EOSC-hub

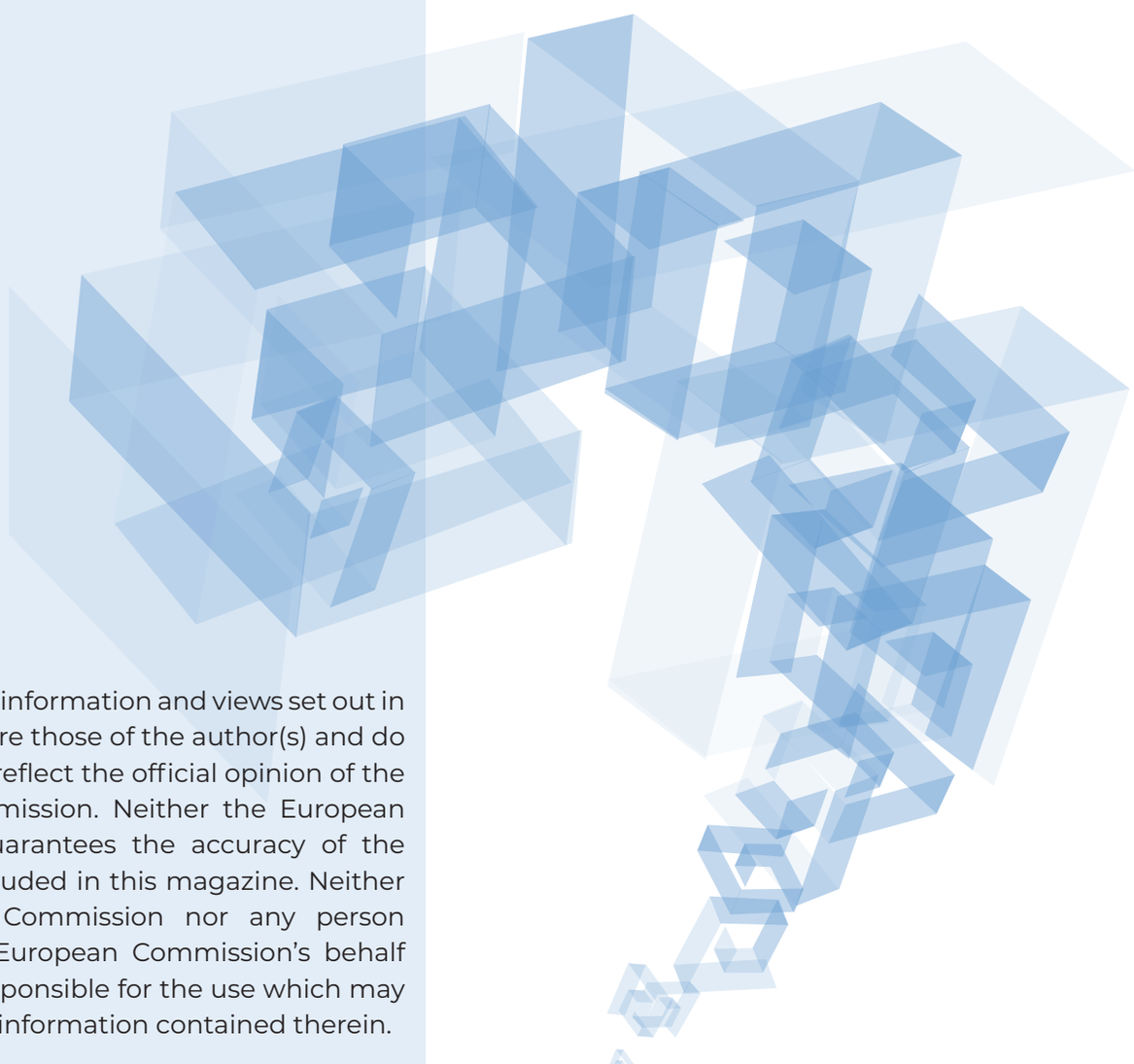
MAGAZINE

ISSUE 2
October 2018

eosc-hub.eu

EOSC-hub Magazine

The EOSC-hub Magazine is a publication of the EOSC-hub project, edited to showcase major results and achievements of the project, collaborations ongoing with other initiatives and updates from the communities. The magazine also provides an overview of the latest highlights from the European Open Science Cloud (EOSC) landscape.



Issue 2 // Table of content

About EOSC-hub

The EOSC-hub project brings together multiple service providers to create the Hub: a single contact point for European researchers and innovators to discover, access, use and reuse a broad spectrum of resources for advanced data-driven research.

