

The Basics of EOSC and FAIR

04 June 2019

The FAIRsFAIR project and [DANS](#) (FAIRsFAIR coordinator) contributed to the "10 FAIR things" sprint by working on the topic "The Basic EOSC and FAIR in 10 Things" article, part of the "[TOP 10 FAIR Data & Software Things](#)", published by [Library Carpentry](#) (a global community teaching software and data skills to people working in library- and information-related roles). Read the full article below:

The Basics of EOSC and FAIR

Audience: Library staff who provide research support & Researchers and research communities

Sprinters: Marjan Grootveld, Frans Huigen, Eliane Fankhauser, Ellen Leenarts ([Data Archiving and Networked Services - DANS](#)) & Paula Andrea Martinez / [ELIXIR-Europe](#)

Brief Description: In the ideal world, everyone would have access to all research outcomes and available knowledge, to use it, build upon it, and expand it to serve societal goals as well as private and personal interests. The ambitious goal of the European Open Science Cloud or EOSC suggests a "location" where this vision would become a reality. But what exactly is this EOSC, where does it come from, who is involved, and why would researchers use it? These Top 10 Things address these questions. Moreover, they link the EOSC to the reusability of research data that will be available. To this end, we link EOSC to the FAIR data principles.



Thing 1 □ EOSC: a vision that affects us all

In 2000, a predecessor of the European Open Science Cloud (EOSC) was endorsed by the [Lisbon European Council](#). Named the *European Research Area*, this system of scientific research programs concentrating on European-wide cooperation in the fields of medical, environmental, industrial, and socioeconomic research. It is within this framework that projects concerning open science, open access and open market for researchers are broadly initiated and funded by the European Commission. In the course of the past twenty years, the initiatives grew, developed and improved even further.

[In his 2015 speech](#), Carlos Moedas - then Commissioner for Research, Science, and Innovation - launched the idea of the European Open (Science) Cloud. It was the commission's vision to initialize a large infrastructure with the aim to support and develop - among others - Open Innovation and Open Science across Europe and beyond.

The EOSC has been officially launched on November 23rd, 2018. By 2020, the large-scale European High Performance Computing, data storage and network infrastructure are envisioned to be largely in place. In the future, more projects will emerge in the EOSC ecosystem, for example synergizing projects like EOSC-secretariat.eu. Also, [five Working Groups](#) have been formed by the EOSC Executive Board, of which one concentrates on FAIR (Findable, Accessible, Interoperable, Reusable) data. **We will discuss FAIR Data in our Things 7 and 8 lateron.**

This EOSC ambition has been incorporated in the [European Horizon 2020 flagship initiative](#), promising "(...) more breakthroughs, discoveries and world-firsts by taking great ideas from the lab to the market", and over [40 countries are involved](#). With funding by this Horizon 2020 initiative, the [EOSC offers](#) 1.7 million European researchers and 70 million professionals in science, technology, the humanities and social sciences a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines by federating existing scientific data infrastructures, currently dispersed across disciplines and the EU Member States. The cross-national collaboration and cooperation is what makes EOSC 'us', it is what defines the European integration of Open Science. The EOSC motto is: [as open as possible, as closed as necessary](#).

Thing 2 □ Buzzwords: a Thing for skeptics

In the context of this ambitious trajectory we can't completely avoid buzzwords. Here is our attempt to clarify and scope a few of them, for the remainder of this Top 10:

EOSC: right, this is the European Open Science Cloud. However,

- E: it is not limited to Europe, witness [the outward arrows](#);
- O: it also supports research findings that cannot be openly shared with everyone, for instance because they relate to individuals, security, or commercial interests;
- S: stimulating citizen science and bringing ideas to the market are express goals of the European Commission;
- C: the cloud is already a metaphor on its own.

FAIR: this is a brilliant* acronym for Findable, Accessible, Interoperable and Reusable, which are all desirable aspects of research data. The "FAIR" concept got worldwide traction in the span of only a few years (see **Thing 7 about the FAIR Principles**). And who could be against it? The proof of the pudding is in the eating, however. Therefore we applaud all initiatives around the world that aim to



operationalize the principles for different scientific disciplines and research methods, as well as all attempts to measure the FAIRness level of data, software and other related concepts.

* *By way of exercise: communication-wise “FAIR” is a brilliant notion. However, some argue that essential aspects of research data are lacking. Check out the plea for [Responsible Data Science. Ensuring Fairness, Accuracy, Confidentiality, Transparency \(FACT\)](#).*

Infrastructure: while not really a buzzword, this term is frequently used and conveniently vague. Think of it as the basic systems and services that an organization uses in order to work effectively. For a country this could include transport and power supplies, and for research for instance libraries, a clean room, data storage systems, single sign-on, and visualization tools. Keep in mind that an infrastructure can - or even should? - include human experts and support staff. For example, to make this explicit the OpenAIRE project calls itself a [socio-technical infrastructure](#).

