EOSC-Pillar

Coordination and Harmonisation of National &Thematic Initiatives to support EOSC

Final Report

Building an EOSC from National Contributions

Authors

Rob Carrillo, Trust-IT Services
Federico Drago, Trust-IT Services
Maria Giuffrida, Trust-IT Services

EOSC-Pillar Final Report

The EOSC-Pillar Annual Report is a publication of the EOSC-Pillar project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 857650 and is produced to showcase the major results and achievements of the project, collaborations ongoing with other initiatives and updates for the wider community.

Table of contents

Foreword

In this final Annual Report of the project, we will review the results that have been presented at our EOSC-Pillar Final Conference: Building an EOSC from National Contributions, held on 25-27 October 2022 in Paris. With the end of the EOSC-Pillar project and as we look back, we should be proud of our achievements. I thank all the partners which have been contributing to these results these past few years.

Together, we have brought forward several achievements such as technical ones - the Federated FAIR Data Space, the use cases, and the training platform. We also have policy achievements, which I think we should also invest in further in the future. This includes the national initiatives that have been consolidated in the project and have been federated in a common effort to bring the results of national activities to the European landscape. We did very much together in this respect, such as surveys and the legal policy analysis. In addition, the guidelines and recommendations developed in the project are very relevant and will be useful for other European projects succeeding EOSC-Pillar. All in all, we should be proud that we have built a federation of initiatives and developed transnational access.

The idea of building this project came to us because we were convinced that the work that was going on at a national level had to be harmonised in a common effort to federate all the activities in the European landscape. Overall, I think that this regional approach has been successful and in the near future, we should invest in similar national initiatives.

The end of EOSC-Pillar is not the end of the activities of course. We have a legacy which will be useful for other projects. Just as an example, in the Horizon Europe project Skills4EOSC, while concentrated in training and similar activities, the approach of federated national activities is the same as what we have done in EOSC-Pillar. We are proud that the activity and ideas developed in EOSC-Pillar will also be used in the next set of projects that will put forward new ideas and initiatives to realise what we think is the EOSC infrastructure that we need to set up at the European level.

With our experience in EOSC-Pillar, we can see that setting up a European infrastructure should be linked to national activities - this is the most important lesson that we learned in our project.

Once again, I want to thank all the partners which have contributed so much to the final results and success of EOSC-Pillar. Thank you all!

Claudia Battista

GARR Director, EOSC-Pillar Project Coordinator

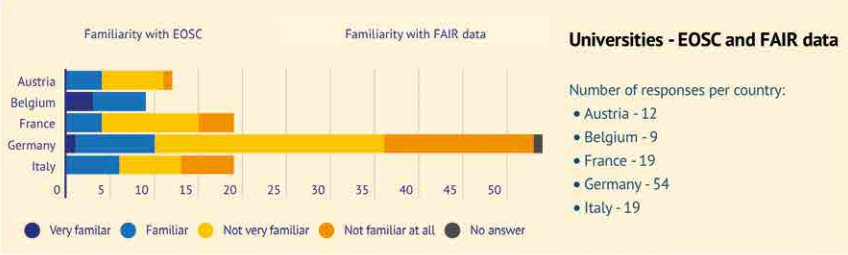
National initiatives survey and advancing monitoring practices in EOSC

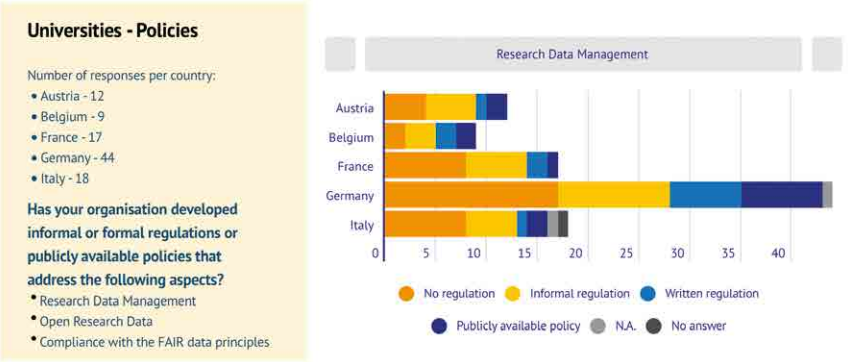
EOSC-Pillar designed and conducted a set of surveys to assess the state of the art of the Open Science landscapes in Austria, Belgium, France, Germany, and Italy, involving about 2,000 institutions across the five EU Member States. The result of this activity laid the foundations of the work within the project as well as contributed to the pan-European landscaping exercise that was happening across various European regions.

