EOSC-Life -D1.3 EOSC-Life EOSC FAIR services deployment for open calls

Helen Parkinson, Philip Gribbon, Ugis Sarkans, Gesa Witt, Andrea Zaliani, Manfred Kohler, Jason Swedlow, Jean-Marie Burel, Morris Swertz, Esther van Enckevort, et al.

▶ **To cite this version:**

Helen Parkinson, Philip Gribbon, Ugis Sarkans, Gesa Witt, Andrea Zaliani, et al.. EOSC-Life -D1.3 EOSC-Life EOSC FAIR services deployment for open calls. D1.3, Fraunhofer-Institut für Offene Kommunikationssysteme (FOKUS Fraunhofer); EMBL; CSIC; VU; BBMRI; KNAW; UVEG; USMI; IMG; UNIMAN; LUMC; EATRIS; UNIMIB; EBI; ECRIN; EMBRC; EMPHASIS (FZJ); ERINHA; INFRAFRONTIER; INRAE; UNIVDUN; HMGU; CERBM; BSCRC; UOULU; CRRMMP. 2022. hal-04161824

HAL Id: hal-04161824
https://hal.science/hal-04161824v1

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

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

D1.3 – EOSC FAIR services deployment for open calls

WP1 – Publishing FAIR RI data Resource
Lead Beneficiary: Fraunhofer, EBI
WP leaders: Philip Gribbon and Helen Parkinson
Contributing partner(s): EMBL, BBMRI, EATRIS, ECRIN, EMBRC, EMPHASIS (FZJ), ERINHA, INFRAFRONTIER, UNIMIB, INRA, UNIVDUN, HMGU, CERBM, BSCRC, UOULU, CRRMMP, CSIC, VU, KNAW, UVEG, USMI, IMG, CNRI, UNIMAN, LUMC, UNITO

Authors of this deliverable: **Helen Parkinson, Phil Gribbon, all authors listed below**

Contractual delivery date: **28 February 2022**
Actual delivery date: **21 February 2022**
H2020-INFRAEOSC-2018-2

Grant agreement no. 824087
Horizon 2020
Type of action: RIA

Authors of this deliverable:
WP1: Helen Parkinson (EMBL-EBI, ELIXIR), Ugis Sarkans (EMBL-EBI, ELIXIR, Euro-BioImaging), Phil Gribbon, Gesa Witt, Andrea Zaliani, Manfred Kohler (Fraunhofer, EU-OPENSCREEN), Jason Swedlow, Jean-Marie Burel (UNIVDUN, EUBI), Morris Swertz, Esther van Enckevort (UMCG, BBMRI), Petr Holub (BBMRI ERIC), Marzia Massimi, Rafaele Matteoni (CNR, INFRAFRONTIER), Holger Maier(HMGU, INFRAFRONTIER), Reetta Hinttala, Anne Heikkinen (UOulu, INFRAFRONTIER), Philipp Gormanns (INFRAFRONTIER GMBH, INFRAFRONTIER), Laurent Vasseur, Sophie Leblanc, Yann Herault (CERBM-GIE, INFRAFRONTIER), Dimitris Kontoyiannis, Christina Chandras, Dimitra Panou (FLEMING, INFRAFRONTIER), José Miguel López Coronado, Rosa Aznar Novella (UVEG, MIRRI), Vincent Robert, Ammar Ben Hadj Amor (KNAW, MIRRI), Serge Casaregola, Jean-Luc Legras, Michel-Yves Mistou (INRA, MIRRI), Paolo Romano (USMI, MIRRI), Isabelle Perseil (INSERM, ERINHA), Romain David (ERINHA), Roland Pieruschka (FZJ, EMPHASIS), Katrina Exter, Marc Portier, Cedric Decruw (VLIZ, EMBRC), S. Canham, C. Ohmann, S. Goryanin (ECRIN-ERIC), Laura Del Cano (CSIC, Instruct), Maddalena Fratelli (IRFMN, EATRIS), Carole Goble, Stuart Owen, Stian Soiland-Reyes (UNIMAN), Nick Juty (UNIMAN, ISBE), Henriette Harmse (EMBL-EBI, ELIXIR), Dario Longo (CNR, Euro-BioImaging)

WP2: Carole Goble, Nick Juty (UNIMAN)

WP6: Susanna Sansone, Allyson L. Lister, Peter McQuilton, Milo Thurston, Ramon Granell, Hossein Mirian (UOXF, ELIXIR), Marco Roos (LUMC), Luiz Bonino (LUMC)

WP1 Open Call Projects Personnel:

APPID 1244 - PombeMine: A FAIR workflow-enabled PomBase in the cloud	Gos Micklem
	Daniela Butano
	Rachel Lyne
	Valerie Wood
APPID 1211 - OmicsDI Cloud	Henning Hermjakob
	Yasset Perez-Riverol
	Gaurhari Dass
APPID 1231 - Macromolecular crystallography data management in the cloud	Edward Daniel
	Lari Lehtiö
	Kristian Koski
	Mirko Maksimainen
	Rikkert Wierenga

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

	Tommi Nyrönen
	Orly Dym
	Joel Sussman
	Jaime Prilusky
	David Hall
APPID 1234 - PDB-REDO Cloud: FAIR protein structures with deep versioning for scientific reproducibility and data provenance tracking	Anastassis Perrakis
	Robbie Joosten
APPID 1228 - Cloudification of BBMRI-ERIC CRC-Cohort and its Digital Pathology Imaging	Petr Holub
	Heimo Müller
	Luca Pireddu
	Francesca Frexia
	Jason Swedlow
	Tomáš Brázdil
APPID 1220 - Cloudification of the IMPC dataset for workflow application	Philipp Gormans
	Jeremy Mason
	Robert Wilson
	Sophie LeBlanc
	Yann Herault
APPID 1209 - Developing a federated XNAT portal for medical image datasets	Sara Zullino
	Kranthi Thej Kandula
	Dario Longo
	Jean-Karim Heriche
	Yi Sun

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

	Walter Dastrù
APPID 1237 - Cloudification of CryoEM data and metadata	Jiri Novacek
	Carlos Óscar Sorzano
	José María Carazo
	Laura del Caño
	Aleš Krenek
	Lukáš Hejtmánek

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

Table of Contents

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

Executive Summary

This deliverable summarises the work of WP1 to deliver Findable, Accessible, Interoperable and Re-usable (FAIR) services in the context of EOSC-Life's funding calls, using these to improve FAIR services, service uptake and to inform sustainable development and future use. We describe service delivery and development around the FAIR principles and present the funded projects which have driven our implementation of FAIR Services. We address sustainability and describe the processes used to engage the EOSC-Life funded projects, as well as future work.

Project Objectives

This deliverable has contributed directly to the following WP1 objectives:

a. Development of cloud compatible FAIR-compliant data resources
b. Advance the evolution of RI repository infrastructure for EOSC (sustainability) and the interfaces between the repositories supporting the RI demonstrators and open calls
c. Implementation of FAIR services and WP1 standards for RI data/data resources

Detailed Report on the Deliverable

1. Introduction

This deliverable describes EOSC-Life's FAIR services implementation in support of WP3 demonstrators, open calls and WP1 calls, all of which fall within the scope of EOSC-Life's service needs and provide a comprehensive set of needs representative of the life sciences community. Due to the pandemic, there were delays in the launching of calls and also contractual changes within the project for later calls, with the result that the timing of the open calls has changed since the plans at the start of the project. WP1 has therefore changed its work in response to this changing situation and we provide information on the calls (Appendix 1) to provide context for our work towards this deliverable.

WP3 calls recruited new community/RI led proposals to EOSC-Life in 2021. There are three relevant calls held by WP3, a broad digital life sciences call[1], a sensitive data call[2] and an academic-industry collaboration[3] call (details provided in Appendix 1), however, the last two calls OC2 and OC3 (comprising just three funded projects) were not complete in time to include analyses in this deliverable and will therefore be addressed as future work.

