

Sustainability of FAIR Life Science Resources and Projects: Lessons Learned from EOSC-Life Research Infrastructures

Authors: Romain David¹, Jean-Marie Burel², Jean-Karim Heriche³, Arina Rybina⁴, Pauline Audergon⁵, Jan-Willem Boiten⁶, Frederik Coppens⁷, Sara Crockett⁸, Katrina Exter⁹, Sven Fahrner¹⁰, Maddalena Fratelli¹¹, Carole Goble¹², Philipp Gormanns¹³, Tobias Grantner¹⁴, Björn Grüning¹⁵, Kim Tamara Gurwitz¹⁶, John Hancock¹⁷, Henriette Harmse¹⁸, Petr Holub¹⁹, Nick Juty²⁰, Geoffrey Karnbach²¹, Emma Karoune²², Antje Keppler²³, Jessica Klemeier²⁴, Carla Lancelotti²⁵, Jean-Luc Legras²⁶, Allyson L. Lister²⁷, Dario Livio Longo²⁸, Rebecca Ludwig²⁹, Bénédicte Madon³⁰, Marzia Massimi³¹, Vera Matser³², Rafaele Matteoni³³, Michaela Th. Mayrhofer³⁴, Christian Ohmann³⁵, Maria Panagiotopoulou³⁶, Helen Parkinson³⁷, Isabelle Perseil³⁸, Claudia Pfander³⁹, Roland Pieruschka⁴⁰, Michael Raess⁴¹, Andreas Rauber⁴², Audrey S Richard⁴³, Paolo Romano⁴⁴, Antonio Rosato⁴⁵, Alex Sánchez-Pla⁴⁶, Susanna-Assunta Sansone⁴⁷, Ugis Sarkans⁴⁸, Beatriz Serrano-Solano⁴⁹, Jing Tang⁵⁰, Ziaurrehman Tanoli⁵¹, Jonathan Tedds⁵², Harald Wagener⁵³, Martin Weise⁵⁴, Hans V. Westerhoff⁵⁵, Rudolf Wittner⁵⁶, Jonathan Ewbank⁵⁷, Niklas Blomberg⁵⁸, Philip Gribbon⁵⁹.

Contact author: Romain David romain.david@erinha.eu +0033 769104341

1. Dr Romain David, European Research Infrastructure on Highly Pathogenic Agents (ERINHA), Belgium, romain.david@erinha.eu, 0000-0003-4073-7456
2. Dr Jean-Marie Burel, University of Dundee, Dundee, GB, j.burel@dundee.ac.uk, 0000-0001-6867-9425
3. Dr Jean-Karim Hériché, European Molecular Biology Laboratory: Heidelberg, Baden-Württemberg, DE, heriche@embl.de, 0000-0001-6867-9425
4. Dr Arina Rybina, Euro-BioImaging ERIC Bio-Hub, European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, Germany, arina.rybina@eurobioimaging.eu, 0000-0002-5609-5710
5. Dr Pauline Audergon, Instruct-ERIC, Oxford House, Parkway Court, John Smith Drive, Oxford, OX4 2JY, UK, pauline@instruct-eric.org, 0000-0002-1624-7388
6. Dr Jan-Willem Boiten, European Advanced Translational Research Infrastructure (EATRIS)/Lygature, 3521 AL, Utrecht, The Netherlands, janwillem.boiten@lygature.org, 0000-0003-0327-638X
7. Dr Frederik Coppens, VIB Data Core, VIB Technologies, VIB, Ghent, Belgium, frederik.coppens@vib.be, 0000-0001-6565-5145
8. Dr. Sara Crockett, BBMRI-ERIC, Biobanking and BioMolecular Resources Research Infrastructure (BBMRI-ERIC), Neue Stiftung Talstrasse 2, 8010 Graz, Austria, sara.crockett@bbmri-eric.eu, 0009-0008-7983-0360
9. Dr Katrina Exter, Flanders Marine Institute, Ostend, BE, katrina.exter@vliz.be, 0000-0002-5911-1536
10. Dr Sven Fahrner, IBG-2: Plant Sciences, Forschungszentrum Jülich, 52428 Jülich, Germany, s.fahrner@fz-juelich.de, 0000-0001-6240-9678
11. Dr Maddalena Fratelli, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milano, Italy, maddalena.fratelli@marionegri.it, 0000-0002-1769-3427
12. Pr Carole Goble, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK, carole.goble@manchester.ac.uk, 0000-0003-1219-2137
13. Mr Philipp Gormanns, INFRAFRONTIER, Ingolstädter Landstraße 1, 85764 Oberschleißheim, philipp.gormanns@infrafrontier.eu, 0000-0001-9823-1621
14. Mr Tobias Grantner, TU Wien, Research Unit Data Science, Vienna, AT, tobias.grantner@tuwien.ac.at, 0000-0001-5104-5004
15. Dr Björn Grüning, Albert-Ludwigs-Universität Freiburg: Freiburg, Baden-Württemberg, DE, bjorn.gruening@gmail.com, 0000-0002-3079-6586
16. Ms Kim Tamara Gurwitz, European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom, kgurwitz@ebi.ac.uk, 0000-0003-1992-5073
17. Pr John Hancock, University of Ljubljana, Slovenia, jmhancock.science@gmail.com, 0000-0003-2991-2217
18. Dr Henriette Harmse, European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom, henriette007@ebi.ac.uk, 0000-0001-7251-9504
19. Dr Petr Holub, BBMRI-ERIC, Biobanking and BioMolecular Resources Research Infrastructure (BBMRI-ERIC), Neue Stiftingtalstrasse 2, 8010 Graz, Austria, petr.holub@bbmri-eric.eu, 0000-0002-5358-616X
20. Dr Nick Juty, Dept of Computer Science, The University of Manchester, Manchester, M13 9PL UK, nick.juty@manchester.ac.uk, 0000-0002-2036-8350

21. Mr Geoffrey Karnbach, TU Wien, Research Unit Data Science, Vienna, AT, geoffrey.karnbach@tuwien.ac.at, 0009-0005-9892-7799
22. Dr Emma Karoune, Investigative Science Team, Historic England, Fort Cumberland Road, Portsmouth, P04 9LD, UK, emma.karoune@historicengland.org.uk, 0000-0002-6576-6053
23. Dr Antje Keppler, Euro-BioImaging ERIC Bio-Hub, European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, Germany, antje.keppler@eurobioimaging.eu, 0000-0003-4358-2269
24. Ms Jessica Klemeier, European Molecular Biology Laboratory: Heidelberg, Baden-Württemberg, DE, jessica.klemeier@embl.de, 0009-0006-9083-3808
25. Dr Carla Lancelotti, Universitat Pompeu Fabra, Department of Humanities, CASEs Research Group and ICREA, Barcelona, Spain, carla.lancelotti@upf.edu, 0000-0003-1099-7329
26. Dr Jean-Luc Legras, Université de Montpellier, INRAE, Institut Agro Montpellier, Montpellier, FR, Jean-luc.legras@inrae.fr, 0000-0002-4006-4389
27. Dr. Allyson L. Lister, Oxford e-Research Centre, Department of Engineering Science, University of Oxford, 7 Keble Road, OX13QG, Oxford, UK, allyson.lister@oerc.ox.ac.uk, 0000-0002-7702-4495
28. Dr. Dario Livio Longo, Istituto di Biostrutture e Bioimmagini, Consiglio Nazionale delle Ricerche, Via Nizza 52, Torino, IT, dariolivio.longo@cnr.it, 0000-0002-6906-9925
29. Dr Rebecca Ludwig, EATRIS ERIC, European Infrastructure for Translational Medicine, Amsterdam, 1081 HZ, Netherlands, rebeccaludwig@eatris.eu, 0000-0001-9877-3287
30. Dr Bénédicte Madon, Escuela Superior de Ingenieros, Universidad de Sevilla/AICIA-Asoc.Invest.Coop.Ind.Andalucia, Camino de los Descubrimientos s/n, Sevilla-41092, Spain, bcg.madon@gmail.com, 0000-0001-8608-3895
31. Dr Marzia Massimi, Istituto di Biochimica e Biologia Cellulare, Consiglio Nazionale delle Ricerche (CNR), I-00015 Monterotondo (Rome), IT, marzia.massimi@cnr.it, 0000-0001-5052-1822
32. Dr Vera Matser, The Alan Turing Institute, British Library, 96 Euston Road, London, England, NW1 2DB, UK, vmatser@turing.ac.uk, 0000-0002-2010-6844
33. Dr Rafaele Matteoni, Istituto di Biochimica e Biologia Cellulare, Consiglio Nazionale delle Ricerche (CNR), I-00015 Monterotondo (Rome), IT, rafaele.matteoni@cnr.it, 0000-0002-0314-5948
34. Dr Michaela Th. Mayrhofer, Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-ERIC), 8010, Graz, Austria, michaela.th.mayrhofer@bbmri-eric.eu, 0000-0001-6932-0473
35. Pr Christian Ohmann, European Clinical Research Infrastructure Network (ECRIN), Kaiserswerther Str. 70, 40477 Düsseldorf, Germany, christianohmann@outlook.de, 0000-0002-5919-1003
36. Dr Maria Panagiotopoulou, European Clinical Research Infrastructure Network (ECRIN), 5-7 rue Watt 75013 Paris, France, maria.panagiotopoulou@ecrin.org, 0000-0002-4221-7254
37. Dr Helen Parkinson, European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom, parkinso@ebi.ac.uk, 0000-0003-3035-4195
38. Dr Isabelle Perseil, INSERM, 101 rue de Tolbiac, 75013 Paris, France, isabelle.perseil@inserm.fr, 0000-0001-9058-9290
39. Dr Claudia Pfander, Euro-BioImaging ERIC Bio-Hub, European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, Germany, claudia.pfander@eurobioimaging.eu, 0000-0002-9574-9553
40. Dr Roland Pieruschka, IBG-2: Plant Sciences, Forschungszentrum Jülich, 52428 Jülich, Germany, r.pieruschka@fz-juelich.de, 0000-0002-9774-2670
41. Dr Michael Raess, INFRAFRONTIER GmbH: Neuherberg / München, DE, michael.raess@infrafrontier.eu, 0000-0002-8759-1186
42. Prof Andreas Rauber, TU Wien, Research Unit Data Science, Vienna, AT, andreas.rauber@tuwien.ac.at, 0000-0002-9272-6225
43. Dr Audrey S. Richard, European Research Infrastructure on Highly Pathogenic Agents (ERINHA), 98 rue du Trône B-1050 Bruxelles, Belgium, audrey.richard@erinha.eu, 0000-0002-0207-0139
44. Dr Paolo Romano, IRCCS Ospedale Policlinico San Martino, Largo Rosanna Benzi 10, 16132 Genova, Italy, paolo.romano@hsanmartino.it, 0000-0003-4694-3883
45. Dr Antonio Rosato, Department of Chemistry and Magnetic Resonance Center, University of Florence: Sesto Fiorentino, IT, antonio.rosato@unifi.it, 0000-0001-6172-0368
46. Dr Alex Sánchez-Pla, Vall d'Hebron Institut de Recerca, Universitat de Barcelona, ES alex.sanchez@vhir.org, 0000-0002-8673-7737
47. Prof. Susanna-Assunta Sansone, Oxford e-Research Centre, Department of Engineering Science, University of Oxford, 7 Keble Road, OX13QG, Oxford, UK, sa.sansone@gmail.com, 0000-0001-5306-5690

48. Dr Ugis Sarkans, European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom, ugis@ebi.ac.uk, 0000-0001-9227-8488
49. Dr Beatriz Serrano-Solano, Euro-BioImaging ERIC Bio-Hub, European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, Germany, beatriz.serrano.solano@eurobioimaging.eu, 0000-0002-5862-6132
50. Dr. Jing Tang, University of Helsinki, Finland, jing.tang@helsinki.fi, 0000-0001-7480-7710
51. Dr Ziaurrehman Tanoli, Institute for Molecular Medicine Finland, University of Helsinki, Finland zia.rehman@helsinki.fi, 0000-0003-2435-9862
52. Dr Jonathan Tedds, ELIXIR Hub, Wellcome Genome Campus, Hinxton, CB10 1SD, UK, jonathan.tedds@elixir-europe.org, 0000-0003-2829-4584
53. Mr Harald Wagener, Group Leader Cloud and HPX, Center for Digital Health, BIH@Charité University Medicine Berlin, harald.wagener@charite.de, 0000-0003-1073-4991
54. Mr Martin Weise, TU Wien, Research Unit Data Science, Vienna, AT, martin.weise@tuwien.ac.at, 0000-0003-4216-302X
55. Mr Hans V. Westerhoff, University of Amsterdam, Free University Amsterdam, University of Manchester, NL, H.v.westerhoff@uva.nl, 0000-0002-0443-6114
56. Mr Rudolf Wittner, BBMRI-ERIC, Biobanking and BioMolecular Resources Research Infrastructure (BBMRI-ERIC), Neue Stiftingtalstrasse 2, 8010 Graz, Austria, rudolf.wittner@bbmri-eric.eu, 0000-0002-0003-2024
57. Dr Jonathan Ewbank, European Research Infrastructure on Highly Pathogenic Agents (ERINHA), 98 rue du Trône B-1050 Bruxelles, Belgium, jonathan.ewbank@erinha.eu, 0000-0002-1257-6862
58. Dr Niklas Blomberg, ELIXIR Hub, Wellcome Genome Campus, Hinxton, CB10 1SD, UK, niklas.blomberg@elixir-europe.org, 0000-0003-4155-5910
59. Dr Philip Gribbon ITMP Head Discovery Research (ITMP Drug Discovery), Germany Philip.Gribbon@itmp.fraunhofer.de, 0000-0001-7655-2459

Keywords: Research data, FAIR principles, Data curation, Sustainability, Research Projects, Open Science, Cloud, Life Sciences, EOSC

Abstract

Life Science (LS) communities must increase the sustainability of their data resources, software tools, and workflows, to allow their long-term re-use by the wider scientific community, especially in future applications involving machine-based analyses. Here, we describe key findings regarding sustainable LS resources, based on experiences from the EOSC-Life project. EOSC-Life involves 13 European LS Research Infrastructures. It lays the foundation for an open, digital and collaborative space for biological and medical research. We describe organisational, technical, financial and legal/ethical challenges that represent the main barriers to sustainability in the LS domain. Using lessons from 27 scientific projects selected via open calls, we demonstrate the efficiency of the EOSC-Life support model for sustainable FAIR data management and explore the complex sustainability needs for sensitive- and industry-related data resources. We describe cross-disciplinary sharing of best practices, and how this contributes to the sustainability of knowledge across communities, through building of training resources. We formulate a set of recommendations, focusing on working with experts, and communicating outcomes, to establish strong credibility and recognition of research projects. Furthermore, demonstrating and practising also permit a more sustainable adoption of reproducible research outputs. Recommended measures include sustainable training methodologies, as well as ensuring high quality metadata to drive the sustainable reusability of scientific objects. Responsibilities and roles to ensure sovereignty, sustainable services and tools must also be attributed. Finally, our analysis demonstrates how data harmonisation facilitates interoperability of tools, data, solutions and a better understanding of concepts, functionalities and semantics in the LSs.

Introduction

Life Science (LS) communities cover multiple scientific domains and carry out a diversity of research, from basic biological studies to applied epidemiological and environmental investigations. This breadth is evidenced by the key role played by LS communities during the COVID-19 pandemic, ranging from fundamental studies of the SARS-CoV-2 virus, the discovery of new therapies and the development of novel vaccines, to establishing and validating methods for contact tracing and wastewater surveillance. The COVID-19 pandemic demonstrated that **LS communities can provide high-quality, reliable data** for reuse by the wider scientific community (see recommendations from Research Data Alliance (RDA) COVID-19 Working Group (WG), 2020). LS communities around the world have, however, voiced the need to improve the sharing of, access to and, ultimately, reuse of data resources in a FAIR (Findable, Accessible, Interoperable and Reusable; Wilkinson et al., 2016) manner. Increased data sharing and accessibility would: i) enhance the value of results generated within a specific domain, and expand their utility by integrating data from related communities; ii) provide a vital data substrate that **supports non-hypothesis-driven** scientific discovery and advances through the application of machine-based methods; and iii) inform decision-making by policymakers (e.g. funders, politicians) and industrial partners.

