

Module 5: Integrating resources into EOSC: What, why, and how?

“Integrating resources into EOSC: What, why, and how?”

Hello, my name is Matti Heikkurinen and I am a Senior Strategy and Innovation Officer at EGI. In this video, I am going to explain the concept of “service onboarding”. However, I think “EOSC integration” is perhaps a more intuitive term. In the end, the goal of onboarding is to make your services available and more broadly usable by integrating it with the European Open Science Cloud (EOSC). So, I will use the term “integration” in the rest of this module, but when reading related material it is good to remember that in this context “onboarding” and “integration” can usually be considered synonyms.

“Learning outcomes”

By the end of this module, you will understand the nature of a federated structure, such as EOSC and the role of IT service management in making such structures more robust, and – last but not least – you will be aware of the integration process to bring new resources into EOSC.

“Introduction”

So, let’s start with the “why”. In a nutshell, well managed EOSC integration can be seen as a **vaccination against customers dragging you to your laptop in the middle of the holiday**, because they can’t use your services.

The effectiveness of this vaccination was analysed in a “clinical study” by Gartner research. The result was that 80% of all the major disruptions in IT service operations were due to “people and processes” issues – instead of failing software or hardware. In other words, doing EOSC integration properly means that you will be spared from four out of five complaint calls. And this is the minimum effectiveness: usually a portion of the remaining ones will be handled by EOSC staff (helpdesk, first level support etc).

So, the bottom line is, a proper EOSC integration process will considerably reduce the time and effort you would normally have to put into firefighting – dealing with upset users – when offering your services to new audiences.

“EOSC as a federation”

The word “federation” may sound archaic or pretentious - somehow abstract when all your users want to do seems to be research. “Federation” is also not something your users would ask about explicitly - they just want to perform research in order to publish articles, create startups or innovate otherwise.

However, if you’re a service provider and don’t want to bring your laptop to the beach, the concept of “federation” and its implementation have some important implications. When users “just want to do research”, your service or resource is usually just one piece of the puzzle. The input data may be a result of complex processing of several data sets from different organisations. This processing may require access to a specialised computing service from another provider and results may be passed on for further processing, using another set of tools and resources. If any of the components of this chain break down, your resource might be seen as being the one “at fault”. Thus managing federation in a consistent manner is an important requirement - even if it tends to be hidden from view.

EOSC is emerging as an interdependent network, where the links between service providers can be formed dynamically - or even automatically. This is made possible by lightweight core components and services that support bringing new services to the EOSC ecosystem and coordinating them. So we could say that EOSC has a “hidden center” that addresses the “hidden requirements” related to federation of resources.

These kinds of federated ecosystems are quite common outside the IT service provision, and usually the level of integration determines the consistency of user experience. The loosest level can be like a hotel star-rating: there is usually some correlation between the number of stars and the actual quality of the experience – but you will almost certainly read the reviews before booking a hotel.

Somewhere in the middle of the integration scale, you might go and get your insurance from an agent or broker, with standardised “negotiation phase”, followed by a provider-specific delivery. In the most integrated case, for example, buying a computer or a car, all the complexity of supply chain management and after-sales service is hidden from you. Branded service centres provide after sales services in a uniform way anywhere in the world.

To genuinely facilitate open science, EOSC needs to be a quite tightly integrated federation. In addition to finding services in the same place (a portal), the user should be able to trust that the quality of the service and the available support will fulfil common, minimal standards.

To fulfil this requirement, it is necessary to consider resource integration as a series of steps that make your resource a part of the highly integrated federation.

“Integrating services or resources”

First a couple of words about the distinction between service or a resource. This used to be an important distinction when - for example - a dataset was usually tied to a physical object: hard drive, storage tape or something similar. Making sure they were Findable, Accessible, Interoperable and Reusable (FAIR) was at least to a degree similar to making sure a library book could be found, accessed and used. Managing a physical copy of a dataset was seen as fundamentally different from managing a service (for example, a service providing weather predictions by crunching weather station data more or less directly). However, this distinction matters much less now, as datasets have - for the most part - become accessible online and services become more and more dependent on large datasets.

Practically everyone considering integrating a service or a resource into EOSC is breaking new ground by dramatically extending the potential user community. The challenge is documenting the new resource that it works in a predictable manner in a completely new platform, supporting completely new use cases.