[1] https://www.eosc-life.eu/opencall/
[2] https://www.eosc-life.eu/sensitivedatacall/
[3] https://www.eosc-life.eu/industrycall/

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

In the execution of this deliverable, we have worked closely with WP3, providing WP1 expert engagement to define their calls and WP1 experts worked with applicants to refine the proposals ensuring they were informed about EOSC-Life outputs and community and also informing WP1 members about the open call applicant's needs for services. Finally, the WP1 experts were involved in technical feasibility analysis of projects which provided insight into needs and state of the art for FAIR services. This was included as a prelude to the external scientific review. As previously, our approach has been to extend existing services in an attempt to be sustainable and to develop new services only when necessary.

In parallel with WP3 open calls (OC1, 2 and 3) and, informed by our collaboration with WP3, we have designed and delivered services, adapted and extended these and provided training and collaborative activities, resulting in improved FAIR cloud services for the calls and also available more widely to an international BioMedical user community, thereby maximising the value of investment in EOSC-Life.

The WP3 call scope and EOSC-Life's scope are necessarily large as it mirrors the complexity and granularity of biomedical science. Therefore, in the interaction and delivery of support for open call projects we have deployed the WP1 data experts (described in D1.1, who have a broad range of competencies across biomedical domains) in the application of FAIR and in delivery of FAIR services. These were expected to cover the majority of needs for the Open Call projects and to date this strategy has been successful, with occasional recruitment of additional experts when needed, for example, to increase the number of experts in sensitive data services in response to WP3 call focus.

WP1 experts were also involved in technical analysis[4] of open call applications with the aim of ensuring funded projects are technically feasible, enabling them to take a forward look at the needs for FAIR services as well as contributing to WP3's technical feasibility analysis. In supporting the WP3 open calls we also note that the EOSC wider landscape is changing, specifically the EOSC approach to FAIR services is maturing and, in our conclusion, we have aligned our work to future EOSC activities demonstrating their relevance and the utility of our approach.

2. Landscape Analysis and Use Cases for EOSC-Life WP3 calls and alignment wider EOSC-Life/Future EOSC activities

EOSC-Life has a need for broad discovery FAIR services given the breadth of the domain. In this deliverable, we present mainly Findable, Accessible and Interoperable services as the majority of the projects had requirements in this space. Re-use is demonstrated by a collection of example scientific workflows based on data available in public resources (imaging, ontological, etc.) and published in the WorkflowHub, a registry of scientific workflows supported and developed by WP2.

To satisfy the 'Findable' and 'Accessible' components of FAIR, leveraging existing (or emerging) domain data repositories, knowledge bases and archives are required, and a metadata registry therefore becomes an essential FAIR service to 'Find' datasets (as described in D1.2). FAIRsharing

[4] https://instruct-eric.com/network/eosc-life/eosclife-digital-life-sciences-open-call-technical-evaluation-guidelines

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

records detailed licensing information allowing users to determine the terms of access, an important component of academic-industry collaboration.

Interoperability services are needed to address biomedical domain complexity and the variable granularity of the datasets, the multispecies nature of projects and the variability in data generating technologies, from genome sequencing for individual samples from marine environments to clinical data. As many of the WP1 call projects combine biomedical domains, for example, spanning research infrastructures the semantic interoperability becomes important and use of semantic services to 'translate' or 'map' the data into a common terminology, or more formally an 'ontology' is needed to annotate meta data and has been a focus of this deliverable.

To guide projects satisfying the re-usable components of FAIR, we have developed a set of scientific workflows mainly as Jupyter notebooks.

Those examples use existing FAIR public resources - e.g. Image Data Resource, HumanMine -, they can be run in Cloud base environments and have been published to the WorkflowHub[5], a registry of scientific workflows supported and developed by WP2.

The workflows have been, when possible, linked to published scientific papers[6], exploring analytical results associated to a SARS-CoV2 study published by Ellinger et al, and have been used and demonstrated during training sessions for the open calls teams and outside the EOSC-Life audiences.

The training materials are publicly available either via guides and/or videos and will continue to evolve as new workflows are added.

Those examples have proven to be useful for open calls projects. We have now actively engaged with some open calls teams to develop scientific workflows e.g. ongoing work with APPID 1244 PombeMine (see below for details of the project).

One of the aims of the work was also to raise awareness amongst the new projects, of the EOSC-Life ecosystem since it became clear during the initial interviews that participants were missing information allowing them to complete their projects.

Another aspect that became clearer is that some open calls require guidance on research data management. To benefit the community and not only the participants of the open calls, some WP1 experts have contributed to the content of RDMkit[7], an online guide containing good data management practises applicable to research projects from the beginning to the end. We did not include the work on scientific workflows in the first phase of the contributions to RDMkit. Plans have already been made with colleagues from WP2 to include such work in a follow-up phase.

In assessing each of the projects' needs we have used the kick off workshops discussions in project preparation, regular feedback via WP1 meetings and WP3 meetings, project specific questionnaires as well as analysis of the proposals submitted, particularly for most recent WP3 calls. We have also arranged 1:1 discussions with each open call project to discuss their plans and FAIR services needs in detail. These discussions also included experts from WP2 to assure that the open call projects utilise the technologies and registries recommended by WP2 when integrating

[5] https://workflowhub.eu/
[6] https://workflowhub.eu/workflows/238
[7] https://rdmkit.elixir-europe.org/

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

their resources into workflows. Moreover, these meetings allowed the direct support by various experts on particular questions from the open call project members. Based on the outcomes from these discussions WP1/2/3 started to organise a FAIR hackathon. In parallel we have used our membership of EOSC association activities to place the semantic services we offer in the wider context of EOSC with the aim of increasing their use and evolving them.

2.1. WP Implementation/support for the WP1/WP3 open calls

Appendix 1 provides a list of eight WP1 and eight WP3 projects related to this deliverable but doesn't include the final three WP3 funded projects. Engagement with projects in both WP1 and WP3 calls followed a common model from the WP1 perspective: Co-design of the call specification, a technical feasibility review from WP1, independent review by a panel of invited experts.

WP3 projects underwent an additional maturation phase in collaboration with relevant EOSC-Life WPs, including WP1. These discussions allowed us to determine service needs, refine the scope of this deliverable and ensured that funded projects were technically feasible, contributing to the success of the WP3 open calls.

In total we present a summary of subsequent improvements, as well as the final objectives for the 8 successful WP1 Open Call projects (Table 1).

Use case	Summary Achievements and final objectives
WP1 Open Call projects	
APPID 1244 PombeMine: A FAIR workflow enabled PomBase in the cloud	A new database, PombeMine, was developed. The resource contains sequence features and curated data including the Gene Ontology, phenotypes, protein domains, physical and genetic interactions, disease associations and orthologue data. A set of predefined editable "template" searches with default parameters now allows researchers to easily access these data and act as starting points for creating new queries. The resources will be added to the FAIRsharing collection and workflows published in Jupyter notebooks will link PombeMine to the IDR through the API.
APPID 1211 OmicsDI Cloud	OmicsDI has been extended to support deployment of a local satellite instance and a pilot site demonstrator[8] is implemented with metadata from external systems biology studies. Data from BioModels in the site is updated daily, while data from Cell Collective and Physiome are periodically updated. Current test datasets on FAIRDOMhub will be