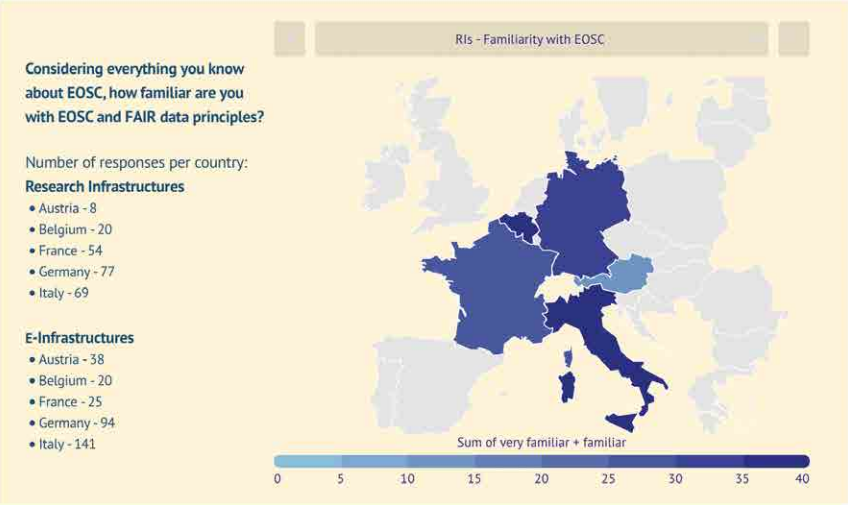
Institutions were identified based on their roles and activities with regard to the European Open Science Cloud (EOSC). The questionnaires were addressed to four types of respondents: funding bodies, universities, research infrastructures, e-infrastructures. To discover and assess gaps, initiate necessary developments, start specific support measures and create opportunities, the survey data available was the key factor for open research data and services. It also provided critical insights for the activities of EOSC-Pillar's policy work and provided input to what would become the EOSC observatory.

EOSC-Pillar's surveying work and experience also provided further insights on how EOSC might have access to reliable and easily accessible data on infrastructures and stakeholders. The Facilitating Data on EOSC session at the EOSC-Pillar Final Conference looked at surveying practices within the community, the challenges on gathering data and providing some insights on how to ensure the success of surveying activities.

See the full presentations from the EOSC-Pillar Final Conference session, Facilitating Data on EOSC: https://www.eosc-pillar.eu/events/final-event/facilitating-data-on-eosc







Open Science Policies for EOSC readiness

One of the areas of EOSC-Pillar's work was on relevant policies and their effectiveness at promoting or hindering the implementation of Open Science and EOSC.

With a stronger drive towards open science across Europe, one of the relevant legal topics to explore was intellectual property rights (IPR). EOSC-Pillar conducted a comparative analysis on IPR policy at the European and national level, also looking at gaps, the application of the IPR considerations. The result was the development of policy recommendations and a Legal compliance guidelines checklist for researchers. So far, the guidelines and recommendations have seen active interest and readership. The checklist has been implemented in several use cases and both the guidelines and recommendations are being fed into the Italian landscape through ICDI.

Another aspect of the work on a national scale was to support National Initiatives and EOSC at the national level. EOSC-Pillar showed how country-based policies and approaches support EOSC readiness and further development and standardisation for a successful EOSC implementation and uptake. EOSC-Pillar also supported national initiatives through an analysis of national initiatives, offering support for the definition or formalisation of policies and governance and coordinating a transversal task force for consulting national initiatives. Among the achievements were a better understanding of coordination opportunities between European Structural Funds and Investments and the EOSC Partnership summarised in the ERA-Learn Policy Brief on Synergies between R&I partnerships & ESIF and the EOSC-Pillar Roadmap for the Consolidation of National Initiatives.