To facilitate research, in addition to FAIR data, there is a need for FAIR research software (Lamprecht et al., 2020, Barker et al., 2022, Chue Hong et al., 2022). Software is used in almost all areas of science and must be sustainable to guarantee the reproducibility and re-usability of the results, data and analyses it generates. This requires long-lived, executable software, which in turn necessitates funds, typically only available through fixed-term project calls that value novelty above utility. This problem has, however, been recognised by funders (Strasser et al., 2022), and has led to the Amsterdam Declaration on Funding Research Software Sustainability that aims to change comprehensively the way funders deal with research software¹.

EOSC-Life is a European project funded under Horizon 2020. It brings together 13 LS Research Infrastructures (RIs) from the European Strategy Forum on Research Infrastructures (ESFRI) to create an open, digital and collaborative space for biological and medical research. The project publishes FAIR data and a catalogue of services provided by participating RIs that enable the management, storage and reuse of data in the European Open Science Cloud (EOSC) (Appleton et al., 2020). The project is framed by a data management plan describing the minimum requirements necessary to initiate research data sharing (Blomberg et al., 2020). The project participants, however, have recognised that merely using this tool will not ensure the long-term and large-scale use of scientific data. EOSC-Life has identified organisational, technical, financial and legal/ethical challenges that represent the main barriers to effective sustainability (definitions in Box-1).

The **organisational challenges** are associated with the non-aligned impact and reward mechanisms operating within academic organisations. Here, pursuing novelty and following new trends are more likely to result in positive funding or tenure decisions compared to investments in community-oriented collaborative initiatives addressing long-term structural

1

<https://future-of-research-software.org/draft-amsterdam-declaration-on-funding-research-software-sustainability/>

issues. The **technical challenges** are related to the need to promote the effective, widespread awareness and implementation of FAIR components, including the need to address sustainability dependencies arising from operating systems, versioning, identifiers, metadata schemata, vocabularies and provenance. These components are often inconsistently applied, hindering interoperability, and reducing opportunities for reuse of both data and software. The **financial challenges** arise from the proliferation of services and LS data repositories (between 2020-2023, for example, the Nucleic Acids Research² Database issue reported 267 new ones), and the increasing size and complexity of datasets and services. Consequently, the resources available to perform basic operations, including curation, storage, computing and access, may be insufficient. **Ethical and Legal challenges** include the need to respect intellectual property rights, and comply with the GDPR³ in a fragmented international legal landscape. How these challenges are to be overcome when a project ends is rarely addressed, leading to a reluctance to reuse resources once projects' primary objectives have been met.

Box-1. Definitions for sustainability types addressed in this paper:

- **Organisational sustainability:** (i) organisations create and support data, as well as services to maintain that support, (ii) organisations continue to support their members who are making the efforts to produce FAIR data, services, software, etc.
- **Technical sustainability:** metadata, data, software, workflows, AAI (Authentication and Authorization Infrastructure), etc., need to be future-proof and flexible enough to continue to be used even as technologies evolve.
- **Financial sustainability:** the financial resources to provide for the human resources and the storage and operational space (cloud infrastructure, supercomputing, computer storage with IT support).
- **Sustainability through continued re-use:** data, tools and software continue to be used because they can be found, accessed, operated on, and re-used, and so do not grow stale. Also, building on existing solutions i.e. established user-bases, as opposed to redundant innovation.
- **Sustainability through training:** tools and software can continue to be used because they are actively disseminated and explained through training, and where the training serves also to identify the new needs of old and new users in emerging communities and emerging scientific domains.

In this paper, we describe resources and services, as well as the associated training and knowledge exchange, which have been supported and/or created within the EOSC-Life project. We emphasise sustainability strategies formulated to ensure long-term access to, and reuse of, these resources and services in a world where open-science policies are being developed and where data and software are both shared and widely accessible. As EOSC-Life is part of the wider EOSC ecosystem, we highlight special considerations and twelve key recommendations (**[R1-R12]**) for resource sustainability. These reflect the scientific diversity of EOSC-Life's constituent communities, as well as external factors, such as the need to align with regulatory requirements related to personal health data.

The radical collaboration framework (McGovern 2018a, Pickering et al. 2021) was used to assess the sustainability of EOSC-Life outputs. We subsequently operationalised the methodology while developing this manuscript (Supplementary_Information_S1).

² <https://www.oxfordjournals.org/nar/database/c/>

³ <https://gdpr-info.eu/>

Sustainability Essentials, Tools and Challenges

PART-1: Community components: “humans and data”

A. EOSC-Life within the wider EOSC landscape

The **EOSC** (Box-2; Figure-1) is a joint initiative from the European Commission (EC), its Member States, Associate Countries and stakeholders from European research communities. It aims to **federate access to research services and data** across scientific disciplines and international borders, under a common governance structure. The EOSC-Life project exists within the wider EOSC ecosystem, creating interdependent sustainability approaches. These commonalities are explored by the EOSC Financial Sustainability, FAIR and Architecture Task Forces⁴ and encourage the use of best practices in EOSC scientific clusters.

Box-2. European Open Science cloud (EOSC) and connected communities

The ambition of the EOSC is to develop an “open multi-disciplinary environment where researchers can publish, find and re-use data, tools and services, thus enabling them to conduct their work better.”. It builds on existing infrastructure and services in a federated ‘system of systems’ approach. In order to link EOSC with efforts of the European Strategy Forum on Research Infrastructures (ESFRI) and other key European RIs, five Science Cluster projects were launched in 2019 and are structuring a large part of the EOSC research landscape (Lamanna et al., 2021):

- ENVRI-FAIR⁵ (environmental sciences)
- EOSC-Life⁶ (life sciences)
- ESCAPE⁷ (astronomy, astroparticle and particle physics)
- PaNOSC⁸ (photon and neutron sciences)
- SSHOC⁹ (social sciences and humanities)

The science cluster projects have built on long-standing interactions between the RIs in the different scientific domains and aim to improve researchers' access to data, tools, and resources, as well as FAIR data management practices. Each science cluster project addresses domain-specific requirements for linking their data resources to EOSC, but all of them also consider intra-domain interoperability and alignment.

⁴ <https://eosc.eu/eosc-task-forces>

⁵ <https://envri.eu/home-envri-fair/>

⁶ <https://www.eosc-life.eu/>

⁷ <https://projectescape.eu/>

⁸ <https://www.panosc.eu/>

⁹ <https://sshopencloud.eu/>

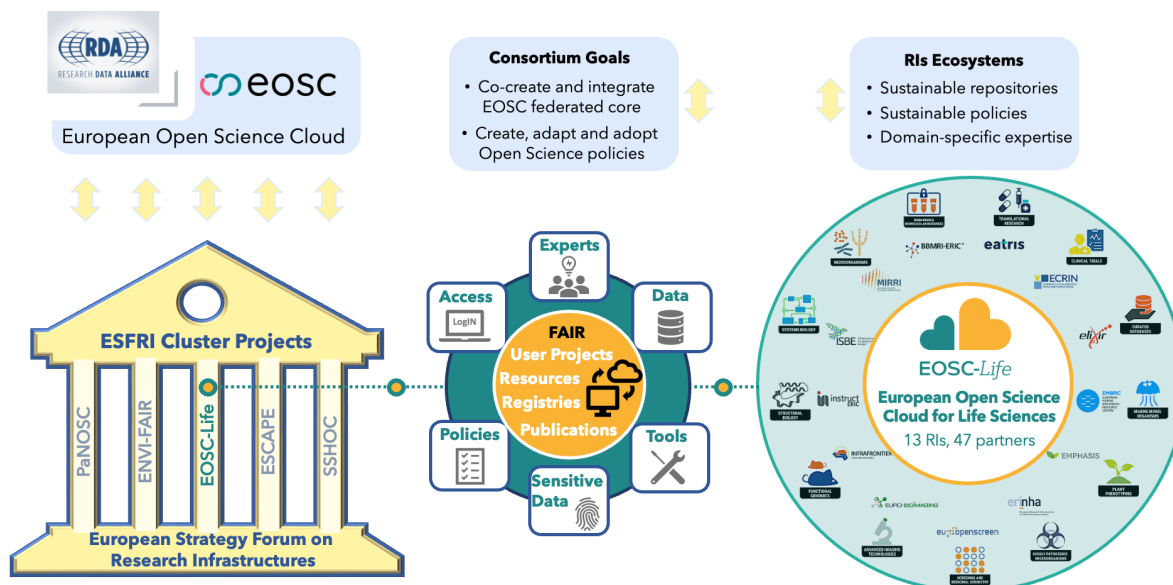


Figure-1: Position and organisation of EOSC-Life in the EOSC and LS RIs ecosystems: EOSC-Life is one of five ESFRI Science Cluster projects (the other four clusters are PaNOSC, ENVI-FAIR, ESCAPE, SSHOC). EOSC brings together 13 LS RIs and domains experts to populate EOSC with LSs FAIR data, tools, resources, harmonised solutions, policies and guidelines, and concrete user projects altogether facilitating interoperability in the EOSC (respective interactions and impact are illustrated by arrows).

EOSC-Life facilitates and harmonises the discovery, access and analysis of data for European and non-European researchers across disciplines. It also provides ways to visualise data creatively, in models simulating reality. To achieve this, the EOSC-Life consortium set out to:

1. Publish data resources and associated metadata from LS RIs;
2. Work with scientists who develop computational methods and packageable analysis tools, and connect the latter to data;
3. Develop guidelines and tools to be able to exploit sensitive data from human research participants in a secure manner.

The **EOSC-Life** consortium has worked to shape and strengthen synergistic partnerships with RI users and external communities through **open calls for user projects**¹⁰ (Haley et al., 2020), so as to:

1. Create RI data resources, to support the work of individual facilities, or multidisciplinary consortia, to make data FAIR in the cloud;
2. Partner with RIs to promote training and staff development;
3. Encourage research communities to engage with EOSC-Life to develop open services [R11];
4. Foster and support collaboration between academia and industry;
5. Deal with sensitive data.

¹⁰ <https://www.eosc-life.eu/calls/>

Domain experts in cloud computing, computational workflows, FAIR data, ethics and data management were assigned to dedicated project teams, supporting the implementation of projects by providing multidisciplinary expertise (Gribbon et al., 2020) **[R1]**. Through open calls, and working with existing RI-managed data resources, EOSC-Life created practical solutions that met the real needs of life scientists. Examples of outcomes, developed or enhanced during EOSC-Life, are listed in the EOSC Association and the RDA Key Exploitable Results report (Nardello et al., 2022) (Table-1) **[R2]**.

Table-1. Key Exploitable Results from EOSC-Life used by communities

I. LS Data in EOSC: FAIR LS data resources for cloud use (organisational sustainability, technical sustainability, sustainable access to data, sustainable data governance)	
COVID-19 Portal https://doi.org/10.25504/FAIRsharing.f3b7a9	Resources focusing on COVID-19-related data
Common Provenance Model https://fairsharing.org/4610	Standard based on W3C PROV, which allows publishing, interlinking and finding provenance information in distributed environments
Clinical Trial Data Repository https://fairsharing.org/3067	Portal allows users to search for studies using a variety of criteria and identify the data objects associated with them.
Ontology Lookup Service (OLS) https://doi.org/10.25504/FAIRsharing.Mkl9RR	Repository for biomedical ontologies that aims to provide a single point of access to the latest ontology versions.
Ontology Cross Reference Service (OxO) https://doi.org/10.25504/FAIRsharing.0c6fea	Database of ontology cross-references (or xrefs).
Sensitive data toolbox https://fairsharing.org/3577	Open, digital and collaborative space for biological and medical research, across LS RIs
II. LS Toolkits in EOSC: eco-system of innovative LS tools and key services/standards in EOSC (sustainability through re-use, technical sustainability)	
FAIRsharing https://doi.org/10.25504/FAIRsharing.2abjs5	Resource on (meta)data standards, inter-related to databases and data policies
Workflowhub https://doi.org/10.25504/FAIRsharing.07cf72	Registry for publishing scientific computational workflows
Schema.org https://doi.org/10.25504/FAIRsharing.hzdq8	Community activity handling schemas for structured data
Bioschemas.org https://fairsharing.org/3517	LS branch of Schema.org, aims to improve data interoperability in LS
RO-Crate https://doi.org/10.25504/FAIRsharing.wUoZKE	Community effort focusing on packaging research data with metadata, complementing richer metadata standards
FAIR Cookbook https://faircookbook.elixir-europe.org	Recipes for the FAIRification journey
Galaxy https://galaxyproject.org/eu/	Open-Source project for FAIR data analysis
Scipion (workflows) https://doi.org/10.25504/FAIRsharing.EsY1WF	Cryo-EM image processing framework
BAND https://forum.eosc-life.eu/t/the-band-a-virtual-desktop-for-bioimage-analysis-in-the-cloud/74	Virtual Desktop for bioimage analysis in the cloud
RDMKit https://rdmkit.elixir-europe.org/	Best practices and guidelines to help you make your data FAIR
III. Harmonisation of Access and Policies across the LS (RIs)	
LS Login https://lifescience-ri.eu/ls-login.html	Common Authentication and Authorisation Infrastructure (AAI) system for LS RIs and communities
ARIA https://aria.services/	User access management system (AMS)
Open Calls Procedures https://zenodo.org/record/4048442	Guidelines for organising, running and integrating Open Calls and user projects into EOSC framework

B. Sustainability and governance

Governance concerns the responsibilities for managing research data and tools over the long-term, in compliance with applicable regulations, as well as the allocation of suitable resources for data availability. **Governance structures** that engage research communities to promote best practices consistently are essential to ensure **long-term accessibility** to research outputs [R8]. Here, research funding agencies play essential roles in catalysing best practices (Jahn et al., 2023). Many of them now require all data to be shared or deposited in specific repositories that apply FAIR and TRUST (Transparency, Responsibility, User-Focus, Sustainability, Technology - Lin et al., 2020) principles. Best practices resulting from consultation with relevant user communities in governing data access, and resource sharing, should be an explicit part of operational management and risk mitigation (including cybersecurity) of research organisations, and of guidelines for infrastructures managing long-term sustainable access to data. These aspects have been highlighted in several high-level reports by, e.g. the Organisation for Economic Co-operation and Development (OECD)¹¹, G7, the Group of Senior Officials on global RIs (GSO)¹² and the EC via ESFRI¹³.

In this context, the core goal of EOSC-Life is to facilitate data interoperability: to allow the interoperation and analysis of data across disciplines, whilst conforming to necessary governance and consent procedures. Data and metadata resources gain **added value** when they are **curated** by experts [R1], well **annotated**, **processed** in sophisticated ways, and **integrated** into or linked to other datasets. For example, the value of sequencing data is enhanced when they are linked to the corresponding organism-level functional data, allowing the connection to be made between genotype and phenotype e.g. for crop performance or human diseases. Such an approach is illustrated by the National Human Genome Research Institute - European Bioinformatics Institute (NHGRI-EBI) catalogue of human genome-wide association studies¹⁴ (Sollis et al., 2023). To facilitate such data integration over time, and therefore promote sustainability through re-use, however, interoperability between more disparate data types needs to be developed. Adequate data management is a key component of this interoperability. Minimising information loss and creating an auditable trail ensures data can be trusted. As interoperability also relies on the granting of permission to access and re-use available data, it is essential that research outputs are accompanied by clear licensing information, and for the licences used to be as permissive as possible [R11]. In addition, the procedures and standards used to achieve added value must be clearly documented and made transparent to stakeholders, as was reported, for example, by the plant science community for the 'MIAPPE'¹⁵ (Minimum Information About a Plant Phenotyping Experiment) standard (Papoutsoglou et al, 2020) [R2]. As regulatory and ethical requirements have to be met, harmonising and simplifying data-related legal frameworks across EOSC would also reduce frictions in data access and re-use [R9].