[8] http://wwwdev.ebi.ac.uk/Tools/omicsdi/

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

	migrated to production status in the final project stages. Simplified algorithms for calculation of OmicsDI impact metrics are now established via an API for use by third party websites from data providers. This API is now used in production mode by two third party sites, BioModels[9] and PRIDE[10]
APPID 1231 Macromolecular crystallography data management in the cloud	The IceBear macromolecular crystallography data management software has been set up in the cPouta environment maintained by the CSC IT Center for Science (ELIXIR Finland) and as a backup instance at University of Oulu. This now allows for images and metadata to be uploaded to the cloud from the automated imaging instruments, elevating Access. Communication and exchange of metadata has been implemented to ExiMC ISPyB of MAXIV synchrotron, improving Findability. The standalone IceBear can be downloaded[11] and the cloud version along with training will be made available in 2022 to collaboration sites starting with Finnish Biocentres and the Weizmann Institute.
APPID 1234 PDB-REDO Cloud: FAIR protein structures with deep versioning for scientific reproducibility and data provenance tracking	The PDB-REDO data has been converted to relevant community standard and FAIR formats including mmCIF, PDB, MTZ and JSON. This elevated FAIRness is supported by full provenance tracking of the entries (PDB models and diffraction data) and software used in calculations. Greatly enhanced documentation of PDB-REDO data has been achieved by adding JSON schemas to all stored metadata further improving findability. Initial deployment include a 3D-beacon for structural data (the AlphaFill databank[12]) at the host institute which will be extended to PDB-REDO data. Users can access the current PDB-REDO data at https://pdb-redo.eu and downloadable data-associated JSON schemas are provided at https://pdb-redo.eu/download-info.html. For each PDB-REDO entry, enhanced provenance data is available at https://pdb-redo.eu/db/{identifier}/versions.json. Links to resources and standards can be found through the FAIRsharing record[13]. Finalisation of the project will include a frontend for dataset creation based on PDB-REDO metadata queries and creation of a data structure describing sets of

[9] https://www.ebi.ac.uk/biomodels/MODEL4780784080
[10] https://www.ebi.ac.uk/pride/archive?keyword=PXD005011
[11] https://www.icebear.fi/releases/
[12] https://alphafill.eu
[13] https://beta.fairsharing.org/3301

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

	PDB(-REDO) entries that can be used for documenting research data in publications and other scientific output
APPID 1228 - Cloudification of BBMRI-ERIC CRC-Cohort and its Digital Pathology Imaging	To support cloudification of whole slide images (WSI), a set of containerized data conversion tools[14,15,16] have been established as part of the project's secure image deposition service. The actual conversion of data sets exploits an open model formalism (openEHR) associated to the OMOP ontology. For data querying, a Locator/Bridgehead component has been deployed whilst dataset export occurs within a Snakemake workflow system and uses secure two-factor authentication. AI-based workflows for have been established for both tile-based and whole-WSI based image analysis. The CRC-Cohort is now part of FAIRsharing.org EOSC-Life collection DOI[17] and workflows cover CRCC modelling are publicly deployed[18], with FAIR CRCC resources registered in the WorkflowHub[19,20]. A publication addressing the needs for WPI anonymization will be submitted at the end of the project
APPID 1220 - Cloudification of the IMPC dataset for workflow application	A cloud native API was developed to support cloudification of mouse phenotype data. The application provides a REST service and supports hypermedia aware browsers. Development was informed by use-case analyses and included an extension of the existing IMPC ETL process so phenotype data could be loaded directly into a distributed highly available database in a scalable manner. This application has been deployed into a Kubernetes cluster and is publicly available. The API will be further iteratively extended to ensure support for downstream workflows management systems. The resource is listed on FAIRsharing[21], and is available for public reuse at https://github.com/mpi2/impc-mousephenotype-api
APPID 1209 - Developing a federated XNAT portal for	To elevate data findability and interoperability, metadata descriptions for preclinical imaging studies (study design,

[14] https://hub.docker.com/repository/docker/ilveroluca/raw2ometiff
[15] https://hub.docker.com/repository/docker/ilveroluca/bioformats2raw
[16] https://hub.docker.com/repository/docker/ilveroluca/crypt4gh
[17] 10.25504/FAIRsharing.N4a3Pj
[18] https://github.com/crs4/crc_cohort_modelling
[19] https://workflowhub.eu/workflows/265?version=1
[20] https://app.lifemonitor.eu/workflow;uuid=cf3aa710-61b9-013a-5604-005056ab5db4
[21] https://fairsharing.org/FAIRsharing.LdoU1Ia

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

medical image datasets	animal species, chemicals, disease model, imaging instrumentation, etc.) were formulated, leading to the creation of new data-tables inside the XNAT core database. The functionality was further extended by a new graphical UI for adding metadata information when creating new project within XNAT. The current version of the XNAT instance has been containerised with Docker. In the next stage of the project, new APIs will the link the frontend and backend for storing metadata information provided by users and also allow queries using metadata terms.
APPID 1237 - Cloudification of CryoEM data and metadata	The team established workflows for real-time submission of the raw and processed data from electron microscope to the cloud. During system evaluation, local computational resources were used to test the data transfer and archive functionality of different cloud providers. Following provider selection, a web browser based frontend application was developed to submit data to the OneData cloud solution directly from instruments. A new backend fs2od application provides for real-time update of the data to the cloud and data management (removal, archival, publication after the embargo period). Documentation covering cloud service installation and setup was summarised in a user manual[22]. Frontend application for data submission is here: https://aperture.ceitec.muni.cz/Shibboleth.sso/Login (requires authentication). Data are currently harvested into the internal (non-public) catalogue and will be made publicly available after publication or expiration of the embargo period (3 years). A tool for automated submission of the data to the public repository (EMPIAR) will be implemented in the final stages of the project as well as workflows adoption at additional electron microscopy laboratories. Besides, progress has been made to export the image processing workflow in CWL using a CryoEM ontology (https://doi.org/10.25504/FAIRsharing.q47l0t), also existing workflow viewer in EMPIAR has been enriched to provide users with quality analysis information.

Table 1: Summary of the achievements so far and final objectives of the WP1 Open Call Projects

The WP3 Open Call received more applications than the WP1 Open Calls (as eligibility for non EOSC-Life associated institutions was wider), 43 applications required input from WP1 data experts in proposal maturation and advice. For the 8 accepted projects[23], this resulted in

[22] https://cryo-em-docs.readthedocs.io/en/latest/user/download_all.html
[23] https://docs.google.com/spreadsheets/d/1qy4uq8GR81WlZNjpVHKaQq8doM6TRCU9kHlz5-g2bFQ/edit#gid=0

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

improved proposal quality and a deeper understanding of the services landscape for EOSC-Life. This is translated into impact in delivery of improved services delivered by EOSC-Life and greater breadth of the projects using services. Projects commenced work in September 2021 with an orientation meeting outlining the support available from the technical WPs', including WP1. Each project was assigned individual contacts from WP1 who could provide guidance on services and resources.

The FAIR Hackathon Background Session for Open call projects was held in December2021. Presentations covered introductory concepts and tools for FAIR data management which may be transferable for use by the project teams in executing their projects

Talk 1: FAIR data beginners guide[24]: This presentation focussed on communicating basic FAIR data concepts illustrated by practical examples of working with metadata from the Marine Biology and Bioimaging domains

Talk 2: FAIRification resources [25]: The use of data FAIRification resources developed by the IMI FAIRplus consortium, who are also represented in EOSC-Life WP1 was outlined, in particular FAIRplus cookbook

Talk 3: Ontologies versus Vocabularies - which is better?[26]: Examination of key features of ontologies and vocabularies

Talk 4: On the road to FAIRness…[27]: Experiences of cloud deployment of FAIR data resources from the Euro-BioImaging domain

2.2. Implementation Findability, Accessibility - FAIRsharing.org

The relevant FAIR principles that FAIR services should meet are:

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

In delivering supporting services, two considerations are important, EOSC-Life is a large collection of research infrastructures with a complex portfolio of services, second that EOSC-Life itself is resident in the EOSC multi-disciplinary landscape. We have therefore chosen to aggregate information across EOSC-Life and FAIRsharing.org, which is part of WP6, and was selected as the project's registry of resources and standards. This is described in detail in D1.2, is publicly available[28] and an EOSC-Life collection of resources was previously created in FAIRsharing. Here we describe updates to the previous deliverable as FAIRsharing has provided new services.