See the full presentations from the EOSC-Pillar Final Conference session, Open Science Policies for EOSC Readiness: https://www.eosc-pillar.eu/events/final-event/policies-track-open-science-policies-for-eosc-readiness

Establishing FAIR Data Services

EOSC-Pillar also worked on the reduction of technical, societal and organisational barriers to ensure research data exploited in the project was FAIR.

This goal was achieved by addressing three main objectives:

- Develop an innovative set of services for a Federated FAIR Data Space (F2DS), starting from the content of existing research data repositories.
- Set up support and training activities, facilitating the dissemination and adoption of FAIR standards for research data management.
- Boost, collate and agglomerate domain-specific ontologies and related metadata as the main basis for cross-domain interoperability.

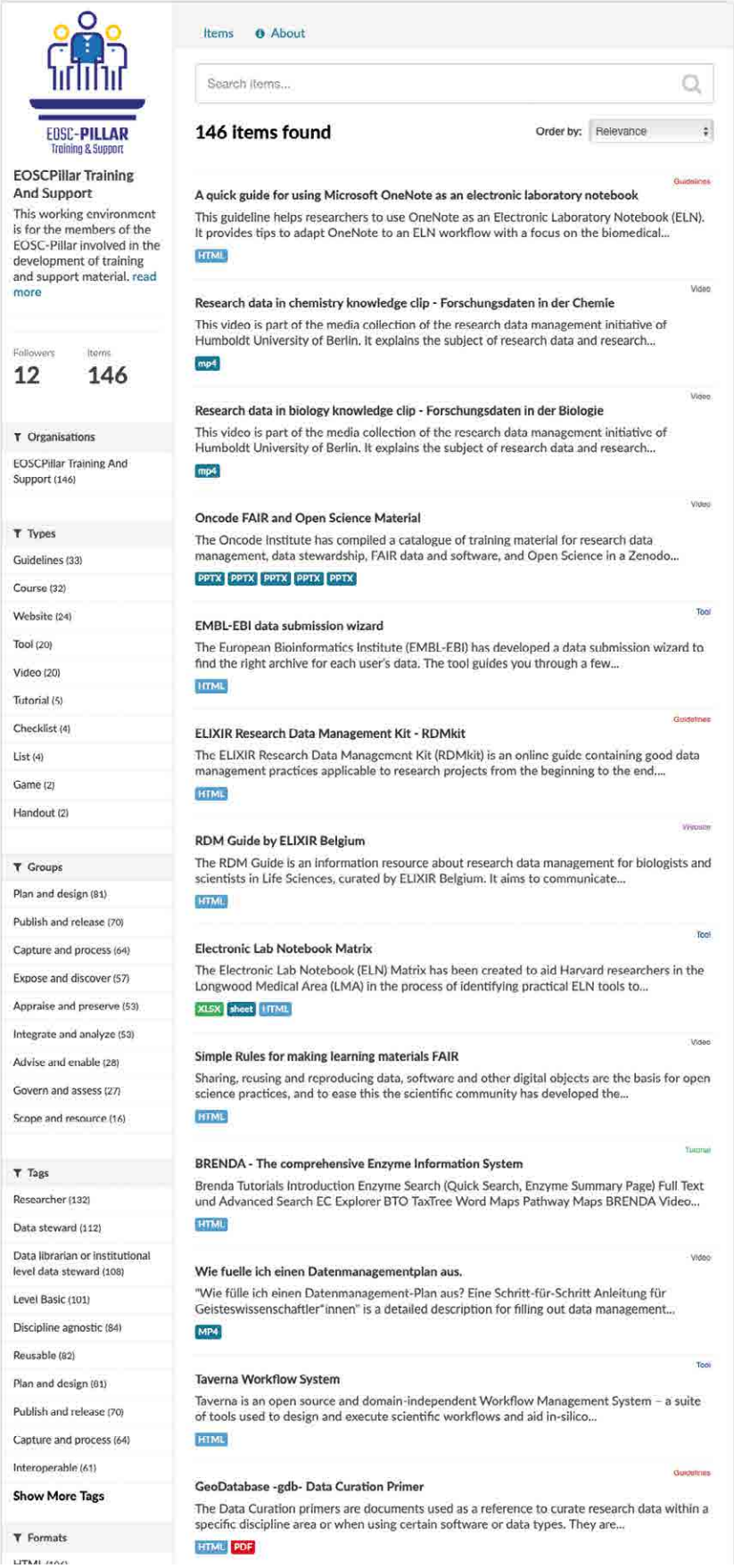
Aside from organising or co-organising both cross-cutting and thematic training sessions for FAIR principles uptake, the project focussed on the creation of the EOSC-Pillar Research Data Management Training and Support Catalogue, an online searchable collection of resources for FAIR data stewardship and RDM.