For researchers, this will mean a broader access to services supporting their scientific discovery and collaboration across disciplinary and geographical boundaries.

The project mobilises providers from the EGI Federation, EUDAT CDI, INDIGO-DataCloud and other major European research infrastructures to deliver a common catalogue of research data, services and software for research.

EOSC-hub is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement 777536.

Start: January 2018

End: December 2020

Total budget: 33 million

100 Partners


SAVE THE DATE
EOSC-hub
Week 2019
9-12 April 2019

EOSC-hub and ESFRIs towards co-creation: the key for the success of EOSC

Per Öster highlights how the ESFRI clusters are at the core of the EOSC-hub vision

The consolidation of ESFRI clusters as EOSC thematic platforms, and the articulation of these platforms with the EOSC-hub will be critical for the success of EOSC. Research infrastructures have spent decades building strong thematic platforms and organising their community of users, and are rightfully expecting that EOSC supports this effort. How can EOSC-hub support these platforms?

We propose here a vision where the EOSC stakeholders, in particular data and service providers, join forces together to establish the Hub as an access channel to data and services - coming both from research infrastructures and e-Infrastructures - made open to the entire research community.

We envision the Hub to be used by researchers and research communities in a bi-directional way:

- to discover and use new generic and thematic services;
- to open up data and services to new users.

Research infrastructures want to remain in control of their own delivery and access channel to serve their organised user base. But, when research infrastructures have an interest to open services beyond their community, the Hub is precisely the right place to promote these resources and facilitate their use.

How is EOSC-hub promoting this vision?

First, the project is pursuing an active engagement strategy with ESFRI clusters to discuss concrete requirements, operational agreements and governance, in the spirit of co-creation. It is crucial to have a genuinely shared ownership of the discussion and co-creation of this common future.

At technical and policy level, the project is making sure that the existing instruments (e.g. marketplace, software and metadata catalogues) are fully operational and fit for purpose, and that the rules of participation for publishing data sets or services are easy to understand and implement. Promoting a service through the EOSC-hub should be possible with very little effort for the service providers!

The question of business models for opening up services also needs to be addressed. The current instruments provided by the EC (e.g. Virtual Access) are primarily designed for access to research infrastructures and not consumption of compute and storage resources. Innovation is needed in this area to create incentives to attract service providers.

With this agenda in mind, EOSC-hub is happy to continue to engage with all interested parties.

Per Öster, Director Research Infrastructures & Policy at CSC and EOSC-hub Project Director
@pererikoster

EOSC in practice – WeNMR

Alexandre Bonvin explains how the WeNMR community builds on EOSC principles to support 12,000 structural biology researchers

1) What are the main research goals of your community/group?

WeNMR is serving the structural biology community at large. Structural biology studies the functions and interactions of proteins, nucleic acids and other biomolecules using experimental methods such as X-ray crystallography, Nuclear Magnetic Resonance (NMR) or cryo-electron microscopy (cryo-EM). All these methods generate data that needs to be processed, analysed and finally converted into three dimensional (3D) structures (or models) of biomolecules using a variety of computational tools and techniques.

Gaining access to 3D structures of biomolecules, their dynamics, and their interactions with other molecules is key to a proper understanding of their function. It also allows you, for example, to rationalise the effect of disease-causing mutations, to engineer better molecules for material, health or food applications and to obtain a starting point for drug design to combat disease. As such, structural biology has a strong socio-economical impact on many application fields from health, to food, to materials.

2) How many people are involved in WeNMR?

The WeNMR collaboration with EOSC-hub involves Utrecht University in the Netherlands, and the University of Florence and INFN Padova in Italy. They are the partners responsible for the operation, maintenance and further developments of the WeNMR thematic services. Those services could however not be supported without the strong commitment of resource providers giving us access to grid, cloud and data storage computing resources. This support has been formalized by a Service Level Agreement with the EGI Federation.

Our user community is however much larger. We have over 12,000 registered users over the years from more than 95 different countries.

3) What are the services you provide, or want to provide, to this collaboration?

The WeNMR thematic services (bit.ly/2KOTyaR) are a suite of web portals, providing user-friendly access to complex computational workflows and tasks. These allow inexperienced and experienced structural biologists to use state-of-the-art software for their data analysis while benefiting from the computational infrastructure provided through the EOSC-hub project. The services make use of high-throughput computing (HTC) resources, but some are also using accelerated computing (GPGPUs) grid resources and cloud computing.

