Developing the EOSC-Pillar RDM Training and Support Catalogue

Paula Oset Garcia[1][0000-0001-9506-766X], Lisana Berberi[2][0000-0002-7632-2466], Leonardo Candela[3][0000-0002-7279-2727], Inge Van Nieuwerburgh[1][0000-0002-7285-1598], Emma Lazzeri[4][0000-0003-0506-046X], Marie Czuray[5][0000-0002-0583-5299]

[1] Ghent University, Ghent University Library, Rozier 9, 9000 Ghent, Belgium
[2] Karlsruhe Institute of Technology, Steinbuch Centre for Computing (SCC), 76128 Karlsruhe, Germany
[3] Institute of Information Science and Technologies "Alessandro Faedo", via G. Moruzzi 1, 56124 Pisa, Italy
[4] Consortium GARR, Via dei Tizii, 6 - 00185 Roma
[5] Vienna University, Vienna University Library and Archive Services, Universitätsring 1, 1010 Wien

Abstract. Today's many research infrastructures and European projects offer training catalogues to store and list multiple forms of learning materials. In EOSC-Pillar project we propose a web application catalogue, which consists of training materials as well as day-to-day operational resources with the aim to support data stewards and other RDM (research data management), FAIR data (findable, accessible, interoperable, reusable) and open science actors. In this paper we briefly describe the scope and technical implementation of the EOSC-Pillar RDM Training and Support Catalogue and how we are addressing current challenges such as metadata standards, controlled vocabularies, curation, quality checking and sustainability.

Keywords: EOSC, Open Science, Research Data Management, FAIR, Training Material, Catalogue, Virtual Research Environment

1 Introduction

The EOSC-Pillar project [1] aims to coordinate national open science efforts in Austria, Belgium, France, Germany and Italy, building on each countries' specificities to ensure a well organised contribution to the European Open Science Cloud (EOSC) [2]. The goal is to facilitate the process of becoming EOSC service providers and contribute to the population of EOSC with useful services of wider European interest, based on the real needs and interests of the European scientific communities.

Some of the challenges addressed by the EOSC-Pillar project are narrowing the existing skills gap [3] [4] and promoting the cultural changes that are needed to realize the EOSC and the European Commission's vision for open science to become the 'modus operandi' within the European research community [5]. Moreover, the project seeks to promote FAIR (findable, accessible, interoperable, reusable) [6] data culture

and practices with specific tasks that tackle training and skills development. To this end, the EOSC-Pillar project has developed an online catalogue [7] with a curated selection of searchable training and support resources about FAIR, open science and research data management (RDM).

1.1 Defining the scope of the catalogue

Since the EOSC-Pillar project was proposed, an increasing number of institutions have appointed data stewards or similar roles to support researchers with their growing (FAIR) data management and open science requirements, and the need to professionalize such support roles has been recognized [8]. The number of projects and initiatives which promote and support the uptake of the FAIR data principles and open science practices has also increased (e.g. FAIRsFAIR [9], the EOSC Regional Projects, the European Research Infrastructures, etc.). These different teams and initiatives have made available an immense wealth of material to train and help data stewards in their role of supporting researchers with open science and good data management practices all along the research data lifecycle. However, these resources are scattered across different institutional or project websites. The added value by EOSC-Pillar is in curating and cataloguing these materials, and in increasing their visibility by making them findable through a central point by means of interoperable metadata.

The catalogue is aimed at teams of data stewards as well as other RDM support professionals in the first place (both new and consolidated teams), providing them with easy to find, readily available tools and solutions. However, the resources will likely be useful for other types of professionals with an active role in RDM, including researchers. The catalogue consists of conventional training materials as well as day-to-day operational resources and readily applicable tools that can be used by data stewards to support researchers. Therefore, a great diversity of resource types are considered: courses and course modules, tools, checklists, decision trees, audio-visual material, games, etc. Items in the EOSC-Pillar RDM Training and Support catalogue are both discipline agnostic as well as discipline-specific.

Catalogue records are added and managed by a small group of curators within the EOSC-Pillar project. The catalogue is integrated within the EOSC-Pillar virtual research environment (VRE) [10] which includes a social networking platform. For each item published in the catalogue, an automatic post is generated in the social networking area. This collaborative platform can be used by the catalogue members to suggest new records, to provide feedback or updates on existing items, to announce relevant training or events, etc.