Thing 3 □ EOSC projects

Clearly, knowledge circulated across national borders long before the notion of EOSC got traction. Browse for example through [20.000 letters that were written by and sent to 17th century scholars](#) who lived in the Dutch Republic. More recently, other ways of slicing through the European academic landscape have been and are being promoted with - often - European project funding:

- Research Infrastructures or RIs cater to researchers in one or a few related disciplines (or [see the formal definition](#)). Examples are [DARIAH](#) for arts and humanities and LTER for ecologists (**more about discipline-specific infrastructures in Thing 4: EOSC for research domains**) [ESFRI](#) (the European Strategy Forum on Research Infrastructures) is the linking pin of these RIs. It is a strategic instrument to develop the scientific integration of Europe and to strengthen its international outreach. Such RIs have been around for forty years, and this video sketches [the interaction between ESFRI and EOSC](#). You can call RIs discipline-specific or vertical, as opposed to...
- e-Infrastructures: projects like [OpenAIRE](#), [FAIRsFAIR](#), [EOSC-hub](#), and [FREYA](#) develop and provide digital services which are cross-disciplinary. Although [e-infrastructures](#) intensely collaborate with research communities or stakeholders like data librarians and publishers to learn about specific requirements they may have, the expertise they share and the services they offer are generic or “horizontal”. OpenAIRE advances Open Science in the broadest sense of the word and provides a platform which acts as a large-scale collection of research outputs easily discoverable and re-usable; [FAIRsFAIR](#) promotes everything FAIR (**see also Thing 8: FAIR in EOSC**); FREYA builds the infrastructure for persistent identifiers that connect researchers, data, grants, publications, and research organizations; and EOSC-hub lets you run large computational analyzes on a distributed network of computing centres (**see also Thing 6: EOSC services, in this document**).

By way of exercise, compare [research infrastructure](#) and [e-infrastructure](#) as defined by [Science Europe](#).

Thing 4 □ EOSC for research domains

When you work in a certain scientific discipline or domain, you probably share certain research methodologies, tools, or terminology with others in this domain. This is why research infrastructures can be very effective instruments for driving research: they offer digital services for a particular domain or a few related domains, and thereby focus on the commonalities. Many of them also provide training, for instance for early-career researchers, or on how to use the services (**see also**



Thing 5 about EOSC training).

Take for example the [Digital Research Infrastructure for the Arts and Humanities \(DARIAH\)](#). Its aim is “to support transnational researchers in all phases of their work: from data acquisition and analysis to publication and archiving. DARIAH and the tools it offers are designed to meet the needs of arts and humanities researchers working across Europe. For example, they could include a musicologist analysing digital recordings, an archaeologist digitally recreating ancient buildings or a historian studying digitized texts to investigate how place names change over time.” This is quite a [success story](#). Even at this already aggregated level, encompassing several disciplines, DARIAH recognizes that “collaboration with other research infrastructures is also supporting the growth of research communities: for example, with the [CLARIN research infrastructure for language resources and technology](#); [OPERAS for the development of open scholarly communication](#); and the [CESSDA consortium of European Social Science Data Archives](#).”

As mentioned in **Thing 3 on EOSC projects**, the domain perspective of Research Infrastructures (RI) is one way to slice through the EOSC-to-be. Especially when we’re open to using the services offered in other domains, by other RIs. And you don’t have to “think big”. An archaeological study from The Netherlands, carried out before the EOSC era, presents a nice, tiny example: one part of the results is preserved in the [4TU.ResearchData long-term repository for technical sciences](#), and the other in the [DANS long-term repository, which by then mainly catered to the social sciences and humanities](#). Can you find the two datasets and the study?

Thing 5 □ EOSC training

EOSC training can be several things:

- Training that supports the implementation and use of EOSC services. You can find out about such training by contacting the services or the infrastructure that provides them. There is more about central hubs for EOSC services and support including training in **Thing 6 on EOSC services**. Examples of training support connected to services: [Jupyter Hub training](#) or metadata-based data discovery [B2Find](#), IT service management [FitSM Foundation Training](#).
- Training that is more generic and supports Open Science, Research Data Management en data management planning. This training targets the research communities and is provided by research institutes or by projects like [FOSTER](#), and OpenAIRE Advance by their network of National Open Access Desks. Training for librarians and data support staff such as data stewards is provided by for instance
 - [CODATA/RDA summer school](#)
 - [RDNL Essentials 4 Data Support](#)
 - The FOSTER Open science training book [has been translated into Spanish and Portuguese as well](#).
- More domain-specific training is offered by research infrastructures and domains e.g.
 - ELIXIR’s [Training and e-Support System](#) for the life sciences
 - [#dariah Teach platform](#) for the digital arts and humanities and the [Digital humanities Course registry](#) (CLARIN and DARIAH)
 - [Data Management Expert Guide](#) for social scientists (CESSDA)
 - [SoBigData training registry](#) for big data and social mining
- Training of trainers. Sharing best practices on training in EOSC and beyond is supported by the [RDA Interest Group Education and Training on research data](#) and the [Community of Practice of training coordinators](#). These are both global networks of training coordinators and they organize regular online and offline events on training topics such as improving impact of training, improving visibility of training events and resources, skills and competences for open



science and data stewards, and training portals.