[24] https://docs.google.com/presentation/d/1C864JwcEc8hyQhXJV9XbbqOZ6CeOQGHls1BBY4jN5nA/edit#slide=id.p1
[25] https://docs.google.com/presentation/d/1uya9oRlm4MkL8T9sk64HZWKa2Aa63aRSrRlVzn_2WWI/edit#slide=id.p1
[26] https://docs.google.com/presentation/d/1z1FXOoFKBjnYOEW8BzvULpTit1kD0tfgssvz9uwzVll/edit
[27] https://docs.google.com/presentation/d/1fYpNix7LD8FRCIO9SGyRDXis_Ux3mImo/edit#slide=id.p1
[28] https://fairsharing.org/

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

In Jan 2022 a new version of FAIRsharing was released and in parallel its EOSC-Life collection[29] was extended with each RI partner requested by WP1 to review, extend and update their information in FAIRsharing. This information was curated by the Oxford partners, who run FAIRsharing and other FAIR-related tasks in WP6, improving FAIRsharing's representation of this collection of over one hundred resources covering thousands of datasets.

The impact of this work is to enrich the EOSC registries by providing up-to-date resource-level metadata in the public domain ensuring that the EOSC-Life datasets are Findable from FAIRsharing and Accessible from the linked data repositories.

New RI-centric pages for 11 of the 13 RIs in EOSC-Life have been created (e.g. ELIXIR[30] and MIRRI[31]). These pages enable attribution at the level of the RIs and the responsible individuals themselves. Each RI page in FAIRsharing showcases its associated repositories, standards and collections and identifies and attributes (via ORCID) related users.

The result is a landscape of interconnected users, RIs, databases and standards that greatly impacts upon the Findability of EOSC-Life resources. DOIs are provided for each resource, ensuring the metadata entered within FAIRsharing is itself identifiable. The outcome of this work is that all Research Infrastructures have their own landing page in FAIRsharing that links each RI to richly-described and uniquely-identified resource metadata records.

These records are just a part of the wider FAIRsharing registry, which is computationally accessible to a wide variety of third party tools and which is interoperable with pan-EOSC activities via the FAIRsharing collaboration with OpenAIRE.

The EOSC-Life Collection is reached via a specific website at fairsharing.org[32] (Figure 1) and supports the addition of standards, vocabularies and FAIR criteria enabling RIs to package and brand their records. An RI-level example is shown in Figure 2. There are now 132 resources available in the EOSC-Life collection; their metadata is more complete and the diversity of the resources has been extended as more than 20 new resources have been added for this release. As this is a live view of the EOSC-Life resource ecosystem, a variety of stages of the resource life cycle are displayed. In-development, Ready and Deprecated resources are marked as such allowing users to assess their relevance.

[29] https://fairsharing.org/EOSCLife
[30] https://beta.fairsharing.org/organisations/839
[31] https://beta.fairsharing.org/organisations/3280
[32] https://fairsharing.org/EOSCLife

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

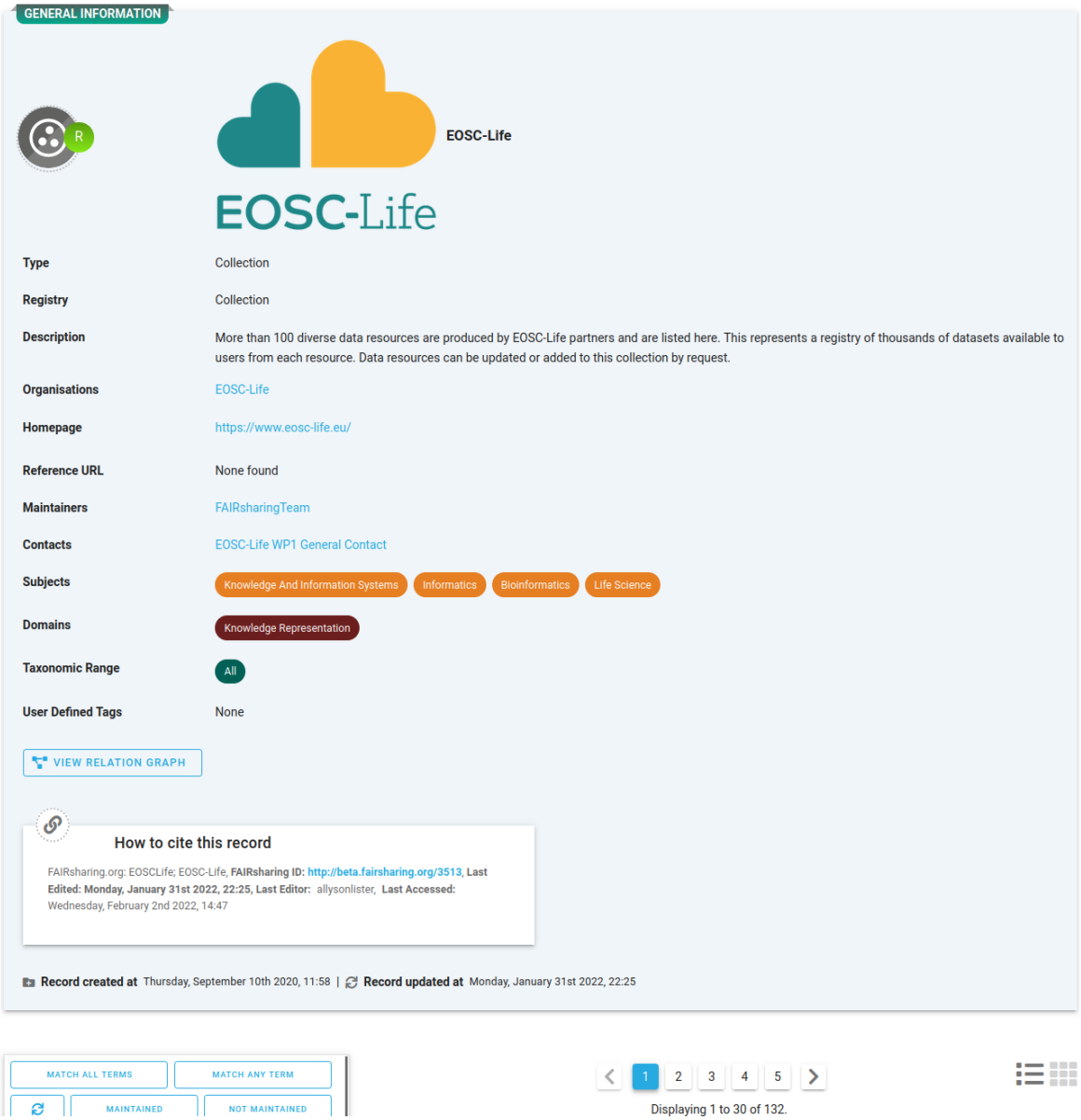

Figure 1: The EOSC-Life collection in FAIRSharing record[33]