As community we have always been proactive in using new technologies in collaboration with EGI and in the context of European projects. We have, for example, piloted the use of GPGPU resources on the grid using the INDIGO-DataCloud udocker solution. Two of our portals are actively using those.

4) What are the computational challenges?

We need to provide user-friendly tools to our users, hiding the complexity of grid/cloud computing and ensure sufficient resources to operate those. WeNMR has a long history of using HTC resources under EGI. Maintaining the quality of our services and support, together with continuously adapting and improving them (e.g. to make use of new compute models, or facilitate their use through the implementation of single-sign-on mechanisms) is a constant challenge.

5) The EOSC is being set up to be Europe's virtual environment for all researchers to store, manage, analyse and re-use data for research, innovation and educational purposes. How will you interact with this environment?

Within EOSC-hub we are planning to connect some of our portals to data repositories such as the ones offered by EUDAT in order to allow user to directly upload and/or download data/results. The data generated by the WeNMR services are however very specific to a user/application and not globally reusable by third parties. This is very different for example from sky images collected by telescopes. We do aim, however, at facilitating data deposition into public repositories where relevant.

6) What opportunities will the EOSC open for your community?

Hopefully the landscape of data and compute resources will become much more unified and transparent to our users. Ease of use is key here.

7) How do you imagine your field in ten years?

In ten years some of the computational approaches that require some level of expertise (and access to resources) will become commodities. The use of the EOSC and its associated HTC/HPC and data resources will become as natural to the new generation of researchers as using a smartphone. We know that some of our portals are already being actively used for education purposes in various bachelor, master and more advanced courses. It is great to see and to know that we are contributing to the training of the next generation of scientists.

WeNMR portals

The WeNMR portals (bit.ly/2KOTyaR) cover different area of structural biology such as NMR structure calculations and data analysis (e.g. the AMBER, Xplor-NIH and FANTEN portals), the fitting of structural models into cryo-EM maps (PowerFit), the analysis of mass spectrometry cross-links (DisVis) or the integrative modelling of biomolecular complexes (HADDOCK).

The WeNMR tools are powered by High-Throughput Compute capabilities provided by the EGI Federation and enhanced with software components developed by the INDIGO DataCloud project.

EOSC-hub at DI4R 2018

The Digital Infrastructures for Research conference is just around the corner!

This year, the event will be held at the University of Lisbon campus in Lisbon, Portugal, from 9 to 11 October. DI4R 2018 is jointly organised by EOSC-hub, GÉANT, OpenAIRE and PRACE and will bring a packed programme of sessions, posters and networking opportunities to Europe's researchers, developers and service providers.

The EOSC-hub project will be present at DI4R with:

Sessions

Tue 9 *Rules of Participation for EOSC* – focusing on conditions that allow services providers to participate in EOSC in a transparent manner.

Tue 9 *Have a CoP of T in our café!* - dedicated to training coordinators and managers to map out training activities of various pan-European, EOSC-related initiatives and strengthen their capacity.

Tue 9 *EOSC Service Architecture: how the services could support the user communities* - giving an overview of the EOSC-hub service architecture and showing users how to build new services starting from the existing ones.

Wed 10 *EOSC-hub service portfolio and applicable policies for service providers and users* - presenting the EOSC-hub service catalogue, the EOSC-hub marketplace, and the rules of participation for becoming a service provider of the Hub.

Wed 10 *Shaping the EOSC service catalogue: what users need* - raising requirements from the audience to identify high-level needs and priorities for potential services of the EOSC catalogue.

Wed 10 *Research with Sensitive Personal Data in the EOSC* - reviewing the state-of-the-art digital services for research with sensitive data, particularly in the field of health.

Wed 10 *Towards the EOSC AAI service for research communities* - introducing the main concepts for meeting research community needs to access EOSC federated resources and providing an opportunity for researchers and infrastructures to share their experiences.

Wed 10 *EOSC from Theory to Practice* - giving an overview of the ongoing efforts towards the establishment of an effective EOSC.

Thu 11 *Digital Innovation Hubs for Industry Engagement* - showcasing the EOSC-hub Digital Industry Hub (DIH) structure and engagement model and gathering existing European DIHs and initiatives to facilitate a closer collaboration with the EOSC.

Thu 11 *The Frontier of Data Discovery* – presenting the current state on data discovery from OpenAIRE, EOSC-hub and from the angle of a given community.

Thu 11 *Building better collaborative national networks to support Open Science* - bringing together national representatives of OpenAIRE, EOSC-hub, GÉANT, PRACE and RDA Europe.