¹¹ <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463>

¹² <http://www.gsogri.org/>

¹³ https://www.esfri.eu/sites/default/files/u4/ESFRI_SCRIPTA_VOL2_web.pdf

¹⁴ <https://www.ebi.ac.uk/gwas/>

¹⁵ <https://www.miappe.org/>

C. Sustainability at multiple scales

Smaller teams and projects from specialist communities, such as those integrated through EOSC-Life Open Calls, often lack access to expertise in FAIR data management, in setting up sustainable hosting resources, or in implementing interoperability using available standards. This means newly established **resources risk being abandoned** after projects are completed. To address this, the EOSC-Life open call process was structured to demonstrate first the utility of its support model for sustainable FAIR data management **[R3]**, and then to explore more complex needs, including those of sensitive- and industry-related data resources. Project applicants were asked, as part of the project design and implementation plan, to identify outcomes that would need to be sustained, and relevant domain experts, and then provided support and guidance **[R1]**. The sustainability of outcomes was analysed at the end of the project, in an iterative lessons-learned process, using “World-Cafe¹⁶” methodology. Supported by EOSC-Life publications **[R2]** and training materials, many user projects contributed to improving the quality, functionality and scope of existing resources, and created new resources, such as data repositories and standards. Significant impacts are described in Supplementary_Information_S2. Overall, this process highlighted the benefit of increasing awareness of existing sustainability solutions, and of training as a path to promote sustainability in small projects. In a fragmented environment of highly distributed data sources, both sustainability and data interoperability can be greatly enhanced by data integration into data warehouses. EOSC-Life supported the initiation of large, centralised platforms, showing how these could be quickly built to meet urgent community and societal needs **[R6]** e.g. the COVID-19 Data Portal¹⁷ (April 2020) enabling researchers to upload, access and analyse COVID-19-related reference data, and specialist datasets, as part of the wider European COVID-19 Data Platform (Supplementary_Information_S3 & **[R6]**).

D. Sustaining FAIR by connecting communities

EOSC-Life provides a framework for 13 LS RIs at different stages of organisational development to establish partnerships, and ensure service interoperability. These RIs represent research communities with different levels of adoption of Open-Science and FAIR practices. Certain RIs have produced white papers that map their possible paths to sustainability, e.g. ELIXIR^{18,19}. The exchange of knowledge among RIs should evolve into common organisational sustainability strategies **[R8]**. For instance, Memoranda of Understanding (e.g. between ELIXIR and BBMRI²⁰, or EU-OPENSSCREEN²¹, EURO-BIOIMAGING²² and INSTRUCT²³)²⁴ have been established to formalise cross-RI partnerships.

RI-associated communities have also learnt about different Open-Science and FAIR practices, through EOSC-Life activities. Initial work highlighted the steps required to ensure

¹⁶ <https://theworldcafe.com/key-concepts-resources/world-cafe-method/>

¹⁷ <https://www.covid19dataportal.org/>

¹⁸ <https://elixir-europe.org/>

¹⁹ <https://f1000research.com/documents/8-1642>

²⁰ <https://www.bbmri-eric.eu/>

²¹ <https://www.eu-openscreen.eu/>

²² <https://www.eurobioimaging.eu/>

²³ <https://instruct-eric.org/>

²⁴

<https://www.eu-openscreen.eu/newsroom/eu-openscreen-news/ansicht/eu-openscreen-has-signed-a-memorandum-of-understanding-mou-with-eurobioimaging-and-instruct-eric.html>

a common use of open formats and ontologies by RIs, and the need to develop Application Programming Interfaces (APIs) for data exchange. As a pilot for interoperable resources, graph databases and knowledge visualisations built around aligned metadata standards were developed for COVID-19 and Mpox (Karki et al., 2023) [R5]. Technical sustainability was addressed by including outcomes in enduring resources such as the FAIR Cookbook²⁵ (Supplementary_Information_S4) and RDMKit²⁶ (ELIXIR, 2021), which integrate findings from multiple academic and industrial projects, provide recipes for FAIRification, and tools for Research Data Management (RDM), respectively. These community projects support Open-Science and ensure that access to materials is sustained beyond a single project [R2]. EOSC-Life also reached out across communities to promote an inclusive framework [R11] (Figure 2) supporting the interoperability and portability of software tools and computational workflows from different domains, through the use of a common set of technologies, such as software containerisation. Gaps in the FAIRification of computational workflows were addressed by creating the **WorkflowHub**²⁷, a registry for **workflows in a highly federated ecosystem** of different workflow managers and community-owned repositories (Goble et al., 2023). By using Research Object (RO)-Crates, lightweight packages of research data with their metadata, and established web approaches (schema.org, JSON-LD, CWL, and GA4GH APIs) [R9], workflows can be exchanged between services for their execution and testing (see LifeMonitor²⁸ in Supplementary_Information_S6). They can also be exported to long-term archives, and described so that they can be reimplemented or executed in alternate systems [R5].

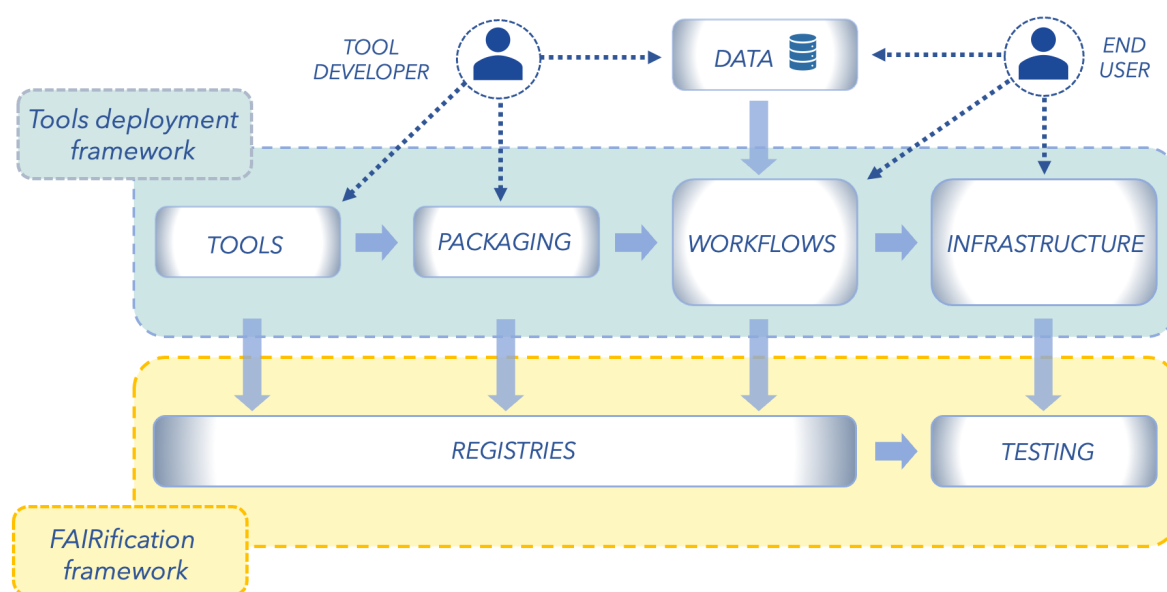


Figure-2: EOSC-Life operational framework, extending from the end user to the EOSC ecosystem. Dashed lines indicate use of and contributions to the framework, while thick arrows indicate process flow within the framework.

²⁵ <https://faircookbook.elixir-europe.org>

²⁶ <https://rdmkit.elixir-europe.org/>

²⁷ <https://workflowhub.eu>

²⁸ <https://app.lifemonitor.eu/>

PART-2: Technical components developed to support sustainability

A. EOSC-Life sustainable Open-Science toolkit for FAIR RDM services

FAIR interoperability services ensure that (meta)data use a formal, accessible, shared and broadly applicable language to represent knowledge, typically relying on vocabularies or ontologies that follow FAIR principles and make qualified references to other (meta)data. They are needed for the re-use of data and tools, but also to address the complexity of the biomedical domain, ranging from the diversity of samples and data-generating technologies to variable granularity of datasets. **Ontologies should be accessible**, e.g. via the Ontology Lookup Service²⁹ (OLS) (see Supplementary_Information_S5), and the annotations must have web-resolvable Persistent and unique IDentifiers (PIDs), as recommended by the EOSC Association Task Force on PID Policy and Implementation³⁰ (EC, 2021) **[R5]**. Mappings between ontologies, vocabulary schema, or coding standards are important to ensure the quality and sustainability of the metadata. EOSC-Life implemented interoperability services such as the Ontology Cross Reference Service³¹ (OxO) in the form of a suite of semantic services, deployed using cloud principles and technologies (Supplementary_Information_S5) **[R9]**. Sustainability has been achieved by extending and evolving pre-existing services, as well as by developing new services to answer multiple user needs. Cloud deployment has become accessible and affordable to all, and some datasets/data resources are being deployed in industrial or secure environments, highlighting the importance of **making interoperability services portable** to increase their sustainability.

B. Sustainability of sharing and shared (meta)data

Data and software must be archived in **online portals** along with rich and semantically-annotated **metadata** to be **findable and usable** over the long-term. Sustainability through re-use requires that the data contain, or be intimately linked to, their own metadata to ensure that the data are useful and understandable, even after they have been downloaded from the portal **[R5]**. Open specifications of data formats, semantics, provenance and machine-actionable data content are the cornerstone of sustainable data interoperability, by making data accessible to all users and services, without tying data exchange and processing to specific hardware, software or groups. Open specifications offer guarantees against obsolescence as they can evolve with community needs, and new implementations can be made available as technologies evolve **[R11]**. EOSC-Life contributed to increase domain coverage, for example, with the Recommended Metadata for Biological Images (REMBI) (Sarkans et al., 2021) or for microbial resources, the development of a unique identification system based on Digital Object Identifier (DOI) (Romano et al., 2022). This identification system enables the users to retrieve microbial metadata (including ecological and legal data) from Data MIRRI dataportal and potentially connect to other data housed in other RI **[R5]**.

²⁹ <https://www.ebi.ac.uk/ols4>

³⁰ <https://www.eosc.eu/advisory-groups/pid-policy-implementation>

³¹ <https://www.ebi.ac.uk/spot/oxo>

C. Curation of FAIR data resources and collections

The constant curation of data resources ensures the quality, accuracy and depth of data, improves their reuse and builds trust in those data resources [R7]. This requires funds to sustain the expert interpretation and classification of vast amounts of information (Bourne et al., 2015; Karp, 2016; Chen et al., 2020). Expert curation is essential for the provision and dissemination of high-quality, accurate data, and the associated metadata, which can be productively used for specific research applications. **Delivering early and efficient curation** processes that reduce human effort through automation is a challenge for sustainability; rewards can be efficient incentives [R10]. Large-scale data curation can be supported by specific biomathematical models used to process and integrate big datasets programmatically, e.g. JWS Online³² and EBI BioModels³³. These approaches often help the discovery of inconsistencies among the datasets, in terms of biological, physical, chemical, or even mathematical incongruities. The registration, dissemination and application of reference collections of curated [R7], FAIR data repositories [R5] are primary goals of EOSC-Life (Perseil et al., 2020) (Box-3).

Box-3. Collection and resource management

Some of the collections described by open standards are provided in support of EOSC-Life Demonstrators and Open Calls projects (Parkinson et al. 2021, 2022). For instance, the PDB-REDO resource (EOSC-Life Open Call project) exploits collections of curated crystallographic datasets to refine, rebuild and validate structural models of biomolecules automatically (Joosten et al., 2014). An extensive collection of open standards and data portals is now available to LS communities. To facilitate discovery and use, these resources are registered in FAIRsharing³⁴.

FAIRsharing collection and EOSC

Embedded in EOSC, and recommended by funders and publishers, FAIRsharing is a curated, informative and educational resource for data and metadata standards that is interrelated to databases and data policies and extends across all disciplines (Sansone et al., 2019). FAIRsharing encourages users to discover, select and exploit these resources with confidence. It also encourages producers to make their resources more sustainable and discoverable, so that they will be more widely adopted and cited. In the context of the sustainable findability of EOSC-Life aligned outputs, a dedicated collection of its 133 data resources are richly described. The descriptors used are also served by FAIRsharing in a machine-readable form to feed the information into EOSC portals and other tools, such as the OpenAIRE Graph (Supplementary_Information_S9). Many of the descriptors themselves (e.g. life cycle status, links to sustainability documentation, and relationships to the organisations who fund and maintain them) are also key to helping both EOSC and the wider research community assess the sustainability of the resources curated within FAIRsharing.

³² <https://jij.biochem.sun.ac.za/>

³³ <https://www.ebi.ac.uk/biomodels/>

³⁴ <https://fairsharing.org/>

D. Provenance for sustainable use

Provenance provides the history of an object. Thus it can be used to assess an object's quality, reliability, usefulness, and trustworthiness, which are all important for its re-use (Moreau, 2013). Providing the provenance itself has to be technically and organisationally sustainable (Box-4).

Box-4. Key aspects for ensuring sustainable provenance (Traceability of Data)

Provenance information is always provided via metadata (**who, what, where, when, how**) and often also consists of heterogeneous data, such as software logs, workflow input files and Standard Operating Procedures. As such, the steps that need to be taken to ensure the sustainability of the provision and the existence of the (meta)data apply equally to the provenance (meta)data. To achieve this sustainability, the following are needed and, in turn, should themselves be sustainable resources:

- **Standards**, e.g. for metadata and data formats, vocabularies, provenance and workflow management and security;
- **Tools for accessing provenance information**, such as the LS AAI, OLS and FAIRsharing;
- Suitable and continually **relevant access technologies and methodologies** for human and machine consumers;
- **Accessibility via online archives/portals**; the adoption of policies concerning how long provenance (meta)data are stored and shared, and how long the necessary AAI (especially for sensitive information) is provided, updated and versioned;
- Appropriate **cryptographic techniques**, e.g. hashes and digital signatures, used in combination with security policies.

EOSC-Life has facilitated the development of the “Common Provenance Model” (CPM) (Wittner et al, 2022). It introduces a framework to handle distributed provenance information [R5]. Such a framework has already been applied in the BY-COVID³⁵ project, and its use in other projects, including in BIOINDUSTRY 4.0³⁶ is planned. The **CPM** also serves as a conceptual foundation beyond open science communities as it is being used for the proprietary ISO 23494³⁷ provenance standard series, currently being developed to provide a standard **sustainability framework** for the biotechnology industry.

The same principles apply to computational work, where the use of modern workflow management systems provides a “retrospective provenance”: the detailed record of the implementation of a computational workflow with information related to every executed process, and the execution environment used (Khan et al., 2019). In this context, the FAIRification of software, which is often initially conceived as a stand-alone tool, would benefit from the software also being designed for inclusion in workflows (Brack et al., 2022). The goal is to develop computational workflows that can be easily used with modern workflow management systems, as well as be retrieved and deployed seamlessly from registries such as the WorkflowHub (Goble et al., 2021 & 2022), while remaining, evolving and being curated in their home repositories [R7]. Workflows become more sustainable thanks to their portability across hosts, and the reduced deployment overhead for new users that such workflow management systems provide. Two main EOSC-Life contributions to the management of provenance are the systematic use of RO-Crate (Soiland-Reyes et al.,

³⁵ <https://by-covid.org/>

³⁶ <https://cordis.europa.eu/project/id/101094287>

³⁷ <https://www.iso.org/standard/80715.html>

2022) and LifeMonitor [R9]. The latter service complements the WorkflowHub by ensuring the correct operation of the workflow over time, through monitoring and then triggering automated workflow tests.