[33] https://fairsharing.org/3513

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

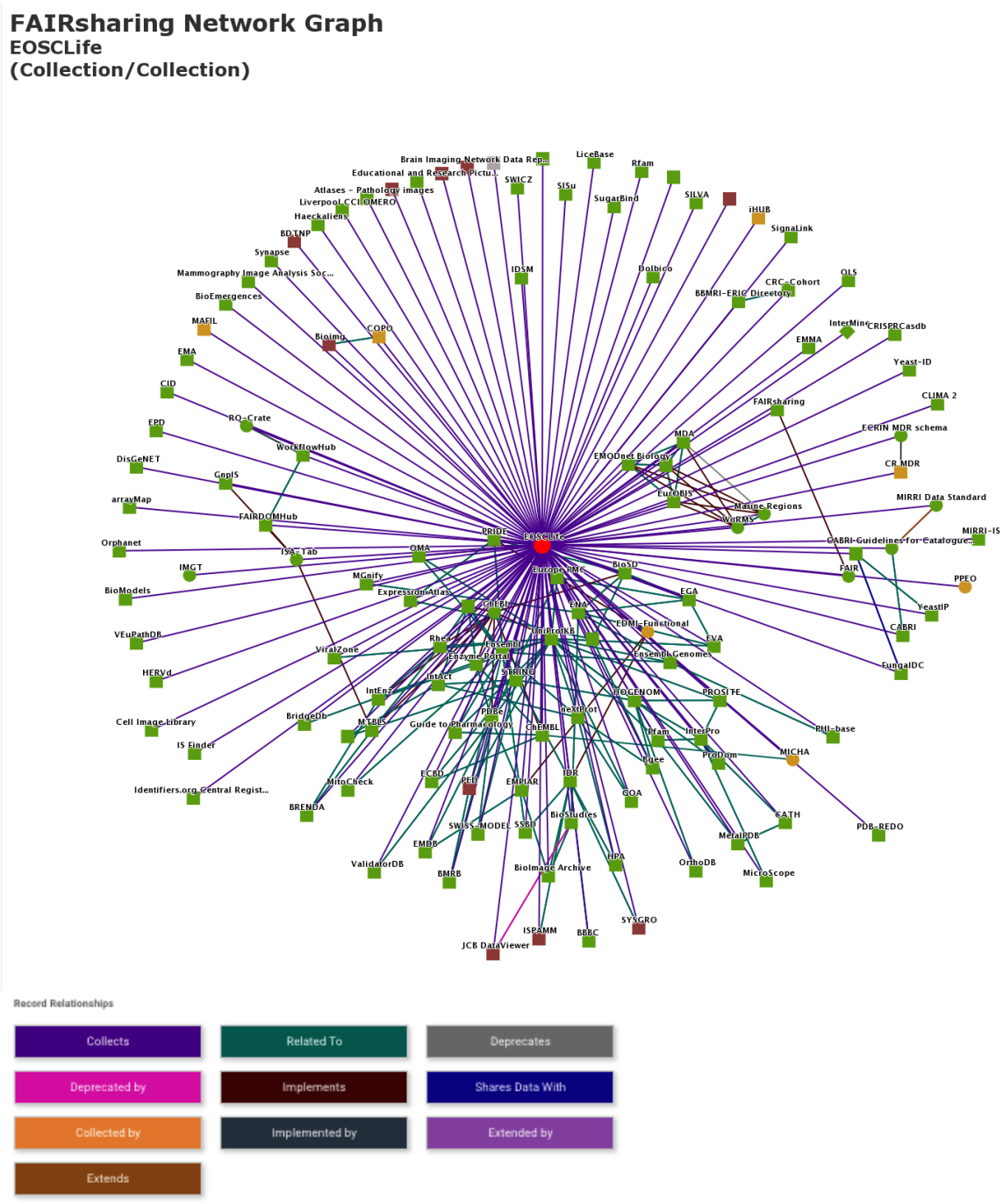

Figure 2: FAIRsharing's graph representation of the EOSC-Life Collection with key showing different types of records[34]

[34] https://fairsharing.org/graph/3513

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

2.3. FAIR Interoperability services

FAIR Interoperability services are required to achieve the following FAIR principles:

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles

I3. (Meta)data include qualified references to other (meta)data

These services are realised in the delivery of a suite of semantic services which are deployed using cloud principles and practice exemplified by EOSC-Life. Sustainability is achieved by extending and evolving pre-existing services and developing new services where these extend the utility for multiple users closely aligned to user needs. Here we describe the work we have completed on FAIR interoperability services contributing to this deliverable

Cloud Deployable Services

Cloud deployable interoperability services are required as data resources become cloud deployable, and also as the pattern of usage has changed. In the past widely accessible services were designed and deployed at a central location. As cloud deployment has become accessible to all at lower cost, and as some datasets/data resources are deployed in industrial or secure settings, then interoperability services must also be portable. There are also considerable benefits for resource developers, for example, portable resources are easier to develop with multisite teams, deployment is often simpler. A key consideration for resource developers in preferring their own cloud deployment is the ability to use their own branding and to manage their own content and content release cycles. For this reason, we have delivered a portable semantic services stack. In doing so we have re-developed previous services to allow us and others to deploy these on public and private clouds:

- Ontology Lookup Service - an ontology discovery and access service containing 273 biomedical ontologies, 7,165,024 terms, 44,851 properties and 493,523 individuals. OLS has a web interface enabling search and visualisation of ontologies and a highly accessed API. More than 35 ontologies are newly available in OLS during EOSC-Life representing user needs during the pandemic and the state of the art of biomedical sciences data. Ontologies are both continually updated and new ontologies appear, having matured to the point that they are needed by users. Existing ontologies are also updated and OLS automatically updates these as this happens, ensuring that users are working with the most recent versions. Content metrics are presented in Table 2 below and new ontologies added during EOSC-Life are listed in Table 3 below.
- Ontology Cross Reference Service - a cross ontology mapping service allowing users to integrate datasets annotated with different ontology terms, a common use case when working data integration across datasets/projects/biomedical domains. Ontologies in OLS are used to deliver cross references in OxO and so cross references from new ontologies are available for use in OxO. OxO presents a convenient way to access cross references.
- Zooma - an ontology to meta data mapping application allowing ontologies to be applied to text, aimed at biomedical curators and data annotators and suggesting terms for the

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

annotation of new datasets based on existing knowledge. By improving the ontology content of OLS we provide a richer term set for curators to use.

Total Ontologies in OLS, 31 January 2022	273
Ontologies last updated in 2022	40
Ontologies last updated in 2021	130
Ontologies last updated in 2020	71
Ontologies last updated in 2019	17
Open issues	159
Software releases	71
OLS usage 2021	4,260,246,514
OxO releases	8
OxO usage	77,400,064
Zooma releases	14
Zooma usage	116,275,605

Table 2: OLS Content and update metrics, accessed 31 January 2022.

Ontology	URL	Domain
REPRODUCE-ME	https://www.ebi.ac.uk/ols/ontologies/reproduceme	Provenance ontology, extending PROV-O and P-Plan ontologie
Cryo Electron Microscopy ontology	https://www.ebi.ac.uk/ols/ontologies/cryoem	An ontology of data types and image processing operations in Cryo Electron Microscopy for Single Particles.
CoVoc	https://www.ebi.ac.uk/ols/ontologies/covoc	The COVID-19 Vocabulary (COVoc) is an ontology containing terms related to the research of the COVID-19 pandemic. This includes host organisms, pathogenicity, gene and gene products, barrier

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

		gestures, treatments and more.
SnomedCT	https://www.ebi.ac.uk/ols/ontologies/snomed	SNOMED CT or SNOMED Clinical Terms is a systematically organised computer processable collection of medical terms providing codes, terms, synonyms and definitions used in clinical documentation and reporting.
PCL	https://www.ebi.ac.uk/ols/ontologies/pcl	Cell types that are provisionally defined by experimental techniques such as single cell transcriptomics rather than a straightforward & coherent set of properties.
LEPAO	https://www.ebi.ac.uk/ols/ontologies/lepao	The Lepidoptera Anatomy Ontology contains terms used for describing the anatomy and phenotype of moths and butterflies in biodiversity research.
EPSO	https://www.ebi.ac.uk/ols/ontologies/epso	An application driven Epilepsy Ontology with official terms from the ILAE
DISDRIV	https://www.ebi.ac.uk/ols/ontologies/disdriv	Drivers of human diseases including environmental, maternal and social exposures.
SCDO	https://www.ebi.ac.uk/ols/ontologies/scdo	An ontology for the standardisation of terminology and integration of knowledge about Sickle Cell Disease.
CDNO	https://www.ebi.ac.uk/ols/ontologies/cdno	N/A
RBO	https://www.ebi.ac.uk/ols/ontologies/rbo	RBO is an ontology for the effects of radiation on biota in terrestrial and space environments.
CLAO	https://www.ebi.ac.uk/ols/ontologies/clao	CLAO is an ontology of anatomical terms employed in morphological descriptions for the Class Collembola (Arthropoda: Hexapoda).
GSSO	https://www.ebi.ac.uk/ols/ontologies/gsso	GSSO is the Gender, Sex, and Sex Orientation ontology, including terms related to gender identity and expression, sexual and romantic identity and orientation, and sexual and reproductive behavior.
FIDEO	https://www.ebi.ac.uk/ols/ontologies/fideo	The Food Interactions with Drugs Evidence Ontology (FIDEO) represents Food-Drug Interactions and underlying interaction mechanisms described in scientific publications, drug and adverse effects