1.2 Accessing the EOSC-Pillar RDM Training and support catalogue

The catalogue is publicly accessible via the EOSC-Pillar website under the “Resources” menu and the “RDM Training and Support catalogue” submenu. A landing page [11] has been created to provide contextual information about the catalogue as well as a link to access it. There are two ways to access and use the RDM Training and Support catalogue:

- Non-registered users can access metadata for all catalogue items by using the URL: https://eosc-pillar.d4science.org/web/eoscpillartrainingandsupport/catalogue.
- "Members" can browse metadata and access the resources behind each item, rate catalogue items and make use of the social networking area to suggest new resources or to comment on existing entries. To become members of the catalogue space, users can either register or sign in with an existing academic or social networking account, by simply clicking on the "Access the VRE" button that can be found on the URL above-mentioned. Signed-in users will be then redirected to the URL: https://eosc-pillar.d4science.org/group/eoscpillartrainingandsupport/home.

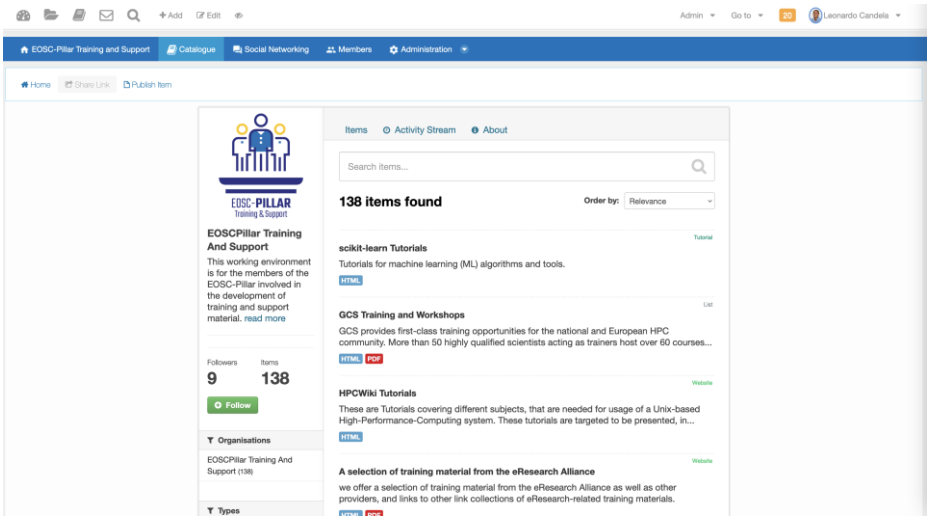

Fig. 1. A screenshot of the EOSC-Pillar RDM Training and Support catalogue search interface within the EOSC-Pillar VRE (admin user view). From this page, a catalogue member can browse the catalogue, access the social networking area and other services of the EOSC-Pillar VRE.

Through the catalogue interface (Fig.1) the search bar allows to look for different resources. The left side panel contains a series of filters based on item metadata which can also be used to browse catalogue content or to fine-tune the search results. For each record, the title and a snapshot of the resource description, together with the resource format(s) are given. The full metadata record can be accessed by clicking on the resource title.

2 Catalogue technical specifications

2.1 Technology

The catalogue has been developed and operated by the D4Science infrastructure [12]. Actually, a dedicated VRE has been deployed to support the needs of the community of practice focusing around the catalogue. The catalogue service is also a component of the EOSC-Pillar Federated Fair Data Space (F2DS) [13] tool set. It relies on CKAN

[14] technology, an open source software that allows to develop and manage open data catalogues. The advantage of this technology is that it can be extended to support specific communities with the definition of custom metadata profiles [15]. The approach to define the community-specific metadata elements is explained in section 2.2.

The technology behind the catalogue offers a series of services that facilitate open science and FAIRness. Each item in the catalogue gets a persistent identifier (namely a persistent uniform resource locator or PURL) and is described with rich and open metadata. The content in the catalogue is accessible via a graphical user interface (GUI), a programming interface based on representational state transfer (RESTful API) and it is also interoperable with other systems (e.g. DCAT and OAI-PMH). Adding new records to the catalogue is possible by harvesting items from other data sources, by using web APIs or by publishing items using a GUI.

