmann, C., et al.: EOSC-Life WP4 toolbox: Categorisation system for resources to be referenced in the toolbox for sharing of sensitive data (Version 1). Available at: <http://doi.org/10.5281/zenodo.4311094> (2020). Accessed 22 November 2021
- [5] OECD: Frascati Manual 2015: Guidelines for collecting and reporting data on research and experimental development, the measurement of scientific, technological and innovation activities. Available at: <http://dx.doi.org/10.1787/9789264239012-en> (2015). Accessed 22 November 2021
- [6] The world university rankings: Subject ranking 2015–2016: Life sciences top 100. Available at: https://www.timeshighereducation.com/world-university-rankings/2016/subject-ranking/life-sciences-0#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/stats (2016). Accessed 22 November 2021
- [7] Clarivate Analytics. Web of Science core collection help: Research areas (Categories/classification). Available at: https://images.webofknowledge.com/images/help/WOS/hp_research_areas_easca.html (2020). Accessed 22 November 2021
- [8] Australian Bureau of Statistics: Australian and New Zealand Standard Research Classification (ANZSRC). National Health and Medical Research Council Fields of Research (FOR). Available at: <https://www.abs.gov.au/Ausstats/abs@.nsf/Latestproducts/4AE1B46AE2048A28CA25741800044242?opendocument> (2008). Accessed 22 November 2021
- [9] Van den Eynden, V.: Research data management team. UK Data Archive: Data life cycle & data management planning. Available at: <https://www.ukdataservice.ac.uk/media/187718/dmplanningdm24apr2013.pdf> (2013). Accessed 22 November 2021
- [10] Ohmann, C, et al.: Classification of processes involved in sharing individual participant data from clinical trials [version 2; peer review: 3 approved]. *F1000Research* 7,138 (2018). Available at: <https://doi.org/10.12688/f1000research.13789.2>. Accessed 22 November 2021
- [11] United Nations, Department of Economic and Social Affairs: Statistics Division: Standard country or area codes for statistical use (M49). Available at: <https://unstats.un.org/unsd/methodology/m49/> (1999). Accessed 22 November 2021
- [12] Hardisty, A., Wittenburg, P. (eds.): Canonical Workflow Framework for Research (CWFR)—position paper—version 2, December 2020. Working paper. Available at: <https://osf.io/9e3vc/>. Accessed 20 June 2021
- [13] Wilkinson, M.D., et al.: Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Scientific Data* 6, Article No. 174 (2019).
- [14] Hettne, K.M., et al.: From FAIR leading practices to FAIR implementation and back: An inclusive approach to FAIR at Leiden University Libraries. *Data Science Journal* 19(40), 1–7 (2020)
- [15] Bahim, C., et al.: The FAIR data maturity model: An approach to harmonise FAIR assessments. *Data Science Journal* 19(41), 1–7 (2020)
- [16] Ohmann, C., et al.: EOSC-Life WP4 toolbox: Pilot study protocol for the evaluation of the categorisation system (Version 1). Available at: <http://doi.org/10.5281/zenodo.4311114> (2020). Accessed 22 November 2021
- [17] Fleiss, J.L.: Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5), 378–382 (1971)
- [18] Peer, L., et al.: Interactions about CWFR at RDA BoF. Available at: <https://osf.io/m2ejt/> (2021). Accessed 22 November 2021
- [19] David, R., et al.: The Research Data Alliance—SHAring Reward and Credit (SHARC) Interest Group: FAIRness literacy: The Achilles' heel of applying FAIR principles. *Data Science Journal* 19(1), 32 (2020)

AUTHOR BIOGRAPHY



Christian Ohmann has a graduation in Mathematics (Ph.D.), an interim examination in medicine and a habilitation in the field of Theoretical Surgery. He was the head of the Coordination Centre for Clinical Trials (KKS) at the University Duesseldorf, Germany (1999–2014) and is now retired. Since 2004 he has been involved in ECRIN, currently as Chair of the Network Committee and of the Independent Certification Board. He has major competence and experience in the field of clinical research and clinical research informatics and has participated in a series of EU-funded projects.
ORCID: 0000-0002-5919-1003



Romain David has a graduation in Oceanography and Data Mining (Ph.D.). He is Research Fellow—Data Manager at ERINHA AISBL (European Research Infrastructure on Highly Pathogenic Agents). He is a specialist in environmental information systems and has 20 years' experience both in terrestrial/marine life sciences engineering and database management—network animation. He is particularly interested in data mining, graph approach and data management with FAIR compliance. He is co-chair of 3 interest groups in the Research Data Alliance on data sharing, sensitive data and data stewardship professionalization and member of the EOSC Association Task Force “Fair metrics and data quality”.
ORCID: 0000-0003-4073-7456



Monica Cano Abadia, Ph.D. (Philosophy), works as ELSI Services Officer at the Ethical, Legal, and Societal Issues (ELSI) unit at BBMRI-ERIC. There, she coordinates the ELSI Helpdesk network and is in charge of the ELSI Knowledge Base. Her research interests at BBMRI include artificial intelligence and gender-related issues in biomedical research.
ORCID: 0000-0002-7726-9222



Florence Bietrix has been Head of Operations at EATRIS since 2017. After a Ph.D. in Life Sciences at Toulouse University, France in 2006, she worked as study director in the contract research organization Avogadro (Fontenilles, France), managing and coordinating pre-clinical studies related to new therapeutics for atherosclerosis and the metabolic syndrome. In 2008 she moved to the Netherlands and worked as postdoctoral fellow in the department of Medical Biochemistry of the Academic Medical Center, Amsterdam. During her post-doctorate, she focused on the development of new therapeutic targets for atherosclerosis and non-alcoholic steatohepatitis. In 2012 she joined EATRIS, the European Infrastructure for Translational Medicine, as Scientific Programme Manager.