E. Traceable legal requirements from the data

Data that were collected from physical sources, e.g. human subjects, lab animals, or field samples, for which legal and policy steps of any type were required, are not by definition usable, unless these steps are documented. In different cases, sustainable access to the original or derived data must be ensured. Proof of compliance with legal requirements is part of the provenance of data [R5]. Clear governance approaches for data use are needed to ensure that the investment in data generation is matched by sustained usage, and to ensure that relevant laws are adhered to [R8]. For controlled access data, such as human genetic data, GA4GH standards can be referenced to annotate each data use case, and provide a machine-readable terminology (Lawson et al, 2021) that can be implemented by multiple resources. EOSC-Life has contributed to the development of these standards, as well as operationalisation of the standards, e.g. through development of the LS AAI system to support machine readable access protocols (“GA4GH passport and visas”, Cabili et al., 2021)

Sustained usage and FAIR implementation require machine-readable (meta)data and software licences. We have found that the more open the licence, the more likely that the data or software will be reused. EOSC-Life has promoted technical sustainability by using permissive licences³⁸ and open-source codebases³⁹, which can be reused and receive contributions from a wide community.

F. Role of cloud providers

In the past, widely accessible services were designed and deployed from a central location. As cloud deployment costs have decreased, and because some datasets/data resources are deployed in industrial and secure settings, the interoperability services must also be portable. This offers software scientists considerable benefits, as portable resources are more agile to be developed with multi-site teams [R6], and their deployment is often simpler, resulting in a reduction in the overall effort to sustain a service.

To enable the deployment of LS workflows in the EOSC, EOSC-Life relies on a technology stack that standardises the software installation process to ensure the reproducibility of computational workflows, and automates software deployment in cloud environments. The key to community adoption has been the availability of free and publicly accessible services and technologies [R11]. Integrating such free-to-use services and technologies has made the research process more efficient, reproducible and collaborative, but has also resulted in some sustainability and reproducibility risks. Currently, components of the computational ecosystem rely on services from commercial entities. For example, we estimate that GitHub⁴⁰ provided services to the bioinformatics community in 2021 with a value of over \$1

³⁸ <https://fossa.com/blog/all-about-permissive-licenses/>

³⁹ <https://pncnmnp.github.io/blogs/oss-guide.html>

⁴⁰ <https://github.com/>

million. Similarly, quay.io⁴¹ provides services with an estimated \$500,000/year in value for BioContainers⁴². Communities work on the assumption that the conditions under which these technologies and services are provided will remain compatible with the research requirements and the companies' abilities, e.g. that they will remain free of charge and freely accessible, and have adequate performance characteristics. This may not always be true (Box-5).

Box-5. Code and software management

Already, over the project's lifespan, we have seen changes in and restrictions to some widely used services. To highlight a few:

- **TravisCI**⁴³: a popular continuous integration system, initially used by the LifeMonitor, switching to a paid model;
- **DockerHub**⁴⁴: a well established repository of container images, which introduced limits to the number of containers that can be pulled without subscription;
- **Conda**⁴⁵: a popular open source package management and environment management system which changed the licence terms of its default channel usage.

Code repositories are particularly at risk, as it has been demonstrated in the past by the disappearance of Google code (2006-2016) and the bundling of junkware/malware by SourceForge⁴⁶ with each download, in an attempt to become profitable. Larger institutions already run their own code-hosting platforms, but most RIs do not have sufficient IT infrastructure or human resources in place to do so.

G. Sensitive data challenges

Across research domains, sensitive data present challenges related to sustainability, cross-domain categorisation, and discovery. Sharing sensitive data within EOSC-Life is particularly challenging when they need to be made available to third parties not contractually bound to the original data controller. The sensitivity of the data may arise not only from their personal nature, but could originate from intellectual property considerations, biohazard concerns, or compliance with the Nagoya Protocol⁴⁷. With regard to Access and Benefit-sharing (Supplementary_Information_S7), and as sharing may be perceived as a risk, EOSC-Life has developed a prototype toolbox (Ohmann et al., 2022) allowing researchers to find existing, reliable resources relevant for sharing sensitive data across all participating RIs (Supplementary_Information_S8), supported by an interdisciplinary categorisation system [R9] (David et al., 2022) for which sustainability is increased by iterative building, validated by iterative consensus (Figure 3). Another example, the Open-Source Secure Data Infrastructure and Processes Platform⁴⁸ (OSSDIP - Weise et al, 2021), was developed as a blueprint that can be deployed, e.g. for educational purposes [R11]. OSSDIP clarifies the rather complex underlying concepts, and reduces the initial burden of deploying such an infrastructure, giving researchers access to sensitive data in

⁴¹ <https://quay.io/>

⁴² <https://biocontainers.pro/>

⁴³ <https://www.travis-ci.com/>

⁴⁴ <https://hub.docker.com/>

⁴⁵ <https://docs.conda.io/en/latest/>

⁴⁶ <https://sourceforge.net/>

⁴⁷ <https://www.cbd.int/abs/>

⁴⁸ <https://www.ifs.tuwien.ac.at/infrastructures/ossdip/>

return.

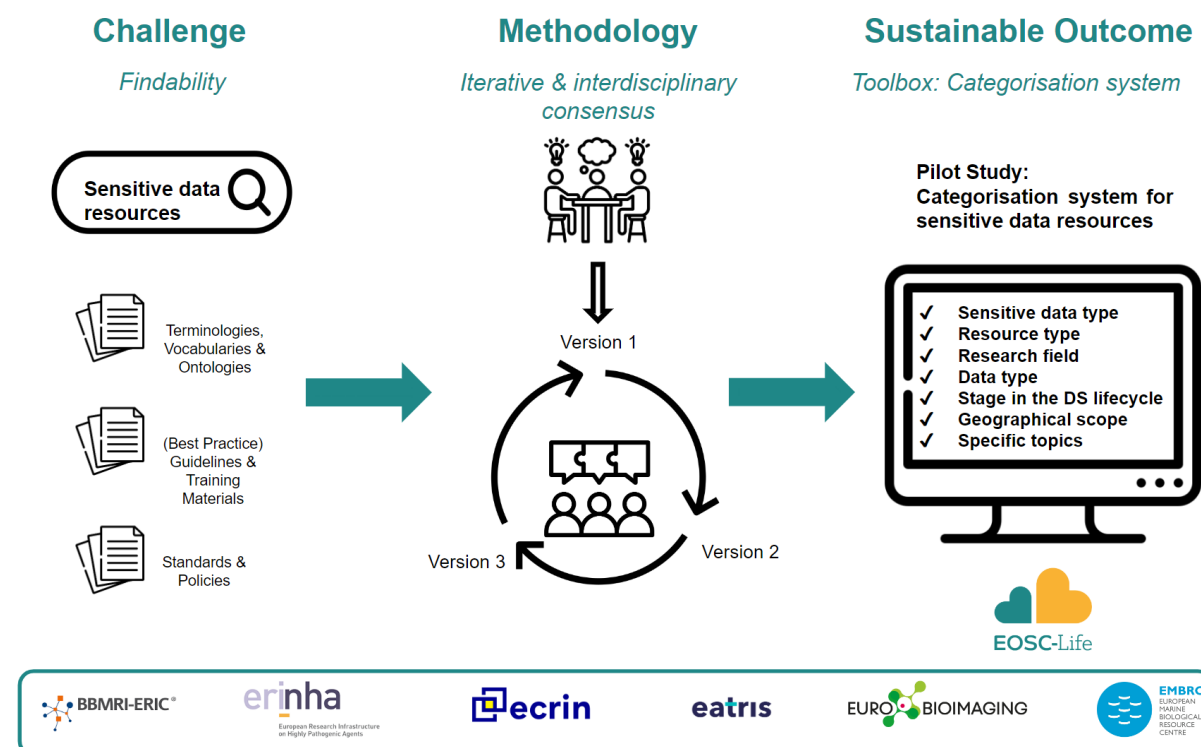


Figure-3: EOSC-Life Sensitive Data toolbox development scheme: One main challenge regarding sensitive data concerns the findability of resources and terminologies used to classify data. An iterative method, involving the subset of the participating LS RIs (BBMRI, ERINHA⁴⁹, ECRIN⁵⁰, EATRIS⁵¹, EURO BIOIMAGING, EMBRC⁵²), with several versions, was used to reach an interdisciplinary consensus on a sustainable categorisation system, and the classification of toolbox resources by sensitive data type, resource type, research field, data type, stage in life cycle, geographical scope and specific topics related to sensitiveness.

H. Skills and capabilities to design, setup and implement

Systematic assessments have shown that both individual projects and RIs display a broad range of maturity levels, in terms of their cloud implementation status, and the technical skills of team members involved in the deployment and management of deployed resources. To address existing gaps, and support the development of organisational expertise **[R1]** to sustain long-term operation, as described below (Part 3 & **[R4]**), training processes and materials were established in EOSC-Life. Having appropriately trained and rewarded personnel not only benefits the RI, but provides individuals with career advancement opportunities **[R10]** and transferable skills which, in turn, support innovation **[R12]** and economic development.

⁴⁹ www.erinha.eu

⁵⁰ <https://ecrin.org/>

⁵¹ <https://eatris.eu/>

⁵² <https://www.embrc.eu/>

PART-3: Driving and sustaining the EOSC-Life community and expertise through training and skill development

Training initiatives were designed to promote community development and cross-RI collaboration, as well as to help LS RIs meet the technical challenges of sustainable FAIR implementation (David et al., 2020) with respect to both the data and training materials themselves⁵³. These initiatives aligned with our aim to promote a cultural change, encouraging users to adopt Open-Science practices. **Training is a community driver**, and not only provides technical expertise enabling researchers to improve the handling of their data (e.g. data management, analysis, annotation) in a sustainable manner, but also reinforces the expertise. Supported by a dedicated training team in EOSC-Life, collaborative training activities were delivered by multiple consortium partners, and RI experts, permitting cross-domain, cross-RI training programmes for diverse user needs **[R4]**.

Domain-specific training activities were also promoted and supported through the “Training Open Calls” (Box-6).

Box-6: BioImage Data Analysis training supported by EOSC-Life Open Calls

The NEUBIAS training school (May 2023)⁵⁴ provides one example of RIs benefiting from the EOSC-Life Training Open Call. This course was the continuation of an initiative that started in 2017, aiming to bring BioImage Data Analysts closer to recent solutions for computing and workflow-based image analysis in the cloud. For the NEUBIAS⁵⁵ project, EOSC-Life support has been essential for sustaining, further developing and running this valuable training initiative.

Image-based data are ubiquitous across multiple LS-RI domains. Because of the multimodality and the large volumes of data, linkage of image data to publication analysis workflows and the sustainability of access and hosting are key ongoing challenges. Training in this field has been provided regularly through EOSC-Life partners and project leads, including FAIR training from European Molecular Biology Laboratory (EMBL)-EBI (Image Analysis and Machine Learning⁵⁶) or FAIR Data training run by national or international consortia e.g. Euro-BioImaging⁵⁷.

Overall, 15 training activities⁵⁸ were established, which benefited 13 LS RIs and 27 user projects. The training calls contributed to the strategic effort to link the work of EOSC-Life to each RI's development plans, improve the visibility of ongoing work on RI-led resources, and address gaps in expertise by carrying out specific activities and capacity building (Box-7) **[R1]**.

To speed up the evolution of the RIs, and the integration of user projects, we employed consultative orientation meetings and hackathons, including the cross-RI FAIR Hackathon for Training on Demonstrators, and Open and Internal Calls project teams. This consortium-wide training was provided to **assess and improve the FAIRness** of the projects, orient the teams based on available examples, share knowledge on solutions and resources provided by EOSC-Life experts **[R1]** and RIs communities. Overall, the initiatives

⁵³ <https://www.eosc-life.eu/services/training/>

⁵⁴ <https://www.eosc-life.eu/news/neubias-defragmentation-training-school/>

⁵⁵ <https://eubias.org/NEUBIAS/venue/>

⁵⁶ <https://www.ebi.ac.uk/training/events/microscopy-data-analysis-0/>

⁵⁷ <https://www.eurobioimaging.eu/news/euro-bioimaging-guide-to-fair-bioimage-data/>

⁵⁸ <https://www.eosc-life.eu/services/open-call-training/>

provided fruitful opportunities and platforms for goal-oriented, hands-on teaching, introducing EOSC-Life topics to consortium teams. Such interactive platforms for consortium-wide exchange facilitated further networking and collaboration among experts from the EOSC-Life community and strengthened capability building in new projects (Example in Supplementary_Information_S10).

Box-7. Sustainability of knowledge through EOSC-Life Training assets

By supporting cross-disciplinary sharing of best practices and methodologies, the consortium contributed to sustainability of knowledge across LS communities through:

- 1) Training on EOSC-Life and RI **cloud-based open access** solutions and resources, services, and expertise in the dissemination, practical uptake and targeted adoption of tools;
- 2) Training related to **Open Science** practices and **FAIR guidelines**, harmonising the requirements across a spectrum of LS disciplines whilst addressing needs in individual branches;
- 3) Establishment of **expert groups** to provide guidance and consultation to broader communities within the framework of Open Science and FAIR Science, resulting in a network that can continue operating after the project;
- 4) Development of collaborative **cross-disciplinary training approaches**, modules and materials, as well as formats for effective training seminars and workshops, including the joint creation of best practice solutions for efficient remote training;
- 5) **Integration of 27 Demonstrators**, namely Open and Internal Calls user projects from different disciplines, into EOSC-Life and across the RI landscape by providing consultation and guidance for the teams in the form of advice from EOSC-Life experts and consortium partners;
- 6) Designation and training of the so-called EOSC-Life **consortium translators**, to create a cohort of professionals that understand the jargon, needs, work culture and drivers of professionals working in other areas of expertise or RIs;
- 7) Creation of a **training community** with a practical understanding of EOSC but also with conceptual and technical knowledge in different LS domains.

This consortium effort in building up knowledge, skills and trust, made possible rapid responses to large-scale scientific, societal, environmental, or other challenges of concern, e.g. pandemics similar to the COVID-19 pandemic [R6]. A compelling EOSC-Life example for implementation of interdisciplinary training, to address challenges that emerged during the coronavirus pandemic, was the creation of an epidemiology mathematical modelling training course⁵⁹, containing content relevant to future pandemic situations.

EOSC-Life's investment in **building a community of experts**, and in providing training to these experts, has already increased organisational and technical sustainability. It is hoped that experts trained in the EOSC-Life consortium will continue working in RIs, taking on the role of master trainers, disseminating their expertise to others. To promote and strengthen the evolution of the cross-RI expert network and support cross-community activities, EOSC-Life also organised the "EOSC-Life translator training series" [R9], which enables the experts to understand the drivers and challenges in the different RIs and data professions, and break down silos [R1].

⁵⁹ <https://www.eosc-life.eu/news/training-modelling-covid-19-epidemics/>

Discussion

Technical and operational sustainability

The long-term preservation of data and tools for re-use currently relies on multiple repositories. In such a fragmented landscape, sustainability depends on broadly distributed funding, but also on adequate linkage and/or cross-referencing between repositories. This is becoming a pressing issue for multi-modal datasets, where individual components are split into data-type specific repositories. Examples of this are single cell omics datasets combining imaging and sequencing, where the sequencing data are submitted to sequence archives, and images are submitted to imaging archives, without necessarily any formal linkage. The wider adoption and implementation of linked data principles⁶⁰ (Bizer, Heath, and Berners-Lee, 2009) could mitigate this. Fragmentation of a project's data also prevents users from accessing them as a whole and fully understanding them. One way forward could be to build dedicated web applications and API access allowing the visualisation and browsing of project-specific data on top of distributed data archives. While some research groups already implement this in their projects, such web applications have tended to have a limited lifetime.

The use of cloud resources in EOSC-Life has been limited to well-established academic clouds mainly due to associated costs. A sustainability threat due to the use of a commercial cloud arises from, on the one hand, the expertise gap created by outsourcing and, on the other hand, a lack of cost control due to vendor lock-in, which is seen as incompatible with the time-limited, fixed budget of most research grants. In the case of sensitive data, the complex legal framework and country derogations means that most institutions are risk averse and researchers refrain from using cloud environments that are not directly under their control.