Lightning talks

Wed 10 *WeNMR activities in the EOSC-Hub* – an introduction to WeNMR thematic portals, their solutions for researchers working in Structural Biology and collaboration with EOSC-hub.

Wed 10 *Preventing security incidents in the EOSC-hub era* - the talk will describe plans for evolution of software vulnerability handling in the EOSC-hub project.

Wed 10 *EOSC-hub market research and business model analysis: call to action* - a short talk to better understand the demand for digital services and resources for research over the coming years.

Trainings

Tue 9 *Planning early, following through: Data Management Planning in the EOSC* – for supporters of research projects and research infrastructures and other stakeholders that are managing research data.

Thu 11 *Security Incident Management in the EOSC era* - a hands-on training focused on technical and higher level aspects of incident response, where participants will analyse Virtual Machine disk images. In the second part, the participants will handle a high-profile incident, coordinated by the Incident Response Task Force..

Demos

Tue 9 *EOSC-hub Business Pilots* - the purpose of this demo is to present the first outcomes of two successful EOSC-hub Business Pilots that are part of the EOSC Digital Industry Hub.

Wed 10 EOSC-hub Marketplace Tool – this demo will show the achievements of the EOSC-hub Marketplace functionalities allowing users to search, discover, order and access EOSC-hub services.

Presentations

Wed 10 *Applications Database: New features for user communities* – a presentation of the new features of EGI AppDB.

Thu 11 *IT Security Management (ISM) in EOSC-hub: policies and global trust* – this presentation will give an overview of the developments made by the EOSC-hub security team in 2018.

And a lot of posters!

We look forward to greeting everyone at #DI4R2018 in Lisbon!

@DI4R_eu

DIGITAL INFRASTRUCTURES for RESEARCH 2018
Serving the user base

Secure services for sensitive data in research

An overview of secure data services offered through EOSC-hub, by Abdulrahman Azab, Francesca Iozzi and Antti Pursula

Research often involves the use of personal data as a basis for the scientific analysis. However, a particular challenge in this area is to use these data resources without violating privacy. And for that we need secure digital infrastructures, compliant with both national and European regulations.

The EOSC-hub project is working on the topic of providing services for sensitive data through two partners: the Sigma2 / University of Oslo in Norway, and the CSC in Finland.

CSC offers the ePouta secure cloud infrastructure, which provides customer organisations a virtual private cloud connected to the customer's infrastructure through a secure virtual private network. This private area of the ePouta cloud is only accessible via customer-provided endpoint, ensuring both full control and flexibility to bring in customer's own software environment. CSC ePouta has 8000 hyperthreading compute cores within about 130 servers. Most of the servers have 256 or 384 GB RAM but there are also high RAM servers up to 1,5 TB RAM, and GPU nodes. There is altogether close to 2 PB of storage within the same infrastructure.

The University of Oslo provides TSD - the Norwegian e-Infrastructure for sensitive data storage and management. TSD provides sensitive data services directly to researchers and groups in the form of SaaS (Software as a Service) and PaaS (Platform as a Service). Each TSD project has a separate VLAN and is accessed through two-factor authentication. Currently there are 507 projects in TSD. TSD supports: Data Storage, Web forms, High Performance Computing (HPC), Audio/Video streaming and analysis, and software management for Windows and Linux platforms.

TSD and ePouta can be used for secure data processing and sharing data reliably with collaborators.

The work within EOSC-hub project aims to widen the services to data discovery through non-sensitive metadata increasing possibilities for data reuse.

Sensitive data services in practice

A research team may want to record videos of interviews, for example, to study both the spoken language and the gestures and facial expressions of the interviewees. This data cannot be shared in open access, as individuals are directly identified from the recordings. The team can upload the data to ePouta or TSD, and invite their collaborators from other institutes to access the data at the same secure platform. In the future the researchers could also publish metadata through EUDAT B2SHARE and B2FIND services, allowing other researchers discover the data set, and ask for access permissions from the original data providers.