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

		databases, and drug interactions compendia. The ontology builds on previous efforts from the FoodOn, DRON, ChEBI, and DIDEO ontologies as well as the Thériaque database. This ontology is maintained at https://gitub.u-bordeaux.fr/erias/fideo, and requests for changes or additions should be filed at the issue tracker there.
PSO	https://www.ebi.ac.uk/ols/ontologies/pso	N/A
FOVT	https://www.ebi.ac.uk/ols/ontologies/fovt	N/A
CLYH	https://www.ebi.ac.uk/ols/ontologies/clyh	Anatomy, development and life cycle stages - planula, polyp, medusa/jellyfish - of the cnidarian hydrozoan species, Clytia hemisphaerica.
AMPFX	https://www.ebi.ac.uk/ols/ontologies/amphx	An ontology for the development and anatomy of Amphioxus (Branchiostoma lanceolatum).
OMO	https://www.ebi.ac.uk/ols/ontologies/omo	An ontology specifies terms that are used to annotate ontology terms for all OBO ontologies. The ontology was developed as part of Information Artifact Ontology (IAO).
CIDO	https://www.ebi.ac.uk/ols/ontologies/cido	The Ontology of Coronavirus Infectious Disease (CIDO) is a community-driven open-source biomedical ontology in the area of coronavirus infectious disease. The CIDO is developed to provide standardized human- and computer-interpretable annotation and representation of various coronavirus infectious diseases, including their etiology, transmission, pathogenesis, diagnosis, prevention, and treatment.
MAXO	https://www.ebi.ac.uk/ols/ontologies/maxo	An ontology to represent medically relevant actions, procedures, therapies, interventions, and recommendations.
CHIRO	https://www.ebi.ac.uk/ols/ontologies/chiro	CHEBI provides a distinct role hierarchy. Chemicals in the structural hierarchy are connected via a 'has role' relation. CHIRO provides links from these roles to useful other classes in other ontologies. This will allow direct connection between chemical structures

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

		(small molecules, drugs) and what they do. This could be formalized using 'capable of', in the same way Uberon and the Cell Ontology link structures to processes.
TXPO	https://www.ebi.ac.uk/ols/ontologies/txpo	Elucidating the mechanism of toxicity is crucial in drug safety evaluations. TOXic Process Ontology (TXPO) systematizes a wide variety of terms involving toxicity courses and processes. The first version of TXPO focuses on liver toxicity. The TXPO contains an is-a hierarchy that is organized into three layers: the top layer contains general terms, mostly derived from the Basic Formal Ontology. The intermediate layer contains biomedical terms in OBO foundry from UBERON, Cell Ontology, NCBI Taxon, ChEBI, Gene Ontology, PATO, OGG, INOH, HINO, NCIT, DOID and Relational ontology (RO). The lower layer contains toxicological terms. In applied work, we have developed a prototype of TOXPILOT, a TOXic Process InterpretabLe knOwledge sysTem. TOXPILOT provides visualization maps of the toxic course, which facilitates capturing the comprehensive picture for understanding toxicity mechanisms. A prototype of TOXPILOT is available: https://toxpilot.nibiohn.go.jp
ECTO	https://www.ebi.ac.uk/ols/ontologies/ecto	ECTO describes exposures to experimental treatments of plants and model organisms (e.g. exposures to modification of diet, lighting levels, temperature); exposures of humans or any other organisms to stressors through a variety of routes, for purposes of public health, environmental monitoring etc, stimuli, natural and experimental, any kind of environmental condition or change in condition that can be experienced by an organism or population of organisms on earth. The scope is very general and can include for example plant treatment regimens, as well as human clinical exposures (although these may better be handled by a more specialized ontology).
LABO	https://www.ebi.ac.uk/ols/ontologies/labo	LABO is an ontology of informational entities formalizing clinical laboratory tests prescriptions and reporting documents.

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

ORNASEQ	https://www.ebi.ac.uk/ols/ontologies/ornaseq	An application ontology designed to annotate next-generation sequencing experiments performed on RNA.
HTN	https://www.ebi.ac.uk/ols/ontologies/htn	An ontology for representing clinical data about hypertension, intended to support classification of patients according to various diagnostic guidelines
SRAO	https://www.ebi.ac.uk/ols/ontologies/srao	The FAIRsharing Subject Ontology (SRAO) is an application ontology for the categorization of research disciplines across all research domains, from the humanities to the natural sciences. It utilizes multiple external vocabularies.
HCAO	https://www.ebi.ac.uk/ols/ontologies/hcao	Application ontology for human cell types, anatomy and development stages for the Human Cell Atlas.
PLANA	https://www.ebi.ac.uk/ols/ontologies/plana	PLANA, the PLANarian Anatomy Ontology, encompasses the anatomy of developmental stages and adult biotypes of *Schmidtea mediterranea*.
IDO-COVID-19	https://www.ebi.ac.uk/ols/ontologies/idocovid19	The COVID-19 Infectious Disease Ontology (IDO-COVID-19) is an extension of the Infectious Disease Ontology (IDO) and the Virus Infectious Disease Ontology (VIDO). IDO-COVID-19 follows OBO Foundry guidelines, employs the Basic Formal Ontology as its starting point, and covers epidemiology, classification, pathogenesis, and treatment of terms used to represent infection by the SARS-CoV-2 virus strain, and the associated COVID-19 disease.
VIDO	https://www.ebi.ac.uk/ols/ontologies/vido	The Virus Infectious Disease Ontology (IDO Virus) is an extension of the Infectious Disease Ontology (IDO). IDO Virus follows OBO Foundry guidelines, employs the Basic Formal Ontology as its starting point, and covers epidemiology, classification, pathogenesis, and treatment of terms used by Virologists, i.e. virus, prion, satellite, viroid, etc.
BCIO	https://www.ebi.ac.uk/ols/ontologies/bcio	The Behaviour Change Intervention Ontology is an ontology for all aspects of human behaviour change interventions and their evaluation.
COVOC	https://www.ebi.ac.uk/ols/ontologies/covoc	The COVID-19 Vocabulary (COVoc) is an ontology containing terms related to the research of the

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

		COVID-19 pandemic. This includes host organisms, pathogenicity, gene and gene products, barrier gestures, treatments and more.
OM	https://www.ebi.ac.uk/ols/ontologies/om	The OM ontology provides classes, instances, and properties that represent the different concepts used for defining and using measures and units. It includes, for instance, common units such as the SI units meter and kilogram, but also units from other systems of units such as the mile or nautical mile. For many application areas it includes more specific units and quantities, such as the unit of the Hubble constant: km/s/Mpc, or the quantity vaselife. OM defines the complete set of concepts in the domain as distinguished in the textual standards. As a result the ontology can answer a wider range of competency questions than the existing approaches do. The following application areas are supported by OM: Geometry; Mechanics; Thermodynamics; Electromagnetism; Fluid mechanics; Chemical physics; Photometry; Radiometry and Radiobiology; Nuclear physics; Astronomy and Astrophysics; Cosmology; Earth science; Meteorology; Material science; Microbiology; Economics; Information technology; Typography; Shipping; Food engineering; Post-harvest; technology; Dynamics of texture and taste; Packaging
ORTH	https://www.ebi.ac.uk/ols/ontologies/orth	The need of a common ontology for describing orthology information in biological research communities has led to the creation of the Orthology Ontology (ORTH). ORTH ontology is designed to describe sequence homology data available in multiple orthology databases on the Web (e.g.: OMA, OrthoDB, HieranoiDB, and etc.). By sequence homology data, we mostly mean gene region, gene and protein centric orthology, paralogy, and xenology information. Depending on the database, the homology information is structured in different ways. ORTH ontology accommodates these disparate data structures namely Hierarchical Orthologous Group (HOG), cluster of homologous sequences and homologous-pairwise relations between sequences. In addition to the specific ORTH terms, this

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

		specification includes terms of the imported ontologies (e.g. Semantic science Integrated Ontology, SIO) which are pertinent to represent the information from various orthology databases in a homogeneous way.