Validating the adaptability of sustainability components developed in EOSC-Life for use in other projects is key to ensuring the project's long-term impact. The recent EC's INFRASERV⁶¹ programmes represent both a significant challenge to models of sustainability for FAIR LS resources, and an opportunity to create more impactful cross-RI initiatives. By focusing on ISIDORE⁶², which is centred on pandemic preparedness (Richard, Stepanyan et Ewbank, 2022), to provide an INFRASERV perspective on the topics raised in this paper, we showed the critical importance of improving FAIRness Literacy, setting up incentives, training researchers, and providing FAIR expert support for sustainable data management **[R1]**, **[R4]**, **[R5]**, **[R10]**, (Supplementary_Information_S11).

Financial Sustainability

Open-Science allows collaborations to take place without the red tape of negotiating access, intellectual property transfer and detailed assignment of ownership. Thus, Open-Science practices are not only effective in bringing people together around common solutions, but are

⁶⁰ <https://www.w3.org/DesignIssues/LinkedData>

⁶¹

https://rea.ec.europa.eu/funding-and-grants/horizon-europe-research-infrastructures/research-infrastructure-services-support-health-research-accelerate-green-and-digital-transformation_en

⁶² <https://isidore-project.eu/>

also unparalleled tools for long-term sustainability in a complex landscape of national and international funding sources [R11].

The network of experts across the RIs and LS domains established as a result of EOSC-Life provides a solid foundation for ensuring the sustainability of knowledge, the reusability of EOSC-Life tools and resources, and the creation and promotion of EOSC services, increasing adaptability for emerging user needs. These experts, e.g. data stewards who have domain-specific skills, and a firm understanding of Open-Science Cloud solutions, are key to technical and operational sustainability within a FAIR and Open-Science framework (Figure-4). Setting aside the question of the acute lack of qualified personnel, the activities and disponibility of these professionals, with their unique skill sets, can only be sustained by ensuring adequate financial resources. This would allow stable employment, and keep cloud infrastructure solutions operational.



Figure-4: Sustainability lessons learned during the EOSC-Life project - key components driving, facilitating and challenging the sustainability process in an iterative way. **Sustainability essentials** are prerequisites including: availability of financial resources, competent experts, technical infrastructures, aligned policies and recognition. These components allow a set of **sustainability tools** operating towards sustainability to be defined: e.g. via integration of user projects (open calls); community training; dissemination addressing specific needs; creating and expanding the experts networks. The main linked **sustainability challenges** are: securing long-term financing to retain expertise and maintain solutions; engaging institutions and communities while harmonising alignment in common strategies and priorities; and enhancing a reward system that supports re-use of available resources.

EOSC-Life does, however, face a general sustainability challenge in that it relies almost completely on EC project funding. It is thus dependent on the **integration of its outputs into new EC proposals** and initiatives. These are constrained by their own limited lifespans and the need to focus on constant innovation and new developments, rather than on the operation of existing assets. This in turn can have a negative impact on research, as it can be easier to obtain funding to reinvent services and resources, rather than to support the long-term operation of existing valuable resources, and their experts (EOSC TF FinSus, 2022).

While, in the case of EOSC-Life, some resources have already been taken up by other projects, for many of its resources, there is still no immediate follow-up funding available. This means the project partners need to find alternative financial means to sustain the resources, as well as the respective expertise. In essence, the EOSC-LIFE project partners have two **options**:

- A partner organisation **takes it on as a core resource**; or
- The resource **becomes the responsibility of a community**, i.e. the means to operate, maintain and develop the resource is given through future grants and/or “in kind” contributions from intrinsically motivated research parties/individuals (in an equivalent model to various open-source communities).

It is important to note that, in either model, full responsibility not only includes the continued technical and operational support for the resource but also the curation of the resource, which ensures that the content is correct and up to date.

There is a risk, however, that if no project partner takes on the responsibility, and no immediate effort is made by the community, awareness of an existing tool can be lost. This leads to resource-intensive reinvention of something similar; it may take months, sometimes years before a tool is taken up again. For instance, the project Biotracks⁶³ provides a standard format for cell migration files and a series of converters that can be used to convert popular tracking software packages to the Biotracks formats. The project was supported by CORBEL (2016-2020). Due to the lack of funding, the project was put on hold. It is only now that the project has been revived as the need for re-use has emerged in the context of Open Microscopy Environment Next-Generation File Formats⁶⁴ (OME-NGFF) (Moore et al., 2021).

To address this issue before it became urgent, the EOSC-Life open calls included a consultation process to advise the many small teams (typically two to five members per participating institution) who were applying. As part of the project submission process, experts from EOSC-Life **analysed the needs of each team and provided a roadmap** with recommendations to ensure successful project implementation, even if the funding application was unsuccessful. The applicants acknowledged the utility of the consultation process. It helped several teams improve FAIR data management strategies that could be adopted even in absence of funding. This process is, however, resource intensive and relies on the availability of experts to provide consultations and the relevant training materials **[R1]**, **[R4]**.

Clearly, reinventing the wheel or developing new solutions inferior to existing ones may not be the best use of funding resources. There are **four critical phases** of resource development requiring specific financial support: **conception, proof of concept development, stabilisation, and community adoption** with long-term sustainability, dissemination, and maintenance **[R6]**. It is challenging to explain to funders and policy makers that the last two are crucially important **[R12]**. But until there is recognition that maintaining existing resources is critical and should be a priority for investments in RIs, maintenance will continue to rely on volunteer efforts and in-kind contributions, solutions that are clearly not sustainable.

⁶³ <https://github.com/CellMigStandOrg/biotracks>

⁶⁴ <https://ngff.openmicroscopy.org>

Therefore, to guarantee sustainability for successful project outcomes, individual RIs need to engage early on in any project with national and European funders, as well as through existing stakeholder networks such as the LS RI Strategy Board and ERIC Forum [R11], to present clearly and repeatedly the outcomes and messages outlined in this paper [R2].

Recommendations

Not all data infrastructures can or should be supported indefinitely. Financial sustainability is achieved when funders (public or private) are willing to **prioritise necessary long-term investments**, based on the expectation that **impact**, in terms of **science** and innovation, **can be achieved**, but also that **societal threats** can be tackled, such as pandemics and other disasters.

While the COVID-19 pandemic has proven the value of LS data and models for practical assessments and policy decisions, their wider potential still remains largely under-exploited. More efforts are needed to broadcast the many ways in which LS data can be used, and to help government representatives, journalists and members of the general public understand LS data. Therefore, new activities may be necessary that not only ensure that these data are sustainable, but also that they are impactful and considered as indispensable. All the work and experiences collected during EOSC-Life allow us to propose the following list of recommendations (Figure-5); even if some seem obvious, they can be difficult to implement if their importance is poorly understood from the start of a collective work:



Figure-5: Recommendations for ensuring and facilitating sustainability are based on 5 pillars shown here, clockwise from the top: to base the network on experts, and disseminate on a broad-scale; and to demonstrate, to act and plan training and improve metadata; to be prepared to act now, but remain

adaptable, and curate data as soon as possible; to strengthen the community with federated governance, harmonised and integrated RIs and sustainability actions rewarded; to be inclusive and open-minded, and to treat innovation and sustainability independently;.

BE RECOGNIZED: Focus on **strong credibility and recognition** for the research done:

- **[R1] Build with the experts:** The EOSC-Life open calls demonstrated that a consultative process, providing expert advice to teams before they submitted an application, resulted in improved adoption of FAIR data management strategies, irrespective of whether the application was funded or not. Keeping in mind the limited number of available experts and training materials, this approach could help future projects with limited funding options find ways to adjust their resources accordingly.
- **[R2] Publish, communicate, disseminate:** Effective communication through dissemination to funders that relays the impact of previous investments and the benefits of continued support is crucial for achieving long-term alignment between funding decision-making and the development plans for RI data/software/workflow resources. In addition, raising awareness for available and accessible tools and resources, i.e., disseminating and sharing FAIR solutions with a broader audience, and increasing their reusability, impact and overall sustainability, are key. Furthermore, peer-reviewed publications, preferably in Open Access journals, remain the main way to sustainably disseminate recognizable and findable statements and research. New publication types, efficient for dissemination and reusability, are emerging for all Digital Objects sustainability, as they are for data, workflows and softwares.

BE PRACTICAL: **Demonstrating, practising and reproducing** permit better and **sustainable adoption** of research outputs:

- **[R3] Demonstrate, not only tell:** EOSC-Life has shown that establishing FAIR resources that meet the needs of parallel LS communities creates the necessary scale to drive financial and operational efficiency and reduce fragmentation. The EOSC-Life strategy of learning by doing, based upon addressing the concrete needs of scientific demonstrators, helps funding organisations better understand the inherent scalability and translatability potential of our approach.
- **[R4] Plan training for acute and future needs:** Developing competent, effective and sustainable training methodologies and formats within the scope of an inclusive cross-disciplinary Open-Science framework is a stepwise, evolutionary process. It requires: a) connecting experts, trainers and users in a goal-oriented manner, such as through open calls, b) creating adaptive platforms with resources that promote targeted dissemination and the exchange of knowledge, enabling a rapid response to the needs of different communities, and c) providing sufficient sustainable funding schemes to keep this network of experts active and functional beyond time-limited tasks. Making investments in sustaining competent human resources and training is essential to ensure the uptake of tools, services and solutions and to integrate other expertise and diverse communities.
- **[R5] Metadata makes FAIR:** Repositories or portals hosting scientific digital objects in a semantically interoperable manner or collections of physical material (biobanks, culture collections) that provide sufficient provenance and other FAIR (meta)data with the objects they share are becoming more trusted sources of scientific resources, promoting their use in the community. For this reason, guaranteeing provenance is

recommended to ensure the practical sustainability of scientific objects and the services that provide them, i.e. allowing sustainability through re-use.

BE READY: Sustainability requires **agility and readiness to catch opportunities:**

- **[R6] Be prepared, agile and act timely:** Agility is often forgotten but is required to address emerging challenges, and this is particularly true in the LS. EOSC-Life has applied its action plan, which has a high level of adaptability, in order to be able to anticipate and provide non-centralized services and expert support during the COVID-19 crisis. To ensure sustainability, we must proactively take advantage of all relevant opportunities.
- **[R7] Curate NOW:** EOSC-Life has invested in projects which prioritise, leverage and extend expert curation processes. These projects build capacity, adding value to existing data and delivering processes which can be re-used and combined with automated techniques for data annotation. The provision of high-quality, curated datasets is also essential for the development of new AI and data-driven initiatives and projects in LS research. Assessing the efficiency of quality processes could help to improve research outputs.

BE UNIFYING: Sustainability of products is **based on clearly identified, recognised and driven communities:**

- **[R8] Establish federated governance:** The long-term responsibility for the sustainability-related activities must be clear from the outset. Responsibilities and roles that ensure sovereignty and access to sustainable services and tools must be organised during the implementation of research activities. Therefore, the provision of adequate funding and planning is vital, particularly concerning semantic interoperability and data quality.
- **[R9] Harmonise and integrate:** It is easier to encourage the long-term adoption of new operating systems by potential users when these can be connected with previously issued products, and if these are backwards compatible with previously issued tools, scripts and software. Harmonisation, reached by iterative consensus, facilitates and guarantees the interoperability of tools, data and solutions, and improves the understanding of concepts, functionalities and semantics shared across communities and disciplines. It is also easier to maintain services if they are fully integrated into the relevant communities' ecosystems.
- **[R10] Reward sustainability:** Further efforts and investments of resources devoted to increasing sustainability are often considered as unprofitable in the short term. They should, however, be promoted and rewarded as a means of increasing the quality of products and when they are recognized by users as being essential. Rewarding contributors for their extra efforts and making them visible in the community contributes to success and cohesion.

BE CLEAR: Inclusiveness and openness, as well as innovation, sustain communities:

- **[R11] Be open AND be inclusive:** Open specifications of data formats and data exchange protocols promote trust and inclusiveness and safeguard against obsolescence. Inclusiveness is critical to maintaining interest, increasing adoption, and integrating newcomers who can sustain and reuse solutions and tools. Community building is a key component for long-term training, good data-sharing literacy and efficient project output dissemination and reuse.

- **[R12] Treat innovation and sustainability independently:** Our recommendation for funders is that they recognize that sustaining the operation of a resource is an activity in its own right. For this reason, they should encourage the reuse and expansion of existing resources and facilitate the development of innovative tools and services. Consultation among stakeholders when developing funding calls can ensure alignment between the funders' aims and the community needs.

Conclusion

The EOSC has formed “a *federated and open multi-disciplinary environment where users can publish, find and re-use data, tools and services for research innovation and educational purposes*”⁶⁵. Through the EOSC-Life project, we set out to create an “EOSC for the Life Sciences”, connecting, and where necessary, further developing data resources, analysis tools and services that allow research communities to collaborate across national and thematic borders. A core part of the project and its sustainability strategy was to work in close partnership with the broader life-science community, via open calls for partnerships. This strategy has been successful. As EOSC-Life comes to its end, the services that emerged are being carried forward by a range of applied projects: BY-COVID is consolidating an Open-Science platform for pandemic preparedness; EOSC4Cancer⁶⁶ will adapt several EOSC-Life solutions for use by the Cancer Mission; and EuroScienceGateway⁶⁷ will continue to develop the tools and software ecosystem initially developed in EOSC-Life together with experts from the earth, environmental and physical sciences.

EOSC-Life has thus helped the life sciences to take significant steps towards turning the EOSC vision into a reality, but much work remains to be done. How can our experiences shape the future development of EOSC and the data environment for Europe's LS RIs? Even if EOSC-Life underlines the importance of close partnerships, and open calls are becoming even more challenging to implement than pre-defined use cases in our experience, these calls are critical for developing and advancing practical applications driven by user needs and for promoting scientific discoveries. Future EOSC developments should build on the capacity of whole-domain projects such as EOSC-Life (and the four other European research clusters), which serve as a nucleus for managing data and connecting interdisciplinary communities. In addition to the INFRASERV projects mentioned above, canSERV⁶⁸ connects experimental facilities that support the analysis of biomedical data in cancer. Other projects are promoting the agroecological transition (AgroServ⁶⁹) and supporting AI-powered image analysis methods (AI4Life⁷⁰). All of these projects build on the tools and experiences supported by EOSC-Life, including RO-Crate, ISA tools⁷¹, FAIRsharing and the FAIR Cookbook. These will continue to populate EOSC with data, workflows and other tools and to sustainably connect different disciplines.

⁶⁵ <https://www.eosc.eu/sria-mar>

⁶⁶ <https://eosc4cancer.eu/>

⁶⁷ <https://galaxyproject.org/projects/esg/>

⁶⁸ <https://www.canserv.eu/>

⁶⁹ <https://emphasis.plant-phenotyping.eu/european-infrastructures/cluster-projects/agroserv>

⁷⁰ <https://ai4life.eurobioimaging.eu>

⁷¹ <https://isa-tools.org/>

Above all, EOSC-Life has given a legacy in the form of the professional, intra- and inter-disciplinary development of competences. The project has helped establish new data management capabilities in several RIs and, by providing training and support, helped build skills in user communities. To be successful, EOSC needs to further develop networks of data managers and skilled data analysts throughout the European research community: Open data and Open Science can only generate value when the people are able to make use of the available opportunities.

Acknowledgements

This research is mainly a product of the EOSC-Life European programme funding from the European Union's Horizon Europe research and innovation programme under grant agreement N°824087.

Complementary support was provided through

- EU funded project AgroServ (grant agreement N°10105 8020),
- EU funded project BY-COVID (grant agreement N°101046203),
- EU funded project DANUBIUS-IP (grant agreement N°101079778),
- EU funded project EMPHASIS-GO (grant agreement N°101079772),
- EU funded project FAIRplus (IMI grant agreement N°802750),
- EU funded project FAIRsharing (Wellcome grant agreement N°212930/Z/18/Z),
- EU funded project ISIDORE (grant agreement N°101046133),
- EU funded project Precision Toxicology (grant agreement N°965406),
- UKRI DASH (grant agreement N°MR/V038966/1).

Special thanks to T. Biro and her radical collaboration team from Research Data Alliance who gave us great inspiration on how to lead this radical collaboration work.