EOSC IT Service Management System (SMS) provides a common framework that helps achieve this goal. It is based on an open standard called FitSM that provides common concepts, requirements and other support components that can be used to improve the perceived quality of services.

In practice, it is not necessary to read the whole standard: the EOSC integration process focuses on two things:

- Describing the function of a service or a resource in such a way that potential users can find and understand it.

Bringing synergy to better data management and research in Europe

- Identifying key roles and processes - such as helpdesk or security contact points - that are specific to the new service. These are the key interfaces between the SMS and the newly integrated resource that make it possible to benefit from the common quality assurance framework.

This process may be difficult to grasp as the management of the resource before EOSC integration has relied on tacit knowledge. Everyone just knows who to ask for help, who to report errors to and who to call in case it seems there are security issues that need attention. It may not have been necessary to document contact points for helpdesk, service owner - or develop resource descriptions that wouldn't rely on shared historical knowledge among the developers.

So, IT Service (Resource) Management documents the complexity, creates standardised installation instructions, collects knowledge about recurring problems and workaround. It will also encourage standardisation of the ways the new service deals with problems while also bringing in dedicated staff to provide support and do troubleshooting.

This is one of the primary mechanisms that allows EOSC integration to act as a vaccine against the phone calls.

“The resource centric view”

As noted earlier, services and resources need to be dealt with in a very similar way.

However, historically there are two schools of thought when it comes to describing EOSC:

1. Everything is FAIR data (meaning “Findable, Accessible, Interoperable and Reusable”)
2. Everything is a service

Even though - as we noted before - this distinction matters less today, the origins will influence the way the resources are usually described and what are the most relevant data items for them. However, the high-level process should be similar.

In the end, the most pragmatic approach is to make distinction between publications, software, and training resources.

In general, the common sub-categories discussed are services, research products or training resources. But if we abstract them into resources and differentiate only where and when necessary, we can use a...

“The (high level) process for connecting resources”

...common high-level process to connect resources to EOSC. The current model is based on the following steps:

- Registering an account for the provider and submitting the provider for inclusion in the group of EOSC providers.
- The checks are basic sanity checks: does the provider look like it is suitable for EOSC, can we verify the information (e.g., address, registration of a legal entity,...).
- Once the provider is registered, the resource needs to be described and submitted for validation.
- And once the resource is validated (e.g., against the Rules of Participation criteria), it can be listed on the marketplace.

- After this, add-on components – like ordering – can be added. The ordering is relevant mainly for resources that can't be used simultaneously by multiple users, for example, a virtual server instance.

“EOSC Profiles”

This unified path is made possible by consolidating the different resource description standards into a single set of EOSC profiles. They are published under creative commons (available through the links listed on the slide) and intended for use in EOSC, but also in other registries that want to use the tools to build their own federations.

Sharing of the approaches and tools means that these “community federations” will have an option of being interoperable or linked with EOSC, since the technical EOSC interoperability is already built in.

As noted in the green boxes, the published profiles include both the provider and resource profiles, so they cover both of the phases discussed in the previous slide.

“Provider profile”

The provider profile includes the information that is shared by all the resources integrated by the provider. These include:

- Basic information: name, website, legal form and so on
- Marketing information: short description and logo
- Classification: flagging whether the provider is an EOSC provider, focused on specific scientific disciplines; it can also include free-format tags
- Location: in essence, physical address information
- Contact information: emails and phone numbers
- Maturity: list of certifications (such as ISO 20 000) and life cycle status (essentially, under construction, operational or being upgraded) – this allows including providers early, so once the services are available the integration process will be quicker.
- Other information can be used to position the provider by listing e.g., participation in major EOSC projects or status as an ESFRI project participant.

Resource profile

The resource profile is similar to the provider profile, sharing some of the components, such as basic, marketing and maturity information. However, some of the items are more complex.

For example, the contact information needs to include details, such as helpdesk and security contacts. Maturity information needs to be more fine-grained to cover phases, such as limited pilot use or gradual retirement (e.g. no longer offered to new users).

The dependencies are also a new aspect, since using the resource may be possible only if the user has access to other EOSC services.

In general, resource profile is the area where the complexities involved in the integration become apparent: extending the user base and making the resource available in a completely new platform context means documenting things that most likely were not documented before. However, the process has recently been automated, speeding things up.

Bringing synergy to better data management and research in Europe



Wrap-up:

This concludes the module introducing EOSC resource integration. You can find a number of relevant links and in-depth resources on this subject in the slide deck.