The Catalogue responded to the needs of data stewards, an emerging professional figure increasingly relevant in the research environment. Many projects and initiatives see the uptake of FAIR principles as an objective, while existing training and support resources were scattered, especially at the beginning of the project's lifetime.

EOSC-Pillar's efforts went into the curation and standardisation of the descriptions of these materials, in order to improve their findability and visibility, also providing a centralised point of access in the Catalogue. The metadata elements were then updated according to the specifications identified in EOSC Future.

See the relevant presentations from the EOSC-Pillar Final Conference:

- Session: Innovative Federated FAIR Data Spaces: https://www.eosc-pillar.eu/events/final-event/innovative-federated-fair-data-spaces
- Presentation: People & Culture - Training & User Support Approaches: https://www.eosc-pillar.eu/events/final-event/people-culture-track-ensuring-participation-community



Federated FAIR Data Space

The Federated FAIR Data Space (F2DS) provides tools for both data producers and data consumers contributing to enhance the overall FAIRness of datasets natively dispersed across heterogeneous repositories by providing innovative services for metadata harmonisation (DCAT), semantic enrichment and services for seamless discovery and access.

The F2DS can enable researchers to search, find and retrieve data using a single access point and tool set, not only saving working time - by masking access methods of different sources and offering a single access point through user and programming interfaces (UI and API) - but also delivering new insights, given the proper combination of selected search criteria and content of one or more data-sets explored in unison.

What has been done in EOSC-Pillar

The F2DS is a unifying data space built by aggregating and enriching datasets from a set of multidisciplinary repositories, i.e. data sources, with the aim to facilitate data discovery and re-use. Although datasets are the primary focus of the resulting data space, other items are managed including repositories and data sources, APIs, metadata schemas and ontologies.