Competing Interests

The authors have no competing interests to declare.

CRedit authorship contribution statement

- Paper Conception: HP, JMB, JKH, NB, PG, RD; Methodology: PG, RD;
- Investigation and meeting discussions: All attendees to the meetings (see rolling notes);
- Figure conception: ARy, JKH, MP, RD;
- Supervision: ARy, JMB, JKH, NB, PG, RD;
- Writing - Original Draft: All authors;
- Figure contributions: ARy, JK, JKH, KE, MP, MTM, RD, SC, US;
- Writing - Review & editing paper: AK, AL, ARo, ARy, BM, BSS, CG, CP, DLL, HH, HP, HVW, JE, JK, JKH, JMB, JWB, KE, KTG, MTM, NB, PG, RD, RM, RP, SAS, SC.

Data Availability statement

CC-BY licence - Data contact:RD. Radical collaboration resources for this paper (PDF) are supplementary material and rolling notes will be available on Zenodo DOI:*after acceptance*;

References

Peer reviewed

Barker M, Chue Hong NP, Katz DS et al., (2022). Introducing the FAIR Principles for research software. *Sci Data* 9, 622 <https://doi.org/10.1038/s41597-022-01710-x>

Bizer C, Heath T, Berners-Lee T, (2009). Linked data - the story so far. *Int. J. Semant. Web Inf. Syst.* 5, 1–22 <http://dx.doi.org/10.4018/jswis.2009081901>

Bourne P, Lorsch J, & Green E, (2015). Perspective: Sustaining the big-data ecosystem. *Nature* 527, S16–S17 <https://doi.org/10.1038/527S16a>

Brack P, Crowther P, Soiland-Reyes S, et al., (2022). Ten simple rules for making a software tool workflow-ready. *PLoS Comput Biol* 18(3): e1009823. <https://doi.org/10.1371/journal.pcbi.1009823>

Cabili MN, Lawson J, Saltzman A., et al., (2021). Empirical validation of an automated approach to data use oversight. *Cell Genom*, 1(2), 100031. <https://doi.org/10.1016/j.xgen.2021.100031>

Chen Q, Britto R, Erill I, et al., (2020). Quality Matters: Biocuration Experts on the Impact of Duplication and Other Data Quality Issues in Biological Databases. *Genomics Proteomics Bioinformatics* 18(2), 91-103. <https://doi.org/10.1016/j.gpb.2018.11.006>

Chue Hong NP, Katz DS, Barker M, et al., FAIR4RS WG (2022). FAIR Principles for Research Software version 1.0 (FAIR4RS Principles v1.0). *Research Data Alliance* <https://doi.org/10.15497/RDA00068>

David R, Richard AS, Connellan C, et al., (2023). Umbrella Data Management Plans to integrate FAIR data : lessons from the ISIDORE and BY-COVID consortia for pandemic preparedness. *Data Science Journal*. IN SECOND REVIEW. Preprint: <https://doi.org/10.5281/zenodo.7998443>

David R, Ohmann C, Boiten JW, et al., (2022). An iterative and interdisciplinary categorisation process towards FAIRer digital resources for sensitive life-sciences data. *Sci Rep* 12, 20989 <https://doi.org/10.1038/s41598-022-25278-z>

David R, Mabile L, Specht A et al., (2020), RDA – SHARing Reward and Credit (SHARC) Interest Group R.D.A., 2020, 'FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles', *Data Science Journal*, 19(1), p.32. DOI: <http://doi.org/10.5334/dsj-2020-032>

European Commission, Directorate-General for Research and Innovation, Schwarzmann U, Fenner M, Hellström M, et al., (2021), PID architecture for the EOSC – Report from the EOSC Executive Board Working Group (WG) Architecture PID Task Force (TF), Publications Office, <https://data.europa.eu/doi/10.2777/525581>

Goble C, Bacall F, Soiland-Reyes S, et al., (2022), 'WorkflowHub – a FAIR registry for workflows'. *F1000Research*, 11. DOI: <https://doi.org/10.7490/f1000research.1118984.1>

Joosten RP, Long F, Murshudov GN, et al., (2014). The PDB_REDO server for macromolecular structure model optimization. *IUCrJ*. 1(Pt 4):213-20 <https://doi.org/10.1107/S2052252514009324>

Karki R, Gadiya Y, Zaliani A, et al., (2023). MpoX Knowledge Graph: A comprehensive representation embedding chemical entities and associated biology of MpoX. *Bioinformatics Advances* (in press)

- Karp PD**, (2016) How much does curation cost? Database (Oxford) 7;2016:baw110
<https://doi.org/10.1093/database/baw110>
- Knapp J, Zeratsky J, Kowitz B**, (2016). Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days. Simon & Schuster. ISBN978-1-5011-2177-7, 288p.
- Khan FZ, Soiland-Reyes S, Sinnott, RO et al.**, (2019). Sharing interoperable workflow provenance: A review of best practices and their practical application in CWLProv. GigaScience 8, giz095.
<https://doi.org/10.1093/gigascience/giz095>
- Kolodkin, AN, Sharma RP, Colangelo AM, et al.**, (2020). 'ROS networks: designs, ageing, Parkinson's disease and precision therapies', *Npj Systems Biology and Applications*, 6.
<https://doi.org/10.1038/s41540-020-00150-w>
- Lamprecht A-L, Garcia L, Kuzak M, et al.**, (2020): Towards FAIR principles for research software. Data Science 3(1) pp. 37–59. <https://doi.org/10.3233/DS-190026>
- Lawson J, Cabili MN, Kerry G., et al.**, (2021). The Data Use Ontology to streamline responsible access to human biomedical datasets Cell Genom Nov 10;1(2):None.
<https://doi.org/10.1016/j.xgen.2021.100028>
- Lin D, Crabtree J, Dillo I, et al.**, (2020), 'The TRUST Principles for digital repositories'. Sci Data 7, 144 (2020). DOI: <https://doi.org/10.1038/s41597-020-0486-7>
- McGovern NY**, (2018a). Radical Collaboration and Research Data Management: An Introduction. Research Library Issues. 296: 6–22. <https://doi.org/10.29242/rli.296.2>
- McGovern NY**, (2018b). Forward Together. Research Library Issues. 2018b; 296(2018): 62–67.
<https://doi.org/10.29242/rli.296.9>
- Moore J, Allan C, Besson, S et al.**, (2021). OME-NGFF: a next-generation file format for expanding bioimaging data-access strategies. Nat Methods 18, 1496–1498.
<https://doi.org/10.1038/s41592-021-01326-w>
- Moreau L, Missier, P** eds. PROV-DM: The PROV Data Model. 30 April 2013, W3C Recommendation. URL: <http://www.w3.org/TR/2013/REC-prov-dm-20130430/>
- Nurnberger A**, (2018) "The Radical Collaboration of RDA and What It Means for Developing Institutional Data Management Services." Research Library Issues, no. 296 23–32.
<https://doi.org/10.29242/rli.296.3>.
- Ohmann C, David R, Abadia MC, et al.**, (2022). Pilot Study on the Intercalibration of a Categorisation System for FAIRer Digital Objects Related to Sensitive Data in the Life Sciences. Data Intelligence 4, 196–211. https://doi.org/10.1162/dint_a_00126
- Papoutsoglou EA, Faria D, Arend D, et al.**, (2020). Enabling reusability of plant phenomic datasets with MIAPPE 1.1. New Phytologist 227, 260–273. <https://doi.org/10.1111/nph.16544>
- Pickering B, Biro T, Austin CC et al.** (2021). Radical collaboration during a global health emergency: development of the RDA COVID-19 data sharing recommendations and guidelines [version 1; peer review: 2 approved, 1 approved with reservations]. Open Res Europe 2021, 1:69
<https://doi.org/10.12688/openreseurope.13369.1>
- RDA COVID-19 WG: RDA COVID-19**, (2020). Recommendations and Guidelines on Data Sharing. (Version 1.0, June 30). Research Data Alliance working group. <https://doi.org/10.15497/rda00052>

- Richard A, Stepanyan D, Ewbank J**, 2022, Pandemic preparedness - Europe launches research consortium. *Nature*. 2022 Mar;603(7900):228. DOI: <https://doi.org/10.1038/d41586-022-00669-4> PMID: 35260850.
- Rocca-Serra, P., Gu, W., Ioannidis, V. et al.** The FAIR Cookbook - the essential resource for and by FAIR doers. *Sci Data* **10**, 292 (2023). <https://doi.org/10.1038/s41597-023-02166-3>
- Ryan, A., Kerppola, M., Verhey-Henke, A. et al.** (undated) How to Conduct Paper Sprints. The University of Michigan School of Public Health. https://sph.umich.edu/cehr/pdf/Paper_Sprint_Manual.pdf
- Sansone, S-A, McQuilton, P, Rocca-Serra, P, et al.**, 2019. 'FAIRsharing as a community approach to standards, repositories and policies'. *Nat Biotechnol*, 37: 358–367. DOI: <https://doi.org/10.1038/s41587-019-0080-8>
- Sarkans, U., Chiu, W., Collinson, L., et al.**, 2021. REMBI: Recommended Metadata for Biological Images—enabling reuse of microscopy data in biology. *Nat Methods* **18**, 1418–1422. <https://doi.org/10.1038/s41592-021-01166-8>
- Scott K**: Radical Candor: How to get what you want by saying what you mean. 1st Edition. Pan Macmillan. 2017. ISBN:9781509845361
- Scott K**: Radical Candor: How to get what you want by saying what you mean. 2nd Edition. Macmillan. 2019. [ISBN1509845356](https://doi.org/10.1038/s41597-023-02166-3)
- Soiland-Reyes, S, Sefton, P, Crosas, M, et al.**, 2022, 'Packaging research artefacts with RO-Crate'. *Data Science* **5** (2). DOI: <https://doi.org/10.3233/ds-210053>
- Sollis E., Mosaku A., Abid A., et al.**, 2023, The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource, *Nucleic Acids Research*, Volume 51, Issue D1, 6 January 2023, Pages D977–D985, <https://doi.org/10.1093/nar/gkac1010>
- Strasser C, Hertweck K, Greenberg J, et al.** (2022) Ten simple rules for funding scientific open source software. *PLoS Comput Biol* **18**(11): e1010627. <https://doi.org/10.1371/journal.pcbi.1010627>
- Tamm JW, Luyet RJ**: Radical Collaboration : Five Essential Skills to Overcome Defensiveness and Build Successful Relationships. 2nd Edition. HarperCollins, 2010. <https://www.oreilly.com/library/view/radical-collaboration/9780062013569/>
- Weise M, Filip Kovacevic F, Popper N, et al.** (2021) OSSDIP: Open Source Secure Data Infrastructure and Processes Supporting Data Visiting. *Data Science Journal* Volume: 21 P4, 202w <https://doi.org/10.5334/dsj-2022-004>
- Wilkinson, MD, Dumontier, M, Aalbersberg, IJ et al.**, 2016. 'The FAIR Guiding Principles for scientific data management and stewardship'. *Scientific Data*, 3: 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>
- Wittner, R., Mascia, C., Gallo, M. et al.**, Lightweight Distributed Provenance Model for Complex Real-World Environments. *Sci Data* **9**, 503 (2022). <https://doi.org/10.1038/s41597-022-01537-6>

Online EOSC and EOSC-Life resources cited [R2]

- Appleton O, Amsi A, Bird I, et al., (2020).** EOSC - a tool for enabling Open Science in Europe. <http://doi.org/10.5281/zenodo.4044010>
- Blomberg N, Sarntivijai S, Mayrhofer MTh., (2020).** 'EOSC-Life Data Management Plan', Zenodo. DOI: <https://doi.org/10.5281/zenodo.4010442>
- Boiten JW, Ohmann C, Adeniran A, et al., (2021a).** EOSC-Life WP4 TOOLBOX: Toolbox for sharing of sensitive data - a concept description. Zenodo. <https://doi.org/10.5281/zenodo.4483694>
- Boiten JW, Ohmann C, Adeniran A, et al., (2021b).** EOSC-Life Guidance and policy on standards and tools to facilitate sharing and reuse of multimodal data (including imaging), cohort integration, and biosamples. <https://doi.org/10.5281/zenodo.4591011>
- ELIXIR (2021)** Research Data Management Kit. A deliverable from the EU-funded **ELIXIR-CONVERGE** project (grant agreement 871075). URL: <https://rdmkit.elixir-europe.org>
- EMDB (the Electron Microscopy Data Bank)** is a public repository for electron cryo-microscopy maps and tomograms of macromolecular complexes and subcellular structures. <https://www.ebi.ac.uk/emdb/>
- EOSC TF sustainability** <https://www.eoscsecretariat.eu/working-groups/sustainability-working-group>
- Goble C, Soiland-Reyes S, Bacall F et al., (2023).** EOSC-Life Implementation of a mechanism for publishing and sharing workflows across instances of the environment. Zenodo. <https://doi.org/10.5281/zenodo.7886545>
- Goble C, Soiland-Reyes S, Bacall F et al., (2021).** Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory. Zenodo. <https://doi.org/10.5281/zenodo.4605654>
- Gribbon P, Parkinson H, Witt G, (2020).** D1.1 EOSC-Life EOSC Cloud feasibility assessment for Demonstrators. Zenodo. <https://doi.org/10.5281/zenodo.4010371>
- Haley N, Leitner F, Bischof J, Audergon, P., (2020).** EOSC-Life Publication of generic guidelines for the organisation of topic-specific Open Calls. Zenodo. <https://doi.org/10.5281/zenodo.4048442>
- Kerfant, C, Ruiz-Pérez, J, García-Granero, J, et al., (2022).** A dataset for assessing phytolith data for implementation of the FAIR data principles. Zenodo. <https://doi.org/10.5281/zenodo.7215804>
- Lamanna G, Bird I, Petzold A et al., (2021).** ESFRI Science Clusters Position Statement on Expectations and Long-Term Commitment in Open Science. <https://doi.org/10.5281/zenodo.4892245>
- Jahn N, Laakso M, Lazzeri E & McQuilton, P., (2023).** Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements (Version 1.0). Zenodo. <https://doi.org/10.5281/zenodo.7728016>
- Nardello I, Asmi A, Buch R, (2022).** Delivering for EOSC - Key Exploitable Results of Horizon 2020 EOSC-related Projects (FULL Report). <https://doi.org/10.5281/zenodo.7401539>
- Parkinson H, Gribbon P, Sarkans U et al., (2021).** EOSC-Life EOSC repository deployment for project demonstrators. <https://doi.org/10.5281/zenodo.4432029>
- Parkinson H, Gribbon P, Sarkans U et al., (2022).** EOSC-Life EOSC FAIR services deployment for open calls. <https://doi.org/10.5281/zenodo.6208249>
- Perseil I, Sansone SA, McQuilton P et al., (2020).** EOSC-Life FAIR Requirements Document. <https://doi.org/10.5281/zenodo.3693883>
- Robertson, Meijer J, Jones B, et al., (2022).** Towards Sustainable Funding Models for the European Open Science Cloud. Financial Sustainability Task Force Progress report <https://www.eosc.eu/sites/default/files/2022-11/financial-sustainability-tf-progress-report-nov-2022.pdf>
- Romano P, Aznar R, López-Coronado JM et al., (2022).** Microbial resources in the European Open Science Cloud: MIRRI contribution to the EOSC-Life project. ECCO XL 2022, Braunschweig, September 27 – 29, 2022. <https://doi.org/10.5281/zenodo.8052705>
- Wilkinson MD, Sansone SA, Grootveld M, et al. (2022).** FAIR Assessment Tools: Towards an "Apples to Apples" Comparisons. Zenodo. <https://doi.org/10.5281/zenodo.7463421>
- Zafeiropoulos H, Beracochea M, Ninidakis S, et al. (2023).** metaGOflow: a workflow for the analysis of marine Genomic Observatories shotgun metagenomics data - use case (v1.0.0) [dataset]. Zenodo. <https://doi.org/10.5281/zenodo.7771821>