2.2 Metadata profile

When the EOSC-Pillar project started, we researched a series of initiatives which were working to define a minimum and recommended set of metadata elements for training resources. We adopted the proposed specifications set in the then recent outputs from the FAIRsFAIR project [16]. Additional details on the metadata profile initially adopted can be consulted on the EOSC-Pillar project Deliverable 5.4 [17].

Throughout the course of the project, some of these initiatives have reached a more developed state of maturity and become accepted by the community. In particular, we highlight the work of the Research Data Alliance (RDA) Interest Group on Education and Training on Handling of Research Data (IG ETHRD) [18]. This group defined a minimal metadata set [19] to aid harmonised discovery of learning resources that has been recently adopted by the EOSC Future project [20] for the prospective EOSC training catalogue. To maximise interoperability, we are currently reviewing the metadata profile of the EOSC-Pillar RDM Training and Support catalogue, aligning it to the current specifications set by the EOSC Future project. A simple mapping between the EOSC-Pillar metadata profile and EOSC Future is presented in Table 1. As can be observed, not all elements in the EOSC-Pillar RDM Training and Support catalogue have a direct equivalent in the EOSC Future specification and vice-versa. We will work to expand the current metadata profile to include these elements and populate them to the extent possible, so this information can eventually flow to the EOSC catalogue.

Table 1. Table showing the mapping between the metadata fields in the EOSC-Pillar RDM Training and support catalogue and the EOSC Future specification, based in the RDA Minimal Metadata for Learning Resources. When an element is not present in one of the two specifications, this is indicated by the letters "NP" in the first and second columns.

EOSC-Pillar RDM Training and Support catalogue	EOSC Future (RDA Minimal Metadata for Learning Resources customized for the EOSC training resources)		
Property	Property	Metadata for existing resources	Metadata for new resources

Title	Title	Mandatory	Mandatory
Description	Abstract/ Description	Recommended	Recommended
PID	NP		
Resource Name	NP		
Resource URL	URL to Resource	Mandatory	Mandatory
Resource format	NP		
NP	Resource URL Type	Optional	Optional
Resource description	NP		
Author	Author(s)	Mandatory	Mandatory
Data Type	Learning Resource Type	Optional	Recommended
Tags	Keyword(s)	Recommended	Recommended
License	License	Recommended	Recommended
NP	Access Rights	Recommended	Mandatory
NP	Version Date	Recommended	Mandatory
Skills group	NP		
Domain	NP		
Level	Expertise Level	Recommended	Mandatory
Target group	Target Group (Audience)	Recommended	Mandatory
Language	Primary Language	Recommended	Mandatory
Country	NP		
NP	Learning Outcome(s)	Recommended	Mandatory

3 Outcomes and future work

Since the EOSC-Pillar RDM Training and Support catalogue was launched, the number of published items and users has increasingly grown. At the time of writing, the EOSC-Pillar RDM Training and Support catalogue lists over 130 records. Since the beginning of 2021, the number of registered users (VRE members) has grown from 54 to 141, and there have been 150 monthly average accesses to the EOSC-Pillar RDM Training and Support VRE. Significant efforts have been undertaken to give visibility to our work, by presenting and demonstrating the catalogue at different events and training courses. Examples of such user engagement activities are the Open Access Belgium week [21], the EOSC-Pillar RDM Training and Support Catalogue webinar [22], the FAIRsFAIR, EOSC-Pillar, EOSC-Synergy and Ghent University Instructor Training Workshop [23], or the recording of an episode for the Stories of Data podcast [24]. Both catalogue content curation and addition as well as promotion are ongoing activities.

In recent years, many European funded projects and European Research Infrastructures have developed training catalogues or are planning to do so. These different projects regularly engage with each other through different platforms such as

the OpenAIRE Community of Practice for Training Coordinators (CoP) [25] or the EOSC Training and Skills Working Group [26]. Through a series of meetings and events [27] [28] [29], existing or under development catalogues have exchanged knowledge and experiences and discussed common challenges.