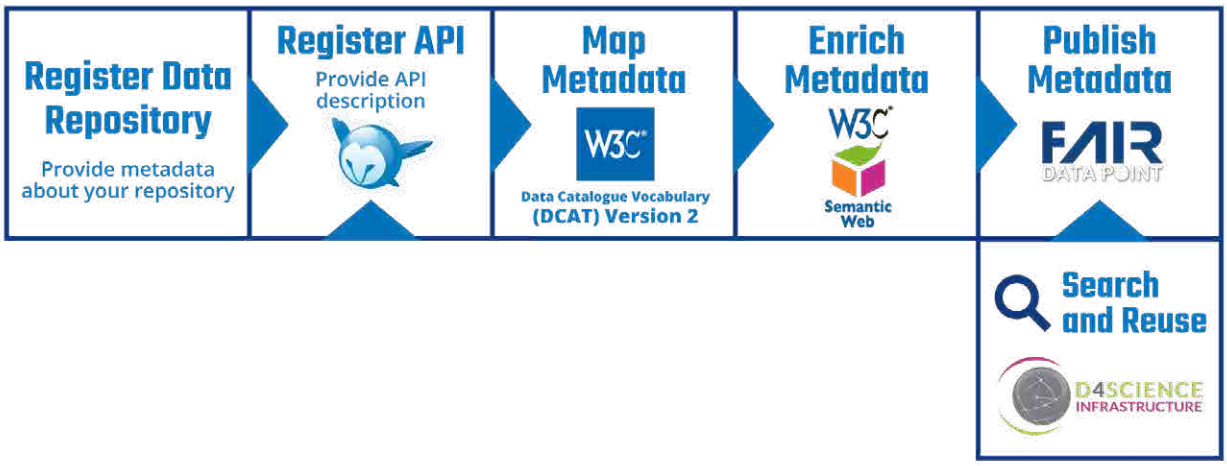


Fig. 1 Pipeline enacted by F2DS Services

It is implemented by two interacting services (see Fig. 2): the Metadata Repository and the Data Catalogue Service.

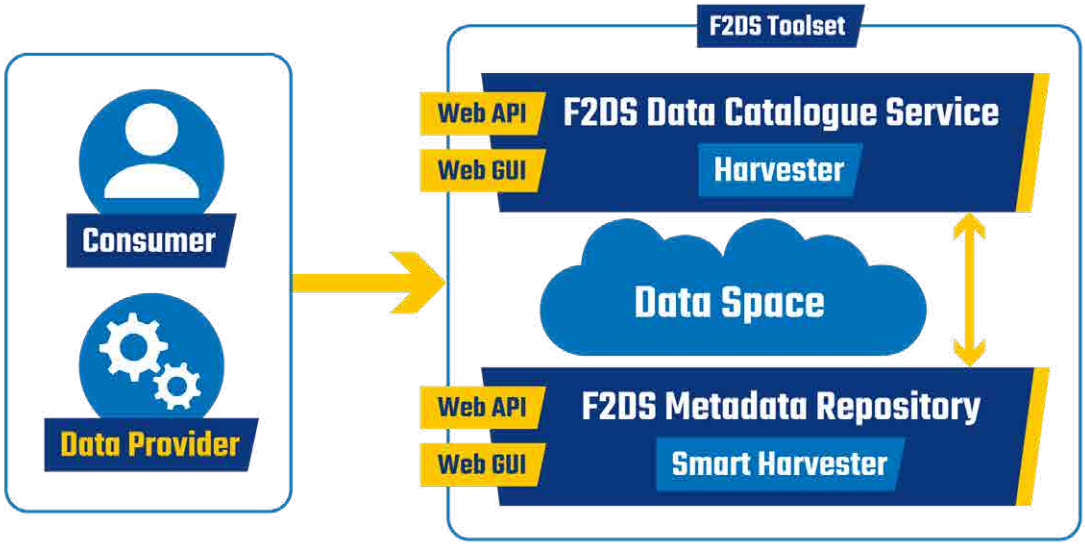


Fig. 2 F2DS Overall Architecture

The Metadata Repository implements a system to create machine actionable API descriptions of repositories and a simple metadata mapping interface to automatically and smartly harvest, convert, enrich and publish metadata describing datasets in a single format (DCAT). The Data Catalogue harvests metadata from the Metadata Repository with a dedicated harvester and makes them discoverable and accessible by a data portal based on the CKAN technology.

Findings and benefits

Data discovery and consumption needs are very heterogeneous in different disciplines and communities, and global solutions have limitations. It is important to cater for communities of practice, and the EOSC-Pillar F2DS toolset addresses this by bringing data closer to diverse users and usage needs.

The main benefits of the F2DS can be summed up as follows:

- Create in no-time data spaces across data resources, institutions, and domains
- No changes required for repositories.
- Make repository content more FAIR.
- Multiple APIs for accessing and processing F2DS content (SPARQL, REST API, and web UI).
- Easily connect any processing services.

See the full presentation from the EOSC-Pillar Final Conference: Innovative Federated FAIR Data Spaces: *https://www.eosc-pillar.eu/events/final-event/innovative-federated-fair-data-spaces*

Horizontal Data Storage and Computing Services

Aside from the Federated FAIR Data Space, EOSC-Pillar developed technological solutions supporting different communities, also supporting the onboarding of new services into the EOSC ecosystem.