Other resources are available at: <https://zenodo.org/communities/eosc-life/>

Initials and Acronyms

- AAI: Authentication and Authorization Infrastructure
- API: Application Programming Interfaces
- BBMRI: Biobanking and Biomolecular Resources Research Infrastructure
- CORBEL: Coordinated Research Infrastructures Building Enduring Life-science Services
- CPM: Common Provenance Model
- CWL: Common Workflow Language
- DOI: Digital Object Identifier
- EATRIS: European infrastructure for translational medicine (European Advanced Translational Research Infrastructure in Medicine)
- EBI: European Bioinformatics Institute
- EC: European Commission
- ECRIN: European Clinical Research Infrastructure Network
- EMBL: European Molecular Biology Laboratory
- EMBRC: European Marine Biological Resource Centre
- EMPHASIS: Effective management of pests and harmful alien species - Integrated solutions
- ERIC: European Research Infrastructure Consortium
- ERINHA: European Research Infrastructure on Highly Pathogenic Agents
- EOSC: European Open Science Cloud
- ESFRI: European Strategy Forum Research Infrastructure
- FAIR: Findable, Accessible, Interoperable and Reusable
- GA4GH: Global Alliance for Genomics and Health
- GDPR: General Data Protection Regulation
- GSO: Group of Senior Officials on global Research Infrastructures
- ISBE: Infrastructure for Systems Biology in Europe
- JSON-LD: JavaScript Object Notation for Linked Data
- LS: Life Sciences
- MIAPPE: Minimum Information About a Plant Phenotyping Experiment
- MIRRI: Microbial Resource Research Infrastructure
- Mpox: Monkeypox
- NEUBIAS: Network of EUropean BioImage Analysts
- NHGRI: National Human Genome Research Institute
- OECD: Organisation for Economic Co-operation and Development
- OLS: Ontology Lookup Service
- OME-NGFF: Open Microscopy Environment Next-Generation File Formats
- OSSDIP: Open Source Secure Data Infrastructure and Processes Platform
- OxO: Ontology Cross Reference Service
- PIDs: Persistent and unique IDentifierS
- RDA: Research Data Alliance
- RDM: Research Data Management
- REMBI: Recommended Metadata for Biological Images
- RI: Research Infrastructure
- SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2
- SRIA MAR: Strategic Research & Innovation Agenda and its Multi-Annual Roadmap
- TRL: Technology Readiness Levels
- TRUST: Transparency, Responsibility, User Focus, Sustainability, Technology
- WG: Working Group
- WP: Work Package

Links in footnotes:

Link name	Link address	In text	Last access
ABS Permits	https://oneworldanalytics.com/abspermits/	81	10/06/2023
AgroServ	https://emphasis.plant-phenotyping.eu/european-infrastructure/s/cluster-projects/agroserv	69	09/06/2023
AI4Life	https://ai4life.eurobioimaging.eu	70	09/06/2023
Amsterdam Declaration on Funding Research Software Sustainability Draft	https://future-of-research-software.org/draft-amsterdam-declaration-on-funding-research-software-sustainability/	1	09/06/2023
BBMRI	https://www.bbmri-eric.eu/	20	09/06/2023
BioContainers	https://biocontainers.pro/	42	09/06/2023
BIOINDUSTRY 4.0	https://cordis.europa.eu/project/id/101094287	36	09/06/2023
Biological systems modelling JWS Online	https://jjj.biochem.sun.ac.za/	32	09/06/2023
Biomodels	https://www.ebi.ac.uk/biomodels/	33	09/06/2023
Biotracks	https://github.com/CellMigStandOrg/biotracks	63	09/06/2023
BY-COVID	https://by-covid.org/	35	09/06/2023
CanSERV	https://www.canserv.eu/	68	09/06/2023
Codebases General Guide For Exploring Large Open Source Codebases	https://pncmnp.github.io/blogs/oss-guide.html	39	09/06/2023
Conda	https://docs.conda.io/en/latest/	45	09/06/2023
Convention on Biological Diversity (CBD) - Recent agreements	https://www.cbd.int/doc/c/e6d3/cd1d/daf663719a03902a9b116c34/cop-15-l-25-en.pdf	82	09/06/2023
COVID-19 data portal	https://www.covid19dataportal.org/	17	09/06/2023
DockerHub	https://hub.docker.com/	44	09/06/2023
EATRIS	https://eatris.eu/	51	09/06/2023
ECRIN	https://ecrin.org/	50	09/06/2023
ELIXIR	https://elixir-europe.org/	18	09/06/2023
ELIXIR's Long-term sustainability plan (2019)	https://f1000research.com/documents/8-1642	19	09/06/2023
EMBRIC	https://www.embric.eu/	52	09/06/2023
EMO BON	https://www.embric.eu/emo-bon	79	09/06/2023
EMPHASIS	https://www.emphasisproject.eu/	73	09/06/2023
ENVRI-FAIR	https://envri.eu/home-envri-fair/	5	09/06/2023
EOSC4CANCER	https://eosc4cancer.eu/	66	09/06/2023
EOSC-Life	https://www.eosc-life.eu/	6	09/06/2023
EOSC-Life Calls	https://www.eosc-life.eu/calls/	10	09/06/2023
EOSC-Life courses websites	https://www.eosc-life.eu/news/training-modelling-covid-19-epidemics/	59	09/06/2023
EOSC-Life Internal call for academia-Industry	https://www.eosc-life.eu/industrycall/	75	09/06/2023
EOSC-Life Open Call training	https://www.eosc-life.eu/services/open-call-training/	57	09/06/2023

EOSC-Life Training	https://www.eosc-life.eu/services/training/	57	09/06/2023
EOSC SRIA MAR	https://www.eosc.eu/sria-mar	65	09/06/2023
EOSC Task Forces	https://eosc.eu/eosc-task-forces	4	09/06/2023
EOSC Task Force PIDs policy and Implementation	https://www.eosc.eu/advisory-groups/pid-policy-implementation	30	09/06/2023
ERINHA	www.erinha.eu	49	09/06/2023
ESCAPE	https://projectescape.eu/	7	09/06/2023
EU-OPENSREEN	https://www.eu-openscreen.eu/	21	09/06/2023
Euro-Biolmaging	https://www.eurobioimaging.eu/	22	09/06/2023
Euro-Biolmaging guide to FAIR	https://www.eurobioimaging.eu/news/euro-bioimaging-guide-to-fair-bioimage-data/	56	09/06/2023
EuroScienceGateway	https://galaxyproject.org/projects/esg/	67	09/06/2023
FAIR Cookbook	https://faircookbook.elixir-europe.org	25	09/06/2023
FAIR Phytolith project	https://open-phytoliths.github.io/FAIR-phytoliths/	77	09/06/2023
FAIRsharing	https://fairsharing.org/	34	09/06/2023
FAIRsharing Community champions program	https://fairsharing.org/community_champions	83	09/06/2023
FAIRsharing educational material	https://fairsharing.org/educational	84	09/06/2023
Fragalysis	https://fragalysis.diamond.ac.uk/viewer/react/landing/	76	09/06/2023
GDPR	https://gdpr-info.eu/	3	09/06/2023
GitHub	https://github.com/	40	09/06/2023
GSO	http://www.gsogri.org/	12	09/06/2023
GWAS studies	https://www.ebi.ac.uk/gwas/docs/methods/summary-statistics	14	09/06/2023
INFRAFRONTIER	https://www.infrafrontier.eu/	78	09/06/2023
INFRASERV	https://rea.ec.europa.eu/funding-and-grants/horizon-europe-research-infrastructures/research-infrastructure-services-support-health-research-accelerate-green-and-digital-transformation_en	61	09/06/2023
INSTRUCT	https://instruct-eric.org/	23	09/06/2023
ISA Tools	https://isa-tools.org/	71	09/06/2023
ISIDORE	https://isidore-project.eu/	62	09/06/2023
ISO 23494	https://www.iso.org/standard/80715.html	37	09/06/2023
Life-Monitor	https://app.lifemonitor.eu/	28	09/06/2023
Linked Data	https://www.w3.org/DesignIssues/LinkedData	60	09/06/2023
Long-Term Sustainability of Research Infrastructures (ESFRI Report)	https://www.esfri.eu/sites/default/files/u4/ESFRI_SCRIPTA_VO_L2_web.pdf	13	09/06/2023
Memorandum of Understanding (MoU) with EuroBioimaging and Instruct-ERIC	https://www.eu-openscreen.eu/newsroom/eu-openscreen-news/ansicht/eu-openscreen-has-signed-a-memorandum-of-understanding-mou-with-eurobioimaging-and-instruct-eric.html	24	09/06/2023
MGNify	https://docs.mgnify.org/src/docs/analysis.html	80	09/06/2023
MIAPPE	https://www.miappe.org/	15	09/06/2023
Microscopy data analysis	https://www.ebi.ac.uk/training/events/microscopy-data-analysis	56	09/06/2023

	-0/		
Microscopy data analysis: Machine Learning and Bioimage archive	https://www.ebi.ac.uk/training/events/microscopy-data-analysis-machine-learning-and-bioimage-archive-2022/	57	09/06/2023
MIRRI	https://www.mirri.org/	74	09/06/2023
Nagoya Protocol on Access and Benefit-sharing	https://www.cbd.int/abs/	47	09/06/2023
NAR	https://www.oxfordjournals.org/nar/database/c/	2	09/06/2023
NEUBIAS	https://eubias.org/NEUBIAS/venue/	55	09/06/2023
NEUBIAS Training school	https://www.eosc-life.eu/news/neubias-defragmentation-training-school/	54	09/06/2023
Ontology Cross Reference Service (OxO)	https://www.ebi.ac.uk/spot/oxo/	31	09/06/2023
Ontology Lookup Service (OLS)	https://www.ebi.ac.uk/ols4	29	09/06/2023
Open Microscopy Environment Next-Generation File Formats (OME-NGFF)	https://ngff.openmicroscopy.org	64	09/06/2023
OSSDIP	https://www.ifs.tuwien.ac.at/infrastructures/ossdip/	48	09/06/2023
PaNOSC	https://www.panosc.eu/	8	09/06/2023
Permissive Licences	https://fossa.com/blog/all-about-permissive-licenses/	38	09/06/2023
Quay.io	https://quay.io/	41	09/06/2023
RDMKit	https://rdmkit.elixir-europe.org/	26	09/06/2023
Recommendation of the Council on Enhancing Access to and Sharing of Data	https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463	11	09/06/2023
Sourceforge	https://sourceforge.net/	46	09/06/2023
SSHOC	https://sshopencloud.eu/	9	09/06/2023
TravisCI	https://www.travis-ci.com/	43	09/06/2023
World-cafe	https://theworldcafe.com/key-concepts-resources/world-cafe-method/	16	09/06/2023
WorkflowHub	https://workflowhub.eu/	27	10/09/2023

Supplementary information

Supplementary_Information_S1 Radical collaboration® method for analysis of sustainability strategies employed in EOSC-Life and identification of best practice, including preparation of this manuscript

We have applied a set of theories and concepts, termed radical collaboration, to identify key elements that support the sustainability of the work and results produced within the framework of our inter-RI project. The choice of taking a radical collaboration approach was made because it intrinsically supports processes involving skilled people working in a collaborative environment. The methodology of radical collaboration was initially established for the purposes of conflict resolution and mediation (Tamm & Luyet, 2010). Further extending the range of application McGovern (2018a, p.6) states, “The concept of radical collaboration means coming together across disparate, but engaged, domains in ways that are often unfamiliar or possibly uncomfortable to member organisations and individuals to identify and solve problems together, to achieve more together than we could separately.” The reuse of radical collaboration as an operational framework in research was also suggested by McGovern (2018b) in a second contribution in that issue of *Research Library Issues*, based on an earlier proposition by Scott (2017 & 2019).

In 2018, Ruttenberg and Waraksa brought out a special issue of *Research Library Issues* that placed a special emphasis on the development of a collaborative thesis for RDM communities. In parallel, Nurnberger (2018) described how the radical collaboration methodology was applied to the experience of forming and maintaining Research Data Alliance (RDA) groups, offering specific recommendations for developing sustainable institutional and inter-RIs RDM services.

When considering issues related to the sustainability of RDM, it is important to recognise that these matters and the underlying research are strongly influenced by the temporary nature of research projects and associated funding. These issues have been put into focus by the publication of the FAIR principles (Wilkinson et al., 2016), which provided a framework for driving reuse of data and software resources. Achieving reuse of scientific products of large collaborative projects, involving many partners, has led to the creation of vital forums and international assemblies, such as the Research Data Alliance (RDA), GO-FAIR initiative, European Open Science Cloud (EOSC), World Data System (WDS), and Committee on Data (CODATA) of the International Science Council (ISC). These forums and assemblies typically aim to both invest in new resources as well as increase the sustainability of outcomes and the reuse of the deliverables and tools resulting from inter-organisation projects.

In her first paper, McGovern (2018a) developed three working concepts that can be used “to guide the process of applying the radical collaboration methodology: inclusive community, distributed digital practices, and productive and sustainable collaboration” [R11]. We used this theoretical framework in our study to assess the durability and sustainability of EOSC-Life outputs. We subsequently applied the methodology shown by Pickering et al, in 2020 in the development of the present paper. Specifically, we asked the project participants to share their perceptions about the sustainability of the work they carried out on their tasks in specifically focused meetings run as sprint events. The first steps were about developing a

detailed plan with willingness to make inputs. The team organised the first meeting to present the focus of the paper, define methodology, and recruit skilled people with a good representation of the different EOSC-Life working teams and RIs. 13 meetings were organised then we run 6 Paper Sprints with iterative invitations through EOSC-Life community for a total of 28 open meetings (integrating also EOSC-Life partners recruited with call for projects).

Jake Knapp, a Google employee presented in 2016 sprints as a method to enable project teams to develop and prototype new products rapidly. These sprints were later adapted to apply the approach in academia, and a helpful paper sprint manual was developed by the University of Michigan School of Public Health Center (Ryan et al., undated). The paper sprint method was ideal for our purpose, as it enabled sections of the research paper to be written mutually and simultaneously by multiple members of a research team. Such a similar collaborative method is also applied to generate content, at a good rate, for RMDKit.

In the first paper sprint, the team wrote the case studies parts with the participants' perspectives, identified relevant examples for sustainability, and planned the next sprint steps. In the 2 secondary sprint(s), we met online to write the remaining portions of the paper based on the results of meetings and exchanges. The final sprints were performed to reorganise the paper and to shorten longer parts.

We focussed on gathering the participants' perspectives on the efficacy of the capacity building process, such as identifying training needs and carrying out the open call process. The capacity building process was implemented to engage and support the development and extension of FAIR resources and services both inside and outside of the EOSC-Life consortium. One point of particular interest was examining the participants' perceptions of how FAIR components were integrated in projects funded by EOSC-Life involving the 13 RIs^{72, 73, 74} how these projects were managed, and how sustainable the project outcomes were. We summarised the FAIR components, including reusable elements of the research data management plan (DMP), Authentication and Authorisation Infrastructures services (AAls), data repositories, data feeds i.e. how data is served in data federations, and data access governance e.g. sensitive data aspects. This summary was prepared by using community-approved terminology and by referring to provenance information. Once the paper sprint process completed, the entire manuscript was subjected to a thorough and iterative review by the involved authors.

Supplementary_Information_S2 Collaborative User projects

The first tranche involved a set of seven demonstrators drawn from the RIs community, placing a focus on achieving sustainable, interoperable collaboration between two or more research infrastructures. The underlying aims were to share experience between resources, to make data more interoperable, and to evaluate sustainability strategies, development of shared processes, improving portability of data, analysis and resources and code.

⁷² ELIXIR, BBMRI-ERIC, EATRIS-ERIC, ECRIN-ERIC, EMBRC, EMPHASIS, ERINHA, EU-OPENSOURCE, Euro-BioImaging, INFRAFRONTIER, Instruct-ERIC, ISBE, MIRRI

⁷³ <https://www.emphasisproject.eu/>

⁷⁴ <https://www.mirri.org/>

Subsequent open calls involved three sub-call topics: i) digital LSs (broad application), ii) sensitive data, and iii) academic-industry collaboration. The aim was to support projects that would populate EOSC with FAIR LS data, cloud-based tools, resources, and workflows from a broad range of LS domains.