A prevailing issue affecting training catalogues is that suitable controlled vocabularies to richly describe data stewardship learning resources are lacking, still under development or have not been agreed upon by the community. In the EOSC-Pillar RDM Training and Support catalogue, it is possible to define controlled vocabularies for each custom (profile) metadata field. Some vocabularies currently used are the FAIR4S framework [30] for "Target group" or the re3data.orgsubject classification [31] for "Domain". The project is aware of alternative possibilities for these and other metadata values (e.g. terms4FAIRskills [32], FAIRSharing Subject Resource Application Ontology SRAO [33], Learning Object Metadata Standard [34]). We will closely follow the evolution of relevant projects, and pay particular attention to decisions made by EOSC Future, to implement the necessary changes and be in line with community agreements and enhance the semantic interoperability of the catalogue.

Another challenge that is common to training catalogues is related to the curation and quality checking of the catalogue items. Some related projects (e.g. SSHOC [35]) organize yearly sprints to review the catalogue content and have developed internal guidelines and documentation. EOSC-Pillar has an embedded user rating system which allows us to gather information about the usefulness of a resource. But, in general, there is a lack of a systematic approach for curation and quality checking of catalogues for training materials. Besides, these tasks are often performed by smaller teams within the project, often without visibility or recognition, leaving the process to a great degree of subjectivity. Efforts are now being started in order to address this challenge by setting a working group on quality assurance of training materials within the OpenAIRE CoP, which aims to co-create guidelines and checklists for the curation and quality checking of training materials.

Finally, the long term sustainability of the results presented is a common challenge arising from the intrinsic time duration and funding availability of the project. In the case of the RDM Training and support catalogue, it will remain accessible to the catalogue administrators and users for at least two years after the project ends. Besides, significant efforts have been put into ensuring syntactic and semantic interoperability, since enabling the machine-to-machine exchange of the catalogue content is the best way to facilitate its future availability.

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme under EOSC-Pillar project (grant agreement No. 857650).

References

[1] "EOSC-Pillar project," [Online]. Available: https://www.eosc-pillar.eu/.

[2] "What the European Open Science Cloud (EOSC) is," [Online]. Available: https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en.

[3] European Commission, Directorate-General for Research and Innovation, "Digital skills for FAIR and Open Science: report from the EOSC Executive Board Skills and Training Working Group," Publications Office, 2021.

[4] European Commission, Directorate-General for Research and Innovation, "Realising the European open science cloud: first report and recommendations of the Commission high level expert group on the European open science cloud," Publications Office, 2016.

[5] European Commission, Directorate-General for Research and Innovation, "Horizon Europe, open science: early knowledge and data sharing, and open collaboration," 2021.

[6] M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. da Silva Santos and Bonino, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, C. t Hoen and Peter A., R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons, "The FAIR Guiding Principles for scientific data management and stewardship.," *Scientific data,* vol. 3, p. 160018, Mar 2016.

[7] "EOSC-Pillar RDM Training and Support catalogue," [Online]. Available: https://eosc-pillar.d4science.org/web/eoscpillartrainingandsupport/catalogue.

[8] European Commission, Directorate-General for Research and Innovation, "Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data," Publications Office, 2018.

[9] "FAIRsFAIR project," [Online]. Available: https://www.fairsfair.eu/.

[10] L. Candela, D. Castelli and P. Pagano, "Virtual Research Environments: An Overview and a Research Agenda," *Data Science Journal,* vol. advpub, pp. GRDI-013, 2013.

[11] "EOSC-Pillar RDM Training and Support catalogue landing page," [Online]. Available: https://www.eosc-pillar.eu/rdm-training-and-support-catalogue.

[12] M. Assante, L. Candela, D. Castelli, R. Cirillo, G. Coro, L. Frosini, L. Lelii, F. Mangiacrapa, P. Pagano, G. Panichi and F. Sinibaldi, "Enacting open science by D4Science," *Future Generation Computer Systems,* vol. 101, pp. 555-563, 2019.

[13] N. Cazenave, L. Candela, L. Berberi, J. van Wezel, A. Hashibon and Y. Le Franc, *EOSC-Pillar D5.1 FAIR Research Data Management tool set,* Zenodo, 2020.

[14] "CKAN homepage," [Online]. Available: https://ckan.org/.