National and regional services should be connected with the EOSC ecosystem: on one hand, this means to to understand which EOSC services may be beneficial for specific national communities and, on the other hand, to allow for scaling at European level by interconnecting the national/regional services with those in EOSC.

The key activity consists therefore in supporting the integration followed by validation and finally providing instances of the federating or enabling services in a production-ready environment. These services may be shared among different national infrastructures involved in EOSC-Pillar. The produced solutions provide support for a wide range of communities and include services that will be part of the EOSC infrastructure layer and will be proposed as candidate services for the inclusion into the EOSC.

A diverse list of services received technical support throughout EOSC-Pillar, mainly in the following areas: onboarding in the EOSC Portal Marketplace; adoption of a federated Authentication and Authorization Infrastructure (AAI); service quality improvement; user expansion. In particular, the following practical results were achieved:

- The use of acceleration resources
  - Onboarded in the EOSC Portal Catalogue & Marketplace, increasing quality (TRL9).
  - Many documentations now available, policies & operation manual written.
- Laniakea@Recas
  - Update following the best practice established on sharing and harmonising Galaxy deployment practices.
  - Use of the official Galaxy Project Ansible roles allowing for faster alignment of Laniakea flavours to the latest Galaxy releases.
- Contribution to Galaxy-E
  - Development of a scientific workflow for ecology in Galaxy.
  - A tutorial and dissemination of the results in conferences.
- Virtual Research Environments developed in the D4Science infrastructure & VRE-based service for research data publishing
  - 12 VREs were created, 500+ registered users.
- Pico2
  - Distributed memory program launched on a HPC cluster from a Galaxy server without writing a script in OAR.
  - Seamless communication between data and calculations through a connection between INRAE dataverse and the VM deployed on GRICAD.
- OIE and Marketplace projects
  - Connection to Dataverse.
  - AAI integration.
  - Maturity assessment provided a better view of the requirements' priorities.
- Readmetrics
  - Set up a dedicated platform (sandbox) for institutions to test the service without having to install it locally.
  - Indigo-IAM integration.

In addition, a model for a national Service Registry was developed in EOSC-Pillar in the Proof of Concept for the Italian community. It was based on the EOSC-Portal Profiles and ready to interoperate with the EOSC Portal Catalogue. This type of registry can be implemented by different actors and operated by any mandated institutions.

See the full presentation from the EOSC-Pillar Final Conference: EOSC-Pillar Contributions to the EOSC Exchange: *https://www.eosc-pillar.eu/events/final-event/national-initiatives-contributions-to-the-eosc-exchange*

National and thematic EOSC use cases

EOSC-Pillar collected use cases based on real-life requirements from scientific communities in Austria, Belgium, France, Germany and Italy. Pilots were run to validate the proposed solutions, designed to be trans-national and general enough to be extended to other communities with minimal changes.

The pilots included resources and cost study, in order to understand their feasibility and propose a viable business model for the resulting services. The EOSC-Pillar use cases were connected to real scientific production to improve the usage of data from a FAIR perspective.

Data FAIRisation is crucial, with different and specific needs for each community (nanotechnologies, environment/ocean, atmosphere, continental surfaces, health, humanities, biodiversity, solid earth). Our objective was to prepare scientific communities to be involved in EOSC.

The status of all use cases as of the end of the project can be explored in the dedicated posters produced for the EOSC-Pillar Final Conference.

- Use case 1 - Defining procedures and services to enforce data provenance for thematic communities and beyond - View poster
- Use case 2 - Agile FAIR data for environment and Earth system communities - View poster
- Use case 3 - Integration of data repositories into EOSC based on communities approaches - View poster
- Use case 4 - Software source code preservation, reference and access - View poster
- Use case 5 - FAIR principles in data life-cycles for Humanities - View poster
- Use case 6 - Exploring reference data through existing computing services for the bioinformatics community - View poster
- Use case 7 - Suitable data formats for seismological big data provisioning via web services - View poster
- Use case 8 - Virtual definition of data sets according to RDA recommendations - View poster
- Use case 9 - Integrating heterogeneous data on cultural heritage - View poster

The initial set of scientific use cases was expanded through an open call for participation, inviting communities and their developers to bring forward thematic services to enhance the portfolio of the project and the EOSC Portal. The open call was carried out between Dec 2020 and Feb 2021, and five service providers were accepted to become part of the project as new use cases. Three of these services also received support from EOSC-Pillar for onboarding into the EOSC Portal, more information can be found on the dedicated poster.