To support the long-term sustainability of EOSC, all stakeholders in the scientific community including SMEs and industry have to be engaged, realise the potential of/benefit from the collaborative cloud ecosystem. EOSC-Life therefore explored ways of collaboration between public and commercial entities for the development and integration of cloud-based solutions. After consultation with industry representatives on an appropriate framework for applications and projects, EOSC-Life launched a dedicated Internal Call for Academia-Industry⁷⁵ collaborations that resulted in support for two feasible projects. These projects involved consultations with EOSC-Life experts as well as the implementation of EOSC-Life resources.

In the digital LSs call, a new online platform, fragalysis⁷⁶, was developed enabling rapid access to data from fragment screens in a collaborative environment. This innovative platform had a major impact in COVID-19 related work by allowing fast release of fragment screens to the worldwide community. From a community aspect, the FAIR Phytoliths⁷⁷ project developed a new ontology and a set of domain-specific FAIR guidelines for a community which has historically poor data sharing practices (Kerfant et al., 2022). Such efficient cross-discipline collaboration beyond the traditional ESFRI clusters has been driven and supported by integrating projects in EOSC-Life that extend into other sectors.

The project *"Integrating several EU-RI datasets with focus on preclinical and discovery research bioimaging"* (Ref PID 14176) demonstrates a collaborative effort of three RIs in the field of biomedical imaging (Euro-BioImaging), animal model strains (INFRAFRONTIER⁷⁸), and biological chemistry (EU-OPENSOURCE). The three RIs joined forces and combined their resources by building novel, open-source tools for integrating and aligning information associated with data belonging to their infrastructures. Another set of examples from open call projects originated from collaborative work among the RIs Euro-BioImaging, ELIXIR, and ISBE. These projects have led to the more thorough integration of data from imaging, the scientific literature, and modelling, and have provided (and exemplified) more validated mathematical models as live data repositories for mammalian cell biology and ageing (Kolodkin et al., 2020).

To highlight the sustainability aspects faced by the various open call projects, we take a closer look at one project, MetaGOflow, first as one of the seven demonstrators then as an Open Call. MetaGOflow (Zafeiropoulos et al., 2023), a workflow for marine genomics observatory data analysis, was built by a team from EMBRC to analyse the eDNA data from its EMO BON⁷⁹ projects that are used to study European marine biodiversity. MetaGOflow is based on an existing platform, MGnify⁸⁰, and provides taxonomic inventories and functional analyses of use to LS researchers in the field of marine biology. It can also be used by researchers in other fields of biology. MetaGOflow promoted sustainability through continued

⁷⁵ <https://www.eosc-life.eu/industrycall/>

⁷⁶ <https://fragalysis.diamond.ac.uk/viewer/react/landing/>

⁷⁷ <https://open-phytoliths.github.io/FAIR-phytoliths/>

⁷⁸ <https://www.infrafrontier.eu/>

⁷⁹ <https://www.embrc.eu/emo-bon>

⁸⁰ <https://docs.mgnify.org/src/docs/analysis.html>

re-use by enhancing the efforts of EMO BON: the eDNA data collected by EMO BON are now turned into scientific results that can be used for academic research and for environmental reporting. The data (raw and analysed) become more interesting to a wider audience and are more frequently used, increasing the chance that they will continue to be provided (organisational and financial sustainability). As the workflow produces full provenance for all its inputs and outputs, the data are more likely to be (i) trusted and (ii) re-used, again promoting its sustainability.

Supplementary_Information_S3 The COVID-19 data portal

During the COVID-19 pandemic, the importance of sustaining tools, services, best practices, and guidelines for sensitive data management was clearly demonstrated. In a crisis of this nature, rapid access to accurate data can enable researchers to assess the severity, spread, and impact of a pandemic and result in the implementation of efficient, effective response strategies. In their recommendations for tackling COVID-19, the Research Data Alliance (RDA) clearly states that data should be deposited in repositories to facilitate quality control, timely sharing, and sustainable access. Whenever possible, these should be trustworthy data repositories that have been certified, are subject to rigorous governance, and are committed to promoting the long-term sustainability of their data holdings (RDA, 2020).

The COVID-19 data portal established in EOSC-Life, contains over 16 million raw and assembled sequences and analysis of SARS-CoV-2 and other coronaviruses. FAIR interoperability standards represent key elements that ensure the sustainability of these resources. Due to the urgent need met by the resources, mid-term sustainability has been achieved through the follow-up funding of the BY-COVID project.

Supplementary_Information_S4 The FAIR Cookbook as a resource to support sustainable FAIR implementation

The FAIR Cookbook (Rocca-Serra et al, 2023) was created and is managed by professionals who routinely work with data in academia, (bio)pharmaceutical companies and information service industries. It is an open and collaborative resource, which documents the fundamental processes and capabilities for provisioning FAIR data and services, and provides real examples and use cases by data-producing projects and organisations. Recommended by Horizon Europe, and anchored to several ELIXIR Nodes for long term sustainability, the FAIR Cookbook serves as practical guidance to improve every-day tasks, and contributes to a curriculum on FAIR data, informing discussions around the necessary changes to deliver FAIR within organisations. The success of the FAIR Cookbook has been to cultivate the collective knowledge, in academia and industry, to timely deliver specialised content, which fills the glaring gap between high-level FAIR Principles and their actual implementation in the LSs.

Supplementary_Information_S5 Ontology services extended and supported by EOSC-Life

The Ontology Lookup Service (OLS) is an ontology discovery and access service containing to date 242 biomedical ontologies, 7,790,912 terms, 42,514 properties, and 22,653 individuals. OLS features a web interface that scientists can use to search for and visualise ontologies as well as a highly accessed API. EOSC-Life has resulted in over 35 additional

ontologies being made available in OLS, representing user needs. Ontologies are continually updated, and new ontologies are being added once they mature to the point that they meet users' needs. The aligned resource, the Ontology Cross Reference Service (OxO), provides cross-ontology mapping service allowing users to integrate datasets that are annotated with different ontology terms, a common use case when working data integration across datasets, projects, or biomedical domains. Ontologies in OLS are used to deliver cross references in OxO offering a convenient way for users to access cross references. Zooma is an ontology used for metadata mapping applications. This allows ontologies to be applied to text by users such as biomedical curators and data annotators.

Supplementary_Information_S6 RO-Crate and LifeMonitor, EOSC-Life tools for sustainable workflows

RO-Crate allows researchers to package and aggregate research artefacts with their metadata and relationships, as well as readily package and exchange workflows, including the needed information regarding provenance. RO-Crate was used in the context of the case study MetaGOFlow. The workflow outputs are, in that study, shared through MGnify and additionally packaged as RO-Crate on the EMO BON Github repository: the data inputs, data outputs, and the full provenance (meta)data. Sustainability through continued re-use as well as organisational sustainability are guaranteed for the foreseeable future.

Building on this, EOSC-Life developed the LifeMonitor, a service that facilitates the maintenance of computational processes and supports their reusability over time through periodic testing and test monitoring. Monitored workflows are less likely to break down over time and consequently are more sustainable. The workflows are registered in WorkflowHub.

Supplementary_Information_S7 Case study on issues related to compliance when gathering digital sequence data

One case study involves the Nagoya Protocol on Access and Benefit-sharing. To gather biological samples in the field to obtain their genetic material, scientists are required to comply with the ABS requirements (where, what, and why they are collecting). They must negotiate the terms by which any benefits that might arise from utilising the genetic material will be shared. The physical samples and derived results must be linked to these permits. At present, the ABS requirements are limited to physical samples: digital sequence information (DSI) are not in scope and so ABS permits⁸¹ are not relevant for the sustained use of archived sequences derived from such sampled material. However, discussions to extend the scope to include DSI have been ongoing, and the outcomes of those will have an impact on how scientists and archives deal with allowing the DSI data to be (legally) sustainably accessible and re-usable. Recent agreements⁸² to develop a multilateral system for benefit sharing from DSI, will increase the burden on the sharing of DSI while at least still recognising the benefits of such DSI for biodiversity conservation. Given the significant impact the convention's decisions will have on the sustainable archiving and sharing of DSI, all parties including scientists, biobanks, and data portals will do their best to link their digital data to the ABS permits of the genetic resources from where those data were obtained. This will allow for sustainability by allowing for (legally) correct data re-use.

⁸¹ <https://oneworldanalytics.com/abspermits/>

⁸² <https://www.cbd.int/doc/c/e6d3/cd1d/daf663719a03902a9b116c34/cop-15-l-25-en.pdf>

Supplementary_Information_S8 Sensitive data toolbox

The toolbox provides links to recommendations, procedures, and best practices, as well as to software tools that support data sharing and reuse. The concept of the toolbox has been submitted to Zenodo (Boiten et al., 2021a & 2021b) and subsequently a prototype has been implemented with a dedicated tagging system. Both the evaluation of the initial tagging system (Ohmann et al., 2022) and the development of the full version (David et al., 2022) have been published.

A considerable degree of dissemination was achieved by the publication and it has been downloaded more than 3000 times, in the 98th percentile of the 435,336 tracked articles of a similar age in all journals. Nevertheless, the toolbox will only become a useful resource for biomedical researchers when the sustainability beyond the project's lifetime can be assured. Two high-level models for sustainability are being evaluated: a) An organisation takes full responsibility (model for the EBI databases), or b) a community takes responsibility, either funded by future grants or through "in kind" contributions from intrinsically motivated research parties/individuals (as is happening in various open-source communities). In either model, not only the technical and operational support of the toolbox should be sustained, but also an editorial board securing the actuality and correctness of the content.

While the joint RI communities in EOSC-Life clearly showed interest in using and maintaining the toolbox, the lack of clarity on its long-term sustainability leads to a reluctance to seriously communicate and disseminate it among the user communities [2]. Similarly, RIs tend to be reluctant to augment the contents of the toolbox to the level required to serve as an authoritative resource. To avoid this catch-22, a group of novice users (i.e. users not exposed to the toolbox before) recruited across the participating RIs will test the toolbox and propose missing content. Any content deemed broadly usable will be added by the editorial board before the end of the project.

Supplementary_Information_S9 The FAIRsharing catalogue

FAIRsharing promotes the FAIR Principles by promoting the value and use of data and metadata standards, and their use by databases. For this reason, the FAIRsharing API serves a growing number of tools and services that use it for look-up, selection and content retrieval of standards and repositories in the: (i) creation of data management plans, (ii) enrichment of guidance and training material, and (iii) assessment of and assistance with FAIRness, work in progress under the EOSC FAIR Metrics and Data Quality Task Force (Wilkinson et al, 2022). Cultivating the collective knowledge, engaging the community, and becoming a core element of the ecosystem are the key essentials for the success and long-term sustainability of any resources. To address this challenge, FAIRsharing has also launched the Community Champion Programme⁸³, a thriving community of domain and discipline experts, which also includes members of EOSC-Life, who: (i) act as advocates to promote the value of standards, databases and policies for digital objects (incl. data, software); (ii) create educational material⁸⁴ describing these resources helping researchers and other stakeholders to find, use and adopt them, and (iii) enrich the content of FAIRsharing, adding and enhancing the description and discoverability of these resources.

⁸³ https://fairsharing.org/community_champions

⁸⁴ <https://fairsharing.org/educational>

Supplementary_Information_S10 Implementation of Open call training programs to build sustainability

Development of reusable training resources to support sustainability of EOSC-Life project outcomes has been one of the high priorities for Open and Internal Calls projects. A step-by-step training and procedures for setting up the framework to build and manage an ontology was produced while developing the first Phytolith ontology with the help of EOSC-Life experts. The materials could be used as orienting and guiding help for other communities aiming to create ontology for their scientific area. Within this collaborative training framework a series of Open Research skills workshops was run through support by EOSC-Life Training Open Calls. Specifically, EOSC-Life experts were brought together to co-create two workshops on standard vocabularies, ontologies and FAIR data. These efforts enabled the project to go above and beyond the goals initially set by the project and to deliver an ontology using sustainable and portable processes used by the community. (Publication in preparation).

Supplementary_Information_S11 Applicability of the EOSC-Life learnings for the sustainability of the ISIDORE project and other connected communities

ISIDORE is the broadest of the current INFRASERV programmes led by the European Research Infrastructure on Highly Pathogenic Agents (ERINHA), bringing together 17 research infrastructure and networks, providing services covering all experimental approaches, and also including regulatory advice and the social sciences. The services are offered under the transnational access model, with funding being provided to the research infrastructure and networks that are then able to offer free-of-charge access to their services and resources to researchers whose applications pass the competitive external review. One can distinguish (meta)data related to the service per se, and that related to any (meta)data produced as part of a scientific service during project implementation. For the former, the ISIDORE consortium invested significant effort into constituting a homogeneous catalogue of services with a harmonised metadata scheme for all of the more than 300 services offered. Furthermore, the scheme was chosen in consultation with the relevant experts involved in the BY-COVID project to ensure sustainability and interoperability. While most of the LS RIs have the possibility to incorporate their well established internal organisational service provision and access management procedures and operational guidelines into the project, or some of the participating networks in the consortium, the ISIDORE project initiated the first occasion to formalise and harmonise the description of their services in a systematic manner. Facilitating the rationalisation of the entire catalogue e.g. reducing duplications and having a common metadata scheme in place will ensure that in the future, new services will necessarily be described using a common approach. The issues regarding (meta)data produced as part of a scientific service provided to an incoming research project are, however, more complicated. In part, this arises from the evolutionary fact that the initial goal of many service providers has been to primarily focus on their responsibility in providing open access to scientific services and resources to users for generating scientific data. Targeted execution of FAIR principles and integration of domain specific mandatory data management plans is currently being adopted as an additional layer of responsibilities - which is also very much driven and facilitated by RIs. In ISIDORE, for example, there is no consortium-wide user's charter detailing how users are expected, or required, to make their data FAIR, nor are instructions given to explain how sustainable data objects and data

management components provided by EOSC/EOSC-Life can be used, while only a part of participating RIs are having a Data Management Plan (DMP) in place which helps the users making their data collected during the project FAIR and bring it to open access data repositories in the respective scientific domain. This is at odds with the requirements specified for Horizon Europe, that beneficiaries should at least prepare a DMP, deposit data in a trusted repository and provide open access to it ('as open as possible, as closed as necessary'), as well as provide information about any research output or any other tools and instruments needed to re-use or validate the data. A programme like ISIDORE partially provides a loop-hole regarding openness of data acquired using Horizon Europe funding. For some users, data deposition can be a challenging process. With their collective expertise, the RIs are in a unique position to help users fulfil the EC's requirements and provide substantial efforts to document their challenges and associated processes (David et al., 2023). Future efforts will for instance build on models such as facilitating the submission of sequence data to the European Nucleotide Archive. This has been proposed by ELIXIR with the support of BY-COVID, a FAIR EOSC-related project. Having such tools in place would then make adherence to open data rules much easier for users, so that this could then reasonably become a condition of service provision. This will, however, require broader awareness of the issues, as well as the provision of training and support, before and during the execution on research projects supported by ISIDORE, or other similar programmes. Working collectively on these issues within ISIDORE, in collaboration with BY-COVID using outputs from EOSC-Life should allow universal adoption of common and sustainable methods to allow RIs to facilitate data deposition by users in a form that is most amenable to future access by interested third parties. For ISIDORE partners, a real inclusive and efficient adoption of sustainable components for FAIR data sharing will depend on sustainable support from expert personnel and continuous training to leverage FAIR data sharing skills and literacy (David et al., 2020).

These strategies rely typically on a combination of income generated by the users' fees, public funding at the European and national level (the latter often directed to the national nodes and/or to support access to the RI services by national researchers), provision of services to third parties such as companies or other research institutions. Subscription fees are used to support central services e.g. the hub.