ORCID: 0000-0002-0499-9435



Jan-Willem van Boiten, Ph.D. (Chemistry), is program manager at Lygature in The Netherlands dedicated to multi-stakeholder biomedical data infrastructure programs, such as TraIT (Translational Research IT), and the overarching national program for health data sharing, Health-RI. Internationally, Jan-Willem is involved in IMI/H2020 projects such as BIGPICTURE, FAIRplus, and Immune-Image, and participating in EOSC-Life on behalf of EATRIS. Before joining Lygature he held various research informatics positions in pharmaceutical industry.

ORCID: 0000-0003-0327-638X



Steve Canham has twenty years' experience working with clinical research systems and data, after previous roles in healthcare and teaching. He now works for ECRIN on several data related H2020 projects, and has a particular interest in managing metadata to promote FAIRness of data.

ORCID: 0000-0002-4409-8834



Maria Luisa Chiusano has a graduation in Biology (Ph.D.), Professor of Molecular Biology (University Federico II of Naples) and Associated Scientist of the Stazione Zoologica (Dohrn). She is an expert in omics and biodata management, responsible for biocomputing and bioinformatics services in a joint effort between two institutes. She is co-founder of the Italian Bioinformatics Society, councilor of the BIG DATA in HEALTH Italian Society, member in working groups for data sharing within EMBRC, EMSO-Italy, and ELIXIR-Italy, and member of the Steering Committee in national and international (included EU-funded) projects.

ORCID: 0000-0002-6296-7132



Walter Dastrù is a research fellow at the University of Torino. He has more than 20 years' experience in MR contrast agents, MRI acquisition and processing of MRI images. In particular he contributed to developing and maintaining software for the analysis of DCE-MRI data.

ORCID: 0000-0001-5757-7749



Arnaud Laroquette, M.Sc. (Biology and Biobanking), has been involved in the improvement of biobank management practices on scientific, quality, data, ethical and legal aspects. He is currently working as an ABS Compliance Officer at the European Marine Biological Resource Centre (EMBRC) in France. He is implementing the Access to genetic resources and the fair and equitable Sharing of Benefit arising from their utilization (ABS) principle.

ORCID: 0000-0001-6541-1394



Dario Longo (Ph.D.) is Head of the Research Unit of Turin of the IBB-CNR in Italy and lecturer in Cancer Imaging for the University of Turin, Italy. He has 20 years of experience in the molecular imaging field and in preclinical research developing novel imaging applications and processing tools. Currently he is responsible for image data management at the Euro-BioImaging Med-Hub. ORCID: 0000-0002-6906-9925



Michaela Th. Mayrhofer is a social scientist by training. Her Ph.D. studied the modes of governance of biobanks. Her research focus today lies on GDPR compliance, the ethics of AI and establishing and promoting ELSI best practices in biomedical research. Since 2019, she has been head of BBMRI-ERIC's ELSI Services & Research unit, member of several (S)EABs and partner in several EU-funded projects.

ORCID: 0000-0001-6932-0473



Maria Panagiotopoulou, Ph.D. (Biotechnology) is Project Manager at the European Clinical Research Infrastructure Network located in Paris, France. She is experienced in the fields of cell biology, nanotechnology, toxicology, genetics, genomics, biomedical imaging and clinical research. Currently, she is coordinating ECRIN's data projects portfolio focusing on developing tools, good practices and guidelines that serve the European clinical research community.

ORCID: 0000-0002-4221-7254



Audrey Richard conducted her Ph.D. in Lille, France, where she investigated the anticancer effects of parvoviruses. In 2013, she joined the Scripps Research Institute in Jupiter, FL, USA, as postdoctoral fellow where she specialized in the attachment and entry mechanisms of emerging and re-emerging infectious viruses, such as filoviruses (Ebola, Marburg viruses) and flaviviruses (Dengue, West Nile, Zika viruses, etc.). She came back to France in 2019 as Scientific Project Manager at the European Research Infrastructure for Highly pathogenic Agents (ERINHA) and became Director of ORCID: Operations in 2021.
ORCID: 0000-0002-0207-0139



Pablo Verde is the leader of the Biostatistics group at the Coordination Center for Clinical Trials and member of the Faculty of Mathematics and Natural Sciences at the Heinrich Heine University Düsseldorf. He has over 30 years of work experience in statistical analysis, and his current topics of research include: statistical software development and Bayesian hierarchical models with applications to meta-analysis of diagnostic test data and generalized evidence synthesis.
ORCID: 0000-0001-8625-7242