- Ethnic and Migrant Minorities' (EMM) Survey Registry - onboarded in EOSC Portal
- MMODA (Multi-Messenger Online Data Analysis) - onboarded in EOSC Portal
- ReadMETRICS
- SIMBAD - onboarded in EOSC Portal
- Thoth

See the full presentation from the EOSC-Pillar Final Conference: National and Thematic EOSC Use Cases Roundtable: https://www.eosc-pillar.eu/events/final-event/eosc-national-thematic-use-cases

The Community

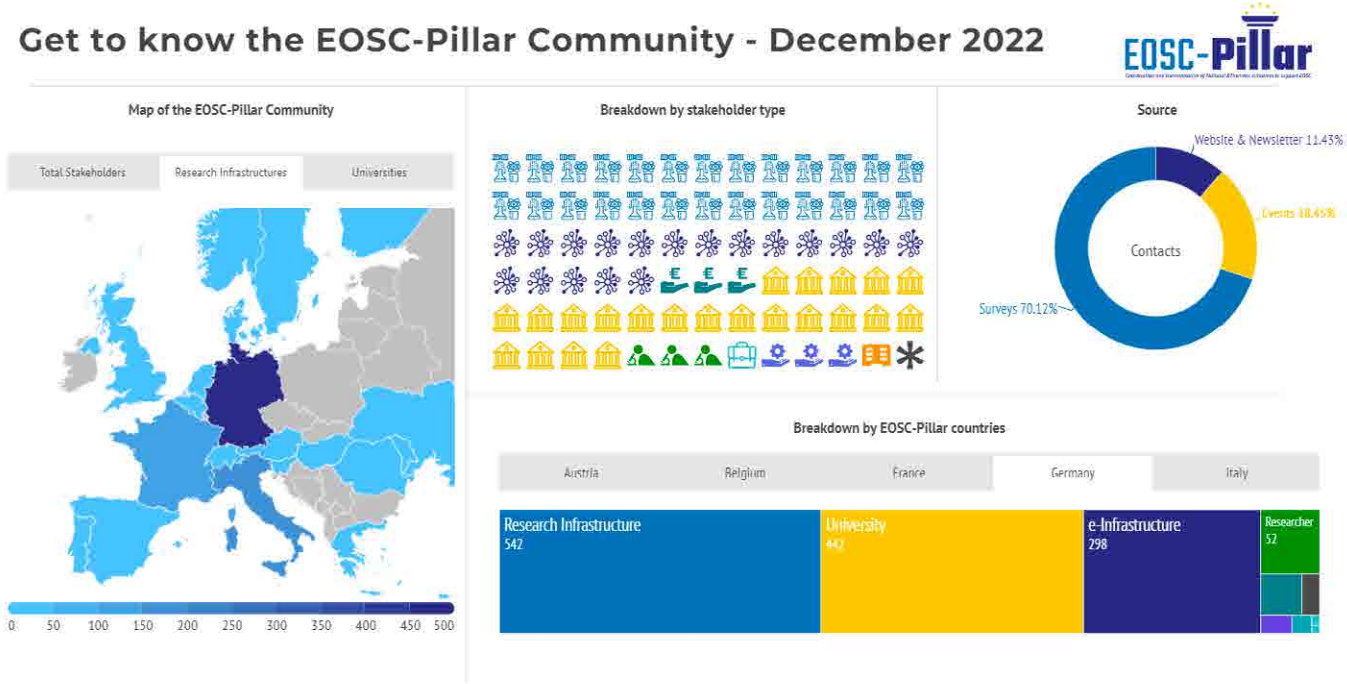
The Human Factor of the European Open Science Cloud has been a key focus for EOSC-Pillar which, through dissemination, outreach and engagement actions, has been able to build a wide community of user researchers, service providers at the national, regional and European level, as well as National funding agencies and EOSC-related and FAIR-related initiatives.