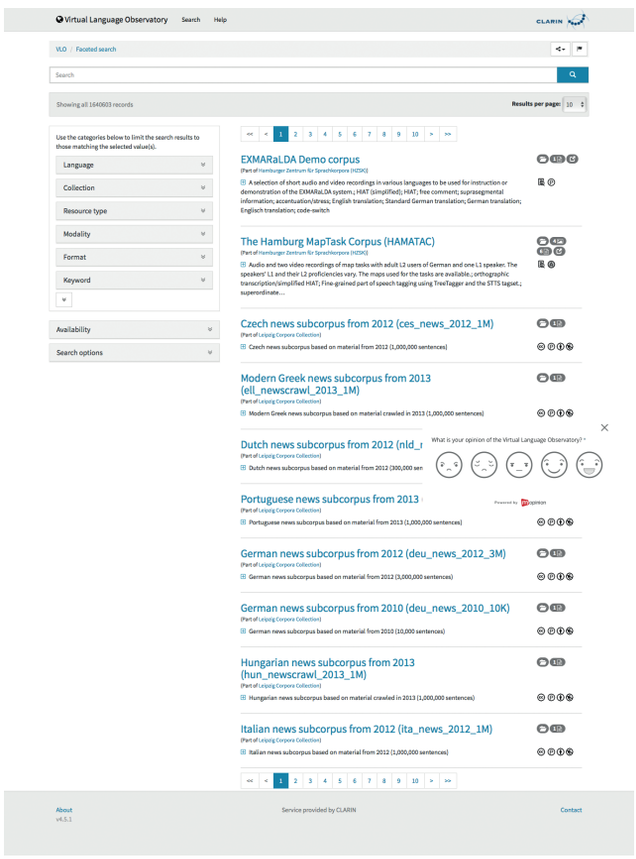
CLARIN's Virtual Language Observatory

Twan Goosen writes about an open digital tool for the Social Sciences and Humanities

The Virtual Language Observatory (VLO) is a service provided by CLARIN offering uniform search and discovery functionality for language resources and tools. The metadata indexed is heterogeneous in terms of content and structure. This metadata is sourced regularly from over fourty CLARIN centres that provide resources or tools of interest to scholars with an interest in language data. On top of that, CLARIN harvests from a few dozen external providers that are also providing relevant content.

The VLO has about 850,000 metadata records that can be searched, browsed and viewed. The addition of another 780,000 records describing cultural heritage objects from Europeana is currently under evaluation.

The VLO is openly accessible via the web to anyone. Researchers can freely enter a search term and/or use a number of pre-defined facets to refine the search results. This method of faceted browsing is as easy as using an online store and allows for quick filtering on basis of object language, nature of the resource, subject or organisations involved.



The VLO is part of CLARIN's Component Metadata Infrastructure and can cope with many different metadata descriptions, as long as they are implemented through (or converted to) the Component Metadata framework. Component Metadata 'profiles' can be defined to contain any number of fields and (sub)components and allows for semantic annotation of all of these, through which interoperability across profiles is achieved. This principle is exploited by the VLO, which maps a wide range of metadata fields to the relatively small set of fields underlying its search and browsing facilities on basis of explicit semantics.

While the VLO has been optimised for language resources and tools, it is not tied to any specific domain. Other fields, especially those that face a strong diversity in terms of metadata formats, could benefit from adopting the component metadata approach. During the EOSC-hub project, CLARIN will work on making the VLO more generically applicable so that any community can adopt it with minimum effort.

Go to vlo.clarin.eu and press 'Take a quick tour' see the VLO's main features demonstrated. You can learn more about CMDI at clarin.eu/cmdi.

Twan Goosen is a developer at CLARIN

EOSC-hub contribution to the EOSC open consultation on Rules of Participation

Tiziana Ferrari summarises the project's recommendations

The EOSC Open Consultation on Rules of Participation and FAIR Data was launched by the High Level Expert Group (HLEG) to gather feedback from the stakeholders on this critical area for EOSC implementation. EOSC-hub responded to the consultation with a contribution focused on Rules for Participation, extracted from the project's ongoing work on the theme encapsulated in the Deliverable 4.1, entitled Operational requirements for the services in the EOSC catalogue.

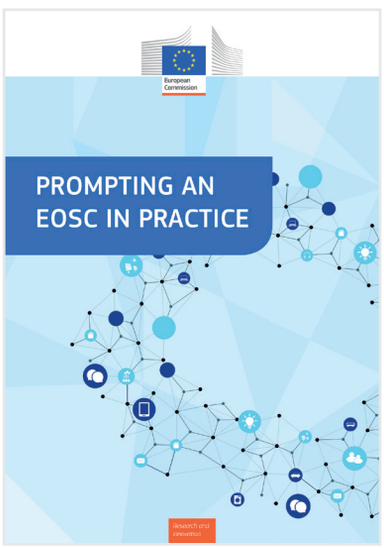
Level of integration

EOSC-hub proposes three different levels of operational levels of integration:

- High integration: for services operated according to the EOSC service management system. The service provider actively participates to the Operations Coordination;
- Medium integration: for services that run with a more mature Service Management Framework;
- Low integration: for services run with a less mature Service Management Framework.