[15] L. Candela, L. Frosini, Y. Le Franc, F. Mangiacrapa and N. Cazenave, *EOSC-Pillar D5.2 FAIR Research Data Management Workbench Operation Report,* Zenodo, 2021.

[16] E. Newbold, G. Kayumbi Kabeya, B. Matthews, J. Davidson, P. Herterich, A. Whyte and L. Molloy, *D6.2 Initial Core Competence Centre Structures,* Zenodo, 2020.

[17] P. Oset Garcia, E. Lazzeri, I. Van Nieuwerburgh, P. von Hartrott and J. Geistberger, *EOSC-Pillar D5.4 FAIR-oriented research data management Support Training and Assessment Activity Report,* Zenodo, 2021.

[18] "Research Data Alliance (RDA) Education and Training on Handling of Research Data Interest Group (IG ETHRD) homepage," [Online]. Available: https://www.rd-alliance.org/groups/education-and-training-handling-research-data.html.

[19] N. Hoebelheinrich, K. Biernacka, M. Brazas, L. J. Castro, N. Fiore, M. Hellström, I. Kuchma, E. Lazzeri, E. Leenarts, P. M. Lavancy, E. Newbold, A. L. Nurnberger, E. Plomp, L. Vaira, C. van Gelder and A. Whyte, *Recommendations for a minimal metadata set to aid harmonised discovery of learning resources,* RD Alliance, 2022.

[20] "EOSC Future wiki page (Training Catalogue - Minimal Metadata for Learning Resources)," [Online]. Available: https://wiki.eoscfuture.eu/display/PUBLIC/Training+Catalogue+-+Minimal+Metadata+for+Learning+Resources.

[21] "Open Access Belgium week (2020)," [Online]. Available: https://www.eosc-pillar.eu/events/eosc-pillar-open-access-week-belgium.

[22] "EOSC-Pillar RDM Training and Support Catalogue webinar," [Online]. Available: https://www.eosc-pillar.eu/events/webinar-eosc-pillar-rdm-training-and-support-catalogue.

[23] "FAIRsFAIR, EOSC-Pillar, EOSC-Synergy and Ghent University Instructor Training Workshop," [Online]. Available: https://www.eosc-pillar.eu/events/instructor-training-online-workshop.

[24] "Stories of data podcast," [Online]. Available: https://www.eosc-pillar.eu/stories-data-open-science-podcast.

[25] "OpenAIRE Community of Practice for Training Coordinators (CoP) homepage," [Online]. Available: https://www.openaire.eu/cop-training.

[26] "EOSC Training and Skills Working Group homepage," [Online]. Available: https://eoscsecretariat.eu/working-groups/skills-training-working-group.

[27] E. Leenarts, C. van Gelder, I. Kuchma, R. Braukmann, L. Molloy, A. Lister, E. Newbold, P. O. Garcia, V. Garnett, L. Vaira, I. van Nieuwerburgh and A. Whyte, *OSF workshop resources RDM training and support catalogues (21 September 2021),* Zenodo, 2021.

[28] E. Newbold, G. Kayumbi, A. Whyte and E. Lazzeri, *Summary Report: Workshop on Harmonising Training Resource Metadata for EOSC Communities,* Zenodo, 2021.

[29] "EOSC5b-TF/FAIRsFAIR Training resource catalogue interoperability workshop event page," [Online]. Available: https://www.fairsfair.eu/events/training-resource-catalogue-interoperability-workshop.

[30] A. Whyte, J. de Vries, E. Leenarts, R. Thorat, E. Kuehn, G. Sipos, V. Cavalli, V. Kalaitzi and K. Ashley, "D7.3: Skills and Capability Framework," 2018.

[31] "The re3data.org subject classification," [Online]. Available: https://www.re3data.org/browse/by-subject/.

[32] "terms4FAIRskills GitHub repository," [Online]. Available: https://terms4fairskills.github.io/.

[33] "FAIRSharing Subject Resource Application Ontology (SRAO) GitHub repository," [Online]. Available: https://github.com/FAIRsharing/subject-ontology.

[34] "Learning object metadata," Wikipedia, Wikimedia Foundation.

[35] "SSHOC project," [Online]. Available: https://sshopencloud.eu/.