Table 3: Novel Ontology content in OLS, accessed 31 January 2022

Major Features added to OLS

OLS has had 71 software releases during the development of this deliverable. Major enhancements included in these releases are the following.

1. An ontology visualisation feature: The ability to define preferred root terms for an ontology that may be different from the upper ontology terms from which the ontology is extended. This enhancement was presented at ICBO 2019 [35]. This allows users to see terms meaningful to them and to avoid seeing terms used to organise the ontology's structure. All terms are still accessible via the API, however, only terms meaningful to the majority of users are visible in the GUI, a significant usability change impacting the majority of web site users.
2. The OLS UI has been redesigned to increase usability. we redesigned 4,260,246,514As an example in the previous UI we used the comma character (,) to separate synonyms (see Figure HH1). This was problematic where a synonym has an embedded comma which several ontologies do. For this reason the UI has been redesigned to make the visual separation between different synonyms clear (see Figure HH2).
3. Ontology creators can choose the granularity of their modelling, for example, an ontology of species may use classes to model to species and may create a strain of a species C57BL6 (mouse strain) as instances rather than classes. In this case for query users may need to query classes and instances to obtain meaningful results which OLS did not support. In particular there was no clear visual distinction between a term and an individual in the tree when visualising the ontology as can be seen in Figure 2. In Figure 3 we therefore demonstrate how we have updated the visualisation of individuals to make the distinction between terms and individuals clearer to users This affects the around 490 000 individuals in OLS and aids users in selecting terms for annotation and also in comparing how ontologies are modelled, useful when integrating annotations to different ontologies for the purpose of interoperability.

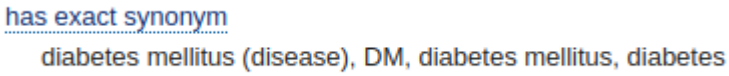

Figure HH1: How synonyms used to be represented in the old version of OLS

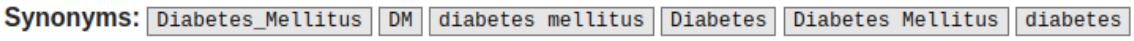

Figure HH2: How synonyms are represented in the redesigned UI of OLS

[35] http://ceur-ws.org/Vol-2931/ICBO_2019_paper_44.pdf

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

Figure HH3: Individuals are indistinguishable from terms.

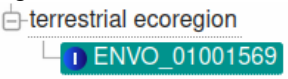

Figure HH4: Blue "I" icon added to make individuals distinguishable from terms.

New Interoperability Service Features

All three of the interoperability services, OLS, Oxo and Zooma have been dockerised and complete documentation is supplied for use of the dockerised services[36]. In order to allow users to customise their self-hosted OLS dockerised instances new features have been added supporting branding, issue tracking, user support, Figure 5 below provides detail of new features. We have also invited users downloading our docker instances to report on their usage, for example: the International 100K Cohorts Initiative which integrates international human cohort data, the Human Cell Atlas, which integrates single cell data and the Monarch Initiative and ontology integration project. We are aware of other downloads identified via our issue tracker; at least three instances of OLS are running in pharma/agrifood companies, but they have chosen not to register these. In total the docker images have been downloaded 3300 times from Docker Hub.

OLS

- `ols.customisation.debrand` — If set to true, removes the EBI header and footer, documentation, and about page
- `ols.customisation.title` — A custom title for your instance, e.g. "My OLS Instance"
- `ols.customisation.short-title` — A shorter version of the custom title, e.g. "MYOLS"
- `ols.customisation.description` — A description of the instance
- `ols.customisation.org` — The organisation hosting your instance
- `ols.customisation.web` — Url of the website for your organization.
- `ols.customisation.twitter` — Handle to the Twitter account of your organisation.
- `ols.customisation.issuesPage` — Url for the issue tracker for your organisation.
- `ols.customisation.supportMail` — Email address where people can contact you.
- `ols.customisation.hideGraphView` — Set to true to hide the graph view
- `ols.customisation.errorMessage` — Message to show on error pages
- `ols.customisation.ontologyAlias` — A custom word or phrase to use instead of "Ontology", e.g. "Data Dictionary"
- `ols.customisation.ontologyAliasPlural` — As `ontologyAlias` but plural, e.g. "Data Dictionaries"
- `ols.customisation.oxoUrl` — The URL of an OxO instance to link to with a trailing slash e.g. `https://www.ebi.ac.uk/spot/oxo/

OxO

- `oxo.customisation.debrand` — If set to true, removes the EBI header and footer, documentation, and about page
- `oxo.customisation.title` — A custom title for your instance, e.g. "My OxO Instance"
- `oxo.customisation.short-title` — A shorter version of the custom title, e.g. "MYOxO"
- `oxo.customisation.description` — A description of the instance
- `oxo.customisation.org` — The organisation hosting your instance

Figure 3: FAIR interoperability services dockerised instance configuration options[37].

[36] https://github.com/EBISPOT/ontotools-docker
[37] https://github.com/EBISPOT/ontotools-docker#customisation

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

Defining 'A FAIR Ontology'

The FAIR principles require ' (Meta)data use vocabularies that follow FAIR principles' - but the definition of a FAIR vocabulary is unclear. In collaboration with others, we have therefore analysed the FAIR vocabulary space within OLS and elsewhere and have published a paper (accepted for the SWAT4LS[38] conference proceedings but not yet available) analysing existing definitions and proposing Features of a Fair Vocabulary (or ontology). We analysed the landscape of vocabularies, refine definitions of FAIR data, FAIR data resources and FAIR vocabularies to reach satisfiable FAIR Vocabulary Features (FVFs). In our analysis we considered whether a Vocabulary is 1) FAIR in terms of its application to FAIR data 2) FAIR in the context of FAIR capable resources 3) FAIR in the context of other vocabularies. We use one of the semantic services, the Ontology Lookup Service (OLS, see above) to review ontology content. We define and test a set of FVFs, compare these to other semantic and community standards and consider how they can be used to evaluate ontologies and how the resulting ontologies can be improved using the FVFs. Three ontologies were evaluated for adherence to the FVFs, the Gene Ontology (GO), International Classification of Diseases 11 (ICD11) and the Experimental Factor Ontology (EFO). These were selected as they are highly used and EFO is compositional, meaning it imports terms from many other ontologies. Three levels of compliance were determined for each FVF: full, partial and no compliance and scores across all 11 FVF were calculated. The FVFs can be used in future to calculate scores for ontologies with the impact of improving their FAIRness in future, for example as an indicator deployed on the Ontology Lookup Service, thus allowing users to make an informed selection in ensuring that data are FAIR. This work is also relevant to FAIR evaluation metrics and tools under development by EOSC-Life's WP6 as it extends the concepts to ontologies as well as data.

2.4. Re-use Services

Public data archives are an essential component of biological research. Publishing data and metadata can be very challenging for multiple reasons e.g. cost, storage. Added-value image data archives like EMPIAR and IDR apply the FAIR principles listed below and provide essential resources for building scientific workflows therefore enabling data reuse.

The relevant FAIR principles are:

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license.

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

Combining public imaging added-values archives, ontology associated resources and data warehouse systems has enabled us to provide meaningful scientific workflows demonstrating real world reuse of data embedded in scientific practice, for example:

[38] http://www.swat4ls.org/

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

- Explore in an interactive manner analytical results associated with a scientific paper e.g. https://workflowhub.eu/workflows/238 (based on IDR[39])
- Allow investigation amongst multiple studies in an added-values archives (based on IDR)
- Search for imaging resources linked to annotations expressed as ontological terms (based on IDR and OXO)
- Answer scientific questions e.g. Diabetes related genes expressed in Pancreas, see https://workflowhub.eu/workflows/242, based on IDR and HumanMine[40].

We used the API of each resource to engage with data analysts/developers. Examples used different programming languages e.g. Python, R. Recent training sessions have highlighted the needs for developers focused materials.

We did not only provide the workflows but also an environment with all the dependencies required so that the examples can be run locally or in existing cloud-based resources e.g. Mybinder[41], Google Colab[42], Galaxy[43] allowing users to simply and quickly reuse data in the location of their choice.