This can then be mapped to two classes of services:

- Access-enabling services › High Integration: the services needed to operate the EOSC itself (internal catalogue). These include all federation-level services (for example: Accounting and Monitoring, the Helpdesk, Configuration Management) that provide the foundation for all researcher-facing services. Services in this class cannot be ordered, as such by users.
- Research-enabling services: services offered to users and research communities by means of a Marketplace (external catalogue). These can be further divided into:
  - *common services* › *Medium Integration*, which can be re-used by other services (e.g. compute, storage and data management services), and



  - *other researchers-enabling services* › *Low Integration* (e.g. a scientific application offered by a Research Infrastructure).

Given the past experience of the EOSC-hub partners and the project's early findings, EOSC-hub foresees the need to support different levels of service management integration, according to the specification of each class.

Maturity Level

Technology Readiness Level (TRL) is a gauge for the maturity of a technology. The EOSC-hub service catalogue is restricted to services considered mature enough to be at production level with TRL 8 and 9.

Our contribution to the consultation recommends that in EOSC TRL8 remains the minimal requirement. Services with TRL 8 and 9 have passed through the previous development states of proof-of-concept, pilot and pre-production, and will have successfully proven to users that the services are mature and fit-for-purpose for their

Tiziana Ferrari is the Project Coordinator of EOSC-hub and Technical Director of the EGI Foundation.

@tferrariEGI

The Hub to be featured during EOSC launch

The official launch of the European Open Science Cloud (EOSC) will take place on 23 November 2018 in Vienna, as part of the programme of the Austrian Presidency of the Council of the European Union.

The EOSC launch is the nexus of more than €300 million worth of funded projects and cross-border collaboration for the past three years. The realisation of the EOSC is an integral part of the broader Digital Single Market strategy that was announced in 2015. EOSC also falls under one of the three key priorities of the Austrian Presidency which is "securing prosperity and competitiveness through digitalisation".

Opening the launch event will be Austrian Minister for Education and Science Heinz Fassmann along with officials from the European Commission.

EOSC-hub's role in the realisation of the EOSC will be described in the dedicated session of the event for introducing the EOSC portal, the new web-platform meant to become the reference point for EOSC-users, -providers, -contributors, led by European Commission Director-General for Communications Networks, Content and Technology (DG CONNECT) Thomas Skordas.

In addition, the contribution of the EOSC-hub project to the implementation of the EOSC Portal, together with OpenAIRE-Advance and the eInfraCentral projects, will be showcased highlighting the EOSC-hub marketplace as the engine for access EOSC services and resources.

The EOSC Portal at a glance

The EOSC Portal will provide general information about EOSC to its stakeholders and the public, including information on the EOSC agenda, policy developments regarding open science and research, EOSC-related funding opportunities and the latest news and relevant events, but most importantly will offer a seamless access to the EOSC resources and services.

The Portal will become the reference point for the 1.7 million European researchers looking for scientific applications, research data exploitation platforms, research data discovery platforms, data management and compute services, computing and storage resources as well as thematic and professional services.

Access to EOSC services and resources will be enabled via the EOSC-hub marketplace.

The launch of the EOSC Portal isn't just a milestone for the European Open Science Cloud, but it also marks the first direct and tangible contribution of EOSC-hub to EOSC.

23 November 2018, 10:00 - 13:30 hrs

University of Vienna

Library, main reading room

How to make your data Open and FAIR?

Ilona von Stein writes about the importance of Data Management Training

What would happen if your data is not trusted by your own research community?

What if your research community generates 2 petabyte of raw data per day, and needs to share it between institutions?

What if reuse of your data takes place in computational workflows, but data archives do not provide formats and annotations to support this?

And what if you cannot find your own data later in your career?

Open and FAIR data

Challenges are ambitious, and the key is to make data Open and FAIR. The FAIR data principles (Findable, Accessible, Interoperable and Reusable) act as a guideline to enhance the reusability of data. And indeed: worldwide there is an astonishing number of tools and services that help you to "do something" with your data. But which tools and services do we have in our EOSC-hub communities that really contribute to making our data open and FAIR?

A typical research data lifecycle includes steps for re-using, creating or capturing data, processing, analysing, publishing, sharing, and preserving. EOSC-hub offers services to support each step and the training resources necessary to make a Data Management Plan – or DMP, for short.