Thing 6 □ EOSC services

This Thing is about digital services; for training as a service, look at **the Thing above**.

When your institute and your research domain lack specific services that would support the research process, or when you're simply curious what's around, there is a wealth of services to explore. Here is an [easy introduction to services that support open and FAIR data](#) (webinar recording plus slides). The services are ordered along the phases of a simplified research data lifecycle.

The [EOSC-hub service catalogue](#) is an obvious place to go to, as long as you're aware that some services target researchers and research communities directly, while other services require administrator expertise. For instance, a "[solution to store and exchange data with colleagues and team members](#)" focuses on researchers, but "[a proxy service that operates as a central hub to connect federated Identity Providers \(IdPs\) with EGI service providers](#)" doesn't. In either case, though, you can rest assured that the service is robust: they all have at least a so-called [technology-readiness level](#) 8 (on a 1-10 scale), as demanded by the European Commission.

EOSC-building projects have started to jointly present their services. See for instance this use case about [complying with Open Science ambitions and the GDPR at the same time](#): how best to manage and share person-related data? OpenAIRE's [Amnesia anonymization tool](#) could help you.

By way of exercise, have a look at [B2FIND](#) and [Zenodo](#). Both are cross-domain repositories for research output. Consider their respective strong points.

Thing 7 □ FAIR principles

The FAIR Data Guiding Principles came into existence during a [workshop in Leiden in 2014](#) where a broad range of stakeholders in the field of research data management and stewardship came together to discuss the improvement of the reusability of research data. The [first official paper on the FAIR Guiding Principles](#), however, was only published in 2016. The FAIR Principles consist of a total number of [fifteen guiding principles](#), every single one of them related to either findability, accessibility, interoperability, or reusability of research data.

Soon the FAIR Principles attracted interest and were widely evaluated and used. Not only data stewardship was discussed on the basis of the FAIR Principles but they were also used as guidelines for scientific disciplines, research infrastructures etc. The FAIR Guiding Principles haven't changed since. However, the [FAIR Metrics Group](#) undertook effort to specify their measurement by defining metrics. After all, FAIR is not a black-and-white quality mark; on the contrary, it makes sense to distinguish between degrees of FAIRness. These metrics are stored on [Github](#); more context about their coming into existence and use is provided in the [paper that now is seen as the follow-up](#) paper of the first FAIR Guiding Principles paper in 2016.

The report [Cost of not having FAIR research data](#) aims to provide an estimate for the EU economy based on a series of indicators extracted from previous studies and analyzed with the help of interviews with experts in the field. Using quantitative methodology and very conservative assumptions, the analysis shows that the minimum cost for the EU is €10.2 billion per year, which will increase over the years if we do not take action. The question, however, is whether the costs of not having FAIR data can really be estimated and expressed in a number.

Today the FAIR Principles and FAIR as a concept are an indispensable part of the community that



deals with research data on a broad scale. Their importance is recognized on a high level which is visible in the uptake of FAIR in EOSC projects (**Thing 3**) and the e-infrastructure project [FAIRsFAIR](#) (**see the next Thing**). FAIR will be one of the main topics along with projects on FAIR in specific research areas such as [FAIRplus](#) and the [ESFRI projects](#) (**as described in Thing 3**).

Thing 8 □ FAIR in EOSC

Early 2017 it became clear that the European Commission had embraced the FAIR data principles. An innovative element of the Horizon 2020 grant scheme at the time was (and still is, in 2019) an [Open Research Data Pilot](#), asking funded projects to make the data underpinning their publications available or “Open” ([infographic](#)). As EC representatives put it in April 2017: “[We are now seeing openness as one component of FAIR data](#) and aim to address all of the FAIR aspects in Horizon 2020”.

By way of exercise: the relation between Open data and FAIR data can be confusing, and we know of researchers who struggle with it when they - have to - write their data management plan. [This publication explores how Open, FAIR, and research data management \(RDM\) connect](#): “The boundaries and intersections between RDM, FAIR and open cover important elements that risk being overlooked if we only focus on one concept.” Read the publication; do you agree with it?