The members of the community were engaged in various project activities via subscriptions to the newsletter, registration via the web platform, registration to EOSC-Pillar webinars, participation at events, social media networks connections and various partners' efforts (such as the National Initiatives Survey) as well as synergies and strategic collaborations.

Ecosystem Overview

An overview of the EOSC-Pillar Community is provided in the image below, and also continuously updated in our community webpage, with details on geographical spread and stakeholder category. At the end of the project, over 3,100 people constitute this large ecosystem

Fig. 3 The EOSC-Pillar Community



Ambassadors Programme

During the last 18 months of the project, the EOSC-Pillar Ambassadors Programme was launched and carried out. The goal of this programme was to firstly raise awareness about EOSC and secondly have people spread the word about it.

The Ambassadors Programme aimed to support everyone willing to promote Open Science and the benefits of EOSC to their community, be it individual researchers or people working at research supporting or funding organisations. The goal was to establish a community of EOSC Ambassadors. Therefore, a complete set of informative materials and multimedia content was created and used for this purpose. The material included flyers, posters, presentations as well as videos and a podcast series: "**Stories of Data – Open.Science.Talk**". All content can be found on the EOSC-Pillar Website and might be adapted, shared and reused for similar purposes in the future. The first season of the podcast series (April-June 2022) was carried out in collaboration with the Horizon 2020 project SSHOC, while the second season (September-December 2022) was entirely dedicated to EOSC-Pillar.

Fig. 4 The Ambassadors Programme Podcast series

See the full presentation from the EOSC-Pillar Final Conference: People & Culture Track: Ensuring the participation of the Community: https://www.eosc-pillar.eu/events/final-event/people-culture-track-ensuring-participation-community

Sustainability and business models

EOSC-Pillar has taken actions to identify and define business models for open science services. One of the critical priorities for the success of EOSC, also highlighted in the Multi-annual roadmap for 2023-2024, is to identify and test a number of funding models to ensure that EOSC services are viable over the long term.

The Business Model Canvas method was applied to 10 Open Science IT services from the five EOSC-Pillar European countries, to develop and evaluate their business models. Workshops and interviews were held with the 10 use cases. Data collected was then analysed to identify sustainable business models for open science. 4 business models were found viable for Open Science IT services, namely advertising, commissioning, pay-per-use and subscription. However, most of the use cases still combine these models with the use of public funding. Indeed, as a matter of fact, various challenges are faced to guarantee sustainability of these models. For instance, it is very often difficult to cover the requests of all the target customers as they may have domain-specific differences in terms of open science understanding, FAIR data definitions or vocabulary. Also, Open Science IT services may find it difficult to communicate the value of their innovation to users and society. Moreover, handling a big amount of data can be very difficult. Considering the value creation aspects, there is a high dependence of Open Science IT services on the support and the infrastructure of their home institutions and there's a shortage of skilled professionals.

The conducted interviews also showed that, especially human resources are expensive, accounting for more than 70% of the costs. To cover these costs a critical number of users is needed for every business model.

Overall, the main results of the whole analysis indicate the following regarding sustainable business models:

- More business models should be combined to provide open science IT services
- Data software and platform as a service models use different business model patterns to cover the operating costs
- All Open Science IT services face different challenges regarding the sustainability of their business models. Therefore funding phases should only support the development phase but also the following phase of introduction to market when a critical mass of users has not been achieved yet
- When the Open Science IT service reaches the growth phases it is generally able to self-fund the initiative using a combination of business models.

See the full presentation from the EOSC-Pillar Final Conference: Sustainability and Business Models: https://www.eosc-pillar.eu/events/final-event/sustainability-track-business-models

www.eosc-pillar.eu

@EoscPillar

/company/eosc-pillar/

bit.ly/3evkLxI

zenodo.org/communities/eosc-pillar/

anchor.fm/eosc-pillar

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 857650.