Data Management Planning

DMPs help researchers and research teams to consider *What* data goes into a project (reuse) and what comes out of it (potential reuse), *How* the team takes care of the data, and *Who* is allowed to do *What* with the data *When*. (And yes, funders increasingly demand DMPs). The information in a DMP covers the research data lifecycle and is essential information for your research support office and IT department or service provider: after all, the DMP specifies

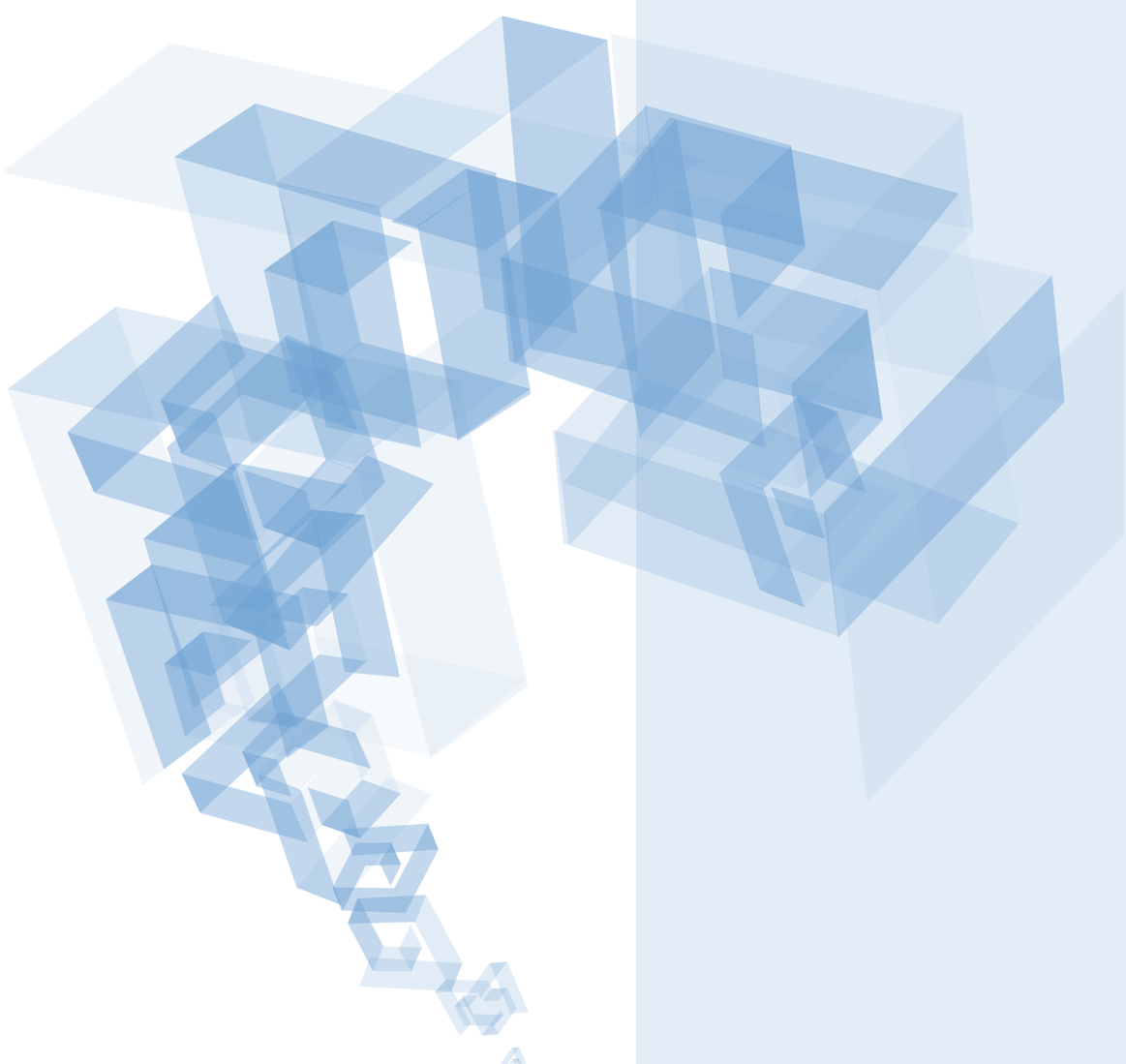
the services and legal support that the project needs to make the data as FAIR and Open as possible.