The EC installed an expert group on FAIR data, which, after an international consultation process, delivered in 2019 a [report with 27 recommendations for turning FAIR into reality](#) (see slide 8 of this [presentation for an overview](#)). 15 of these are considered “high priority” and provide key concepts for FAIR Digital Objects and a FAIR ecosystem, which is then implemented through changes in research culture, technology and skills. Metrics, incentives and investment will drive culture change and implementation.

The FAIRsFAIR project contributes to the implementation of several of these recommendations. For instance, two work packages and [a competence framework](#) with several trainings address the skills-related recommendations “Professionalize data science & stewardship roles” (rec. 10) and “Implement curriculum frameworks and training (rec. 11). This is done by linking with other parties active in Open Science and FAIR such as the bottom-up initiative [GO-FAIR](#), Open Science promoting project [FOSTER](#), the Committee on Data of the International Science Council [CODATA](#), the [Research Data Alliance](#), and the [European University Association](#). (**see also Thing 5 on EOSC training**). FAIRsFAIR also addresses certification of FAIR services (recommendations 9 and 13), by strengthening the network of trustworthy digital repositories (**have a look at the next Thing, about trust**). Outcomes will feed into the work of the [EOSC FAIR Working group](#), which was set up by the EC. Furthermore, FAIRsFAIR will contribute to the [Rules of Participation in the EOSC](#), for which another EOSC Working Group has been installed. These rules will guarantee an open, secure and cost-effective federated EOSC with services of documented quality, taking into account specificities of scientific disciplines as well as the variety of service providers.

Thing 9 □ Scientific integrity and trust

The EOSC aims to make science more open. This aim for research openness is in itself a goal to be achieved, but in opening up science, multiple other things will be achieved as well. One important aspect of research output openness entails the research’ replicability. If data collection methods are explicit and publicly available, the gathered data can among other things be made ready for reuse. We talked about it **FAIR-Things 7 and 8**: making your data FAIR is crucial.

Replicability makes research output reliable and trustworthy. This Thing of scientific integrity and



trust surrounds just that: by making your data FAIR, for example in the context of the EOSC, you disseminate your research methods, hypotheses, assumptions and conclusions, making clear the thought process behind your research. In doing so, you enable fellow future researchers to learn from what you have done.

The blog [Retraction Watch](#) collects examples of bad practices in the world of research data. But how do we prevent those described cases to happen? One method is using Trustworthy repositories, that help to **make** and to **keep** data FAIR:

“make”: by providing a persistent identifier, supporting metadata standards, supporting findability through their public catalogue, providing clear licences

“keep”: by preserving the data and keeping them usable in the long run because they know about sustainable file formats. In the [Guidelines on FAIR Data Management in Horizon 2020](#), the European Commission states: *“Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.”* (cited from [the OpenAIRE initiative](#)). One way to make sure that your data is stored at such a certified repository, is to look at the CoreTrustSeal certification.

The EOSC aims to create a [trusted environment](#) for the storage, processing, and reuse of research data. However, FAIR tells us something about the research data itself that are being processed. Additionally, we would want to store this FAIR somewhere trustworthy. So, trust and FAIR go hand in hand.

By way of exercise: watch this [video tutorial about FAIR data in trustworthy repositories](#). Do you agree with the recommendations? And which one of the [CoreTrustSeal requirements](#) do you think are most important?

Thing 10 □ Open science: the rest of the world

All this talking and [_Thinking_](#) about EOSC and Europe: there are clearly no [Schengen](#)-like borders around it. Science and researchers have many ways of organizing themselves through research infrastructures and collaboration initiatives, and the EOSC is not the only one. Let’s close off our 10 Things with a wider look: how about the rest of the world? Here are some examples of the same scale (or even bigger!) as the EOSC. This is obviously not an exhaustive list:

- In Latin America, there is the [_Consejo Latinoamericano de Ciencias Sociales_](#) (CLACSO). Their organizational structure is similar to that of the EOSC: [about 700 institutions from 52 countries](#) collaborate and cooperate in empowering open science on their continent. The main drivers of this initiative are universities and government organizations.
- [The UNESCO Global Open Access Portal \(GOAP\)](#) is funded by the Governments of Colombia, Denmark, Norway, and the United States Department of State. The portal presents a current snapshot of the status of Open Access (OA) to scientific information in [158 countries](#) worldwide.
- Consider the [Open and Collaborative Science \(OCS\)](#) as well, organized by [the Global South](#) countries - a network similar to CLACSO. The organization has formulated four specific goals in the framework of Open Science.
- On top of all these publicly funded initiatives, one example of a [not-for-profit](#) one is the existing [Open Data Science Cloud](#), part of the Open Commons Consortium. Their cloud provides infrastructures and tools for researchers to analyze terabytes and petabytes worth of datasets.



Source: <https://librarycarpentry.org/Top-10-FAIR/2019/05/31/eosc-fair/>