Several of these example workflows have been published or, in the process of being published to the WorkflowHub[44], a public registry, developed by colleagues from WP2, which facilitates discovery and re-use. Using such a registry means that workflows like the associated data and metadata are given a Digital Object Identifier (DOI), license file etc. increasing the FAIRness of the scientific research.

We plan to add more examples including projects that recently joined the EOSC ecosystem.

3. Conclusions

As we have described we have enhanced the capacity for FAIR services in terms of increasing the number of FAIR experts and disseminating their work in engaging during EOSC-Life's calls, working collaboratively to deliver new datasets, software and services. We have delivered FAIR services addressing all components of F.A.I.R and have developed FAIR services in a sustainable manner by collaborating across the biomedical community to integrate services into the FAIRsharing registry. We have delivered portable interoperability services and used the portable services in our own development practice. We have addressed FAIR at scale both in terms of the users we have reached through our services and the improvements we have made to services.

[39] https://idr.openmicroscopy.org/
[40] https://www.humanmine.org/
[41] https://mybinder.org/
[42] https://colab.research.google.com/
[43] https://galaxyproject.org/
[44] https://workflowhub.eu/

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

4. Next Steps

In future work we will continue to refine the services described in this deliverable based on the needs from the WP3 two remaining calls. We note that as these have industrial and sensitive data themes that these are likely to have additional needs for portability. In addition, we will seek feedback from the wider user community, including EOSC association working groups on emerging needs with the aim of sustaining future usage and development.

We also have planned developments in progress, for example, we will add support for multilingual ontologies to the interoperability services so that when an ontology supports multiple languages, there will be a dropdown to allow users to choose a language. In the case where not all labels in the ontology are available in the chosen language, labels will still be displayed in English. Besides being able to view multilingual ontologies in their different languages, this update will also enable searches in different languages necessary for some health datasets and will support lay users, and users in resource poor settings, who are often not working in English.

4.1. Dissemination

Dissemination for WP1 takes two forms, both intra- and extra- project activities. Given the scale of EOSC-Life and the outward facing activities related to Open Calls we have mainly focussed on intra project activities. These include:

Hackathons:

FAIR Hackathon WP1 and WP3 EOSC-Life demonstrators, Berlin, 2020. This allowed interaction between WP3 demonstrators and provided the first set of use cases for WP1 services.

FAIR Hackathon WP1 and WP3 EOSC-Life projects, a future 7 Feb 2022 is being prepared, questionnaires have been circulated (Appendix 1 below) and we will both demonstrate and recruit new features which will be documented as future user stories.

WP3 Requested presentations/QA sessions on WP1 services for WP3 Open Calls

A Conference paper was accepted for the Fair Vocabulary Features Paper at the Semantic Applications and Tools for Health Care and Life Sciences (SWAT4HCLS[45]) workshop, Leiden, The Netherlands, January 10-13, 2022, an oral presentation was made and the conference paper will be submitted as a full paper for publication in the Journal of Biomedical Semantics (in preparation).

OLS and OxO were presented in an invited industrial tutorial to an agrifood company demonstrating their value for industrial users.

[45] http://www.swat4ls.org/

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

Abbreviations

Abbreviation	Full name
API	Definition
DCAT2	Application Programming Interface
DDI	Data Catalogue Vocabulary version 2
DUO	Cross Domain Integration (DDI-CDI), a specification aimed at helping implementers integrate data across domain and institutional boundaries
EMPIAR	Data Use Ontology - a Global Alliance for Genomics and Health standard
EOSC	Electron Microscopy Public Image Archive
FAIR	European Open Science Cloud
FDP	Findable, Accessible, Interoperable, Re-usable
IDR	FAIR data point
IPD	Image Data Resource
RI	Individual participant data
OC	Research infrastructures within EOSC-Life

Delivery and Schedule

The delivery is delayed from August 2021 to February 2022 due to the Covid-19 pandemic which prevented in person meetings and due to staff being assigned to Covid-19 projects or on sick leave therefore temporarily delaying work on this deliverable. A Contract amendment was made rescheduling delivery in February 2022 which has been met.

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

Appendices

Appendix 1.

Funded projects from WP3 Open Calls with summary information from the WP1 perspective used to inform our analysis of tools requirement

Call	**Project Title**	**WP1 Scope Analysis Summary**
WP3 First, open digital data	Integrating EU-RI datasets for preclinical and discovery research bioimaging	Images, animal models (mice), drug screening chemistry, data integration
	A workflow for marine Genomic Observatories data analysis	Marine species, data readiness for analysis, species taxonomies
	PDB-REDO-cloud: A flexible and scalable engine for computational structural biology	Protein structural data analyses
	Expression Atlas' RNA-Seq and Microarray analysis pipelines migration to workflow environments for cloud deployment and reproducibility	Multispecies, multi variable gene expression data cloud based workflows
	Increasing the FAIRness of Phytolith Data (plant silica bodies)	FAIRness analysis for existing datasets, standardisation, repository development
	Reference Data Resource - RefGenie, promoting the re-use of reference genome data	Genomics data reproducibility improvements using provenance models
	Open Source Secure Data Infrastructure and Processes for Life Sciences (OSSDIP4LIFE)	Sensitive data analysis platform extended for life sciences exposing data subsets preserving privacy

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

	Towards FAIR data for X ray-based structure-guided drug design	Tools for automated harvesting, validation, and deposition of such Macromolecular Crystallography data and metadata which should considerably promote re-use and interoperability
WP1	PombeMine, a FAIR enabled workflow in the cloud	Genomic and functional information
	OmicsDI cloud	Genomic and functional genomics data
	Macromolecular crystallography data management in the cloud	Protein structures
	PDB-REDO - FAIR protein structures with deep versioning and cloud portability for scientific reproducibility and data provenance tracking	Protein structures
	Cloudification of the IMPC dataset for workflow application	Phenotypic and image data packaged for workflow access.
	Developing a federated XNAT portal for medical image datasets	Image data access
	Cloudification of CryoEM data and metadata	Image data access
	Cloudification of the BBMRI-ERIC Colorectal cancer cohort data digital pathology imaging data	Disease and image data access

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

Appendix 2.

FAIR self-assessment service session, FAIR Hackathon, December 2021

	Self-Assessment Questions (Self-Check)
Session 1	**Before starting and during each step of your work** 1. Can I describe the stages of my data life-cycle e.g. in a bulleted list? 2. Do I know what software, protocols, instruments I will be using at each stage of that life-cycle, 3. …can I make a list of what provenance information I need to record for each of these softwares, protocols, and instruments? 4. Can I provide that information to external users, or are they only available on a piece of paper on my desk? **When you are ready to publish your data** 1. Have I gathered the provenance information at a fine enough granular level? a. → have I decided what granular level I need? b. → could someone recreate my results from my raw data using this provenance information? 2. Have I gathered that information together with the data they describe? a. → am I ready to submit this all to my data catalogue/archive? 3. How much provenance information can I provide as metadata in the data catalogue where I will put my data, … 4. …do I need to provide provenance information as files and documents to be included with the data in the catalogue? Do I know how I will do that?

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

Session 2	**Have I made my data Findable?** 1. Have I published my data on a catalogue or data portal? 2. Have I described my data with lots of useful metadata taken from controlled vocabularies? 3. Have I linked my scientific publication, my data paper, and my data in the catalogue? 4. Have I provided both raw and processed data? **Have I made my data Interoperable?** 1. Am I using open access data file formats? 2. Can my data be accessed programmatically? Is my data formatted to allow that? 3. Am I using controlled vocabularies to describe my parameters, locations, terms, species, instruments... **Have I made my data Accessible?** 1. Have I chosen the best catalogue or portal for my type of data? 2. Will I remember to update the record when details change (especially contact details; and dataset updates or addenda)? **Have I made my data Reusable?** 1. Did I make it clear what the usage licence is? 2. Can I make my data open access? 3. Is the provenance for each stage of my data life cycle available? 4. Can someone else take my raw data and produce the same results using the methods I have documented in my provenance?

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