Training team of EOSC-hub

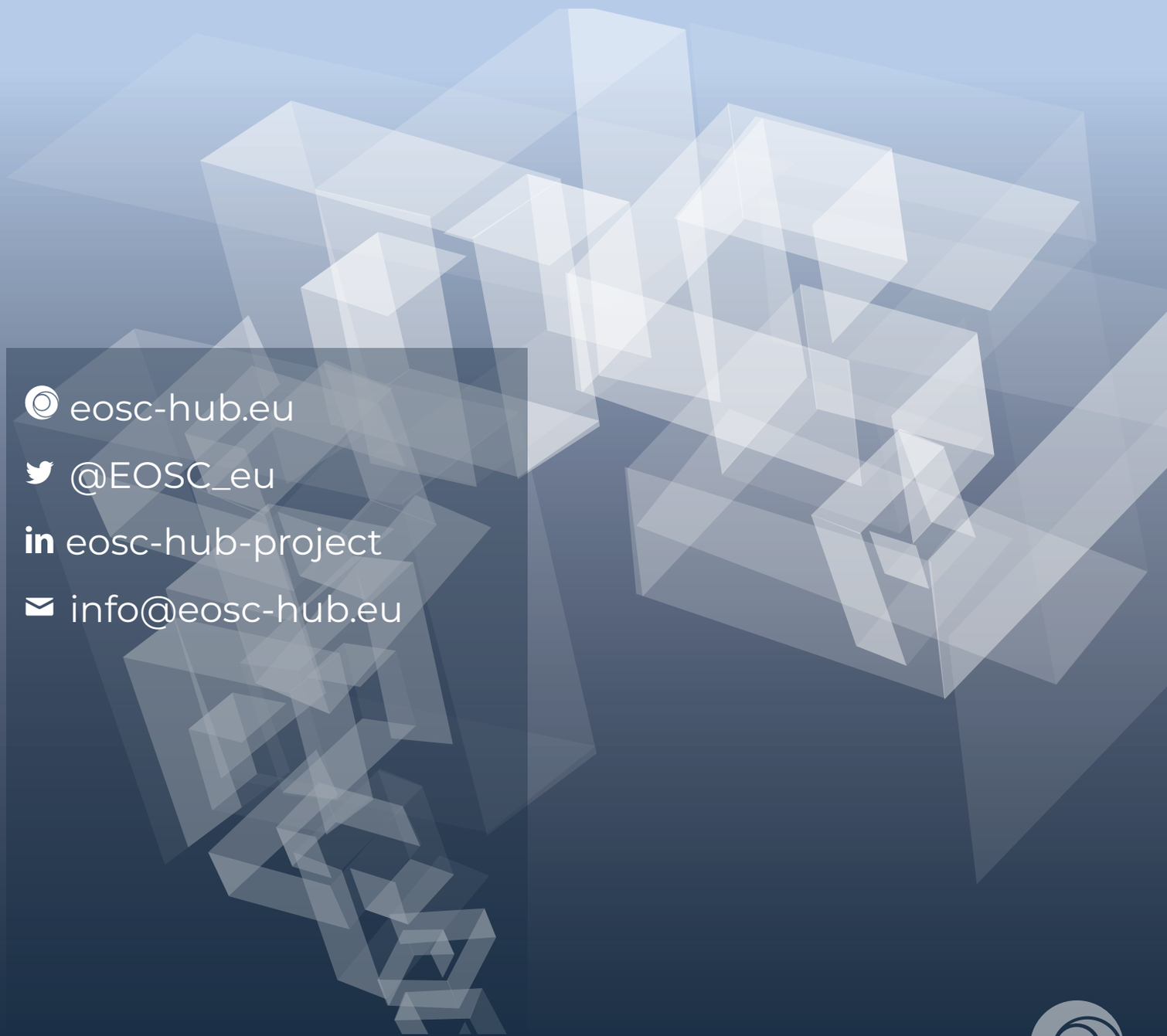
The training team of EOSC-hub, DANS and CCFE, supports researchers in proper management of research data. In collaboration with OpenAIRE-Advance, the team provides information about the *Why* and *How* of the creation of DMPs and about the DMP review process of major research funders, such as the Horizon 2020 programme. Next to support on DMP planning, EOSC-hub also offers training on particular services.

Ilona von Stein is project leader and policy officer at DANS

More information about training:
eosc-hub.eu/training-material
training@mailman.eosc-hub.eu



EOSC-hub



eosc-hub.eu

@EOSC_eu

in eosc-hub-project

info@eosc-hub.eu

EOSC-hub receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777536.