

31 Jan 2019 | News

Europe wrestles with a plan to build the 'Amazon of science'

[Science Cloud](#)[News](#)[Events](#)[Reports](#)[Become a member](#)

The European Open Science Cloud is a giant effort to provide a single point of access to all scientific data. But getting all the infrastructures to integrate and engendering a culture of sharing is a daunting task, say those involved in its creation

By [Éanna Kelly](#)



The EU hopes it has found the way of the future for research, with its plan to create an enormous virtual repository providing access to the collective data from all publicly funded research on the continent in the European Open Science Cloud (EOSC).

The ambition is to seamlessly link the electronic resources of research institutions and university libraries and make them accessible through a single portal.

The European Commission, which is putting €600 million into the early stages of the multi-billion euro effort, says the cloud will transform research, with scientists

having a vast collection of resources on tap.

It certainly ranks as one of the most ambitious projects in science today, requiring technological wizardry, a lot of money and coordination. But that should not daunt anyone, Juan Bicarregui, one of the leading public faces of the project, explained at a conference held in the Royal Geographical Society in London on Wednesday.

"It's not a case of throwing away everything we have, and starting anew. That would be infeasible. We have a whole lot of working infrastructure already and we're going to build on that and help it interoperate. It's an incremental process," said Bicarregui, who is head of data division at the UK's Science and Technology Facilities Council, and member of the [11-person EOSC Executive Board \(https://sciencebusiness.net/science-cloud/news/eu-picks-team-11-run-giant-european-open-science-cloud\)](https://sciencebusiness.net/science-cloud/news/eu-picks-team-11-run-giant-european-open-science-cloud) set up by the commission to lead the project.

EOSC, announced in April 2016 and set to start in 2020, will be a cost-saving efficiency and an equalising force, its supporters say. It will amount to a federation of existing and future research data infrastructures, in which access to data and data processing services is granted via "a federating core." The aim is that scientists will be a few clicks away from access to the vast stores of data from any lab or any scientific discipline in Europe.

Bicarregui suggested the final product should look something like the world's largest online retailer. "Look at the way we buy something from Amazon. I can have a selection from many different suppliers and I hardly notice who the suppliers are," he said.

Other analogies are available. "There are as many theories of how the EOSC should look as there are people working on it," said Jonathan Taylor, head of data management at the European Spallation Source.

One delegate suggested it would be like Airbnb. Another that it could be like a car insurance comparison website, sharing some features of Uber, then quickly corrected himself, "Oh they've been getting bad press recently, so let's not say Uber."

The animating idea behind EOSC is that the rate of new scientific data being generated is exploding and out of control. As things stand, no two research

facilities have a common approach for managing this.

“Modern research requires you to have access to multiple infrastructures,” said Andrew Smith, head of external relations with ELIXIR, a European research infrastructure which manages biological data. A researcher in Italy, for example, studying brain function, might need a model based on fish genes collected in a lab in Portugal, he said.

Curating data

Lots of public labs now run cloud services for researchers, either in house or through private sector cloud providers.

The modern lab needs a robust data policy, and has to worry about curation “early”, Taylor said. “Working without a data policy is a minefield”, and can lead to fights down the line about who owns data. “And it’s no good collecting 10 petabytes of data [if] you can only store a half a petabyte.”

Labs are concentrating on managing their own data, or collaborating with one another on specific themed projects, rather than on combining their electronic resources into a single online access point.

Now, with funding from the European Commission, some of those individual efforts in fields like astronomy and physics are beginning to dovetail.

But a project like EOSC is clearly a bigger challenge for Europe, with its vast, disparate and multilingual research efforts, than it would be for American states working in a single language.

In addition, everyone is starting from a different level. “It’s like Strictly Come Dancing, where you have one professional dancer and one new dancer trying to perform together,” said Ron Dekker, director of the Consortium of European Social Science Data Archives in Norway.

A big to-do list

A range of practical matters — legal, financial, technological — need to be resolved before EOSC gets off the ground.

The Commission says it will support the core functions of the plan up to 2020. After that EOSC could be financed by a mix of funding, including fees from national funders and revenues from users.

The business model is a big question, but aside from that, "We need to worry about data stewardship, ethics and data protection," said Bicarregui. "It's very nice to say we're going to be open and have fair access, but there are going to be some constraints."

Rules will need to be established to clarify the roles and responsibilities of the funding agencies, the data custodians, the cloud service providers and the researchers who use cloud-based data.

For Susan Daenke, a structural biologist and coordinator at Instruct, a European research infrastructure, thinking about the effort required is overwhelming.

It took two years for her own lab to develop communication protocol for a "seamless interchange of data" with a synchrotron facility. "If you talk about getting all the infrastructures to integrate with EOSC, that daunts me," she said. "It's several orders of magnitude above what we tried to do."

Bicarregui agreed that with a very large number of projects are contributing to the final EOSC piece it, "could become a very confusing picture unless there is proper communication."

"Synchronising is the issue: we have to learn how to talk to each other," concurred Rudolf Dimper, head of the technical infrastructure division at the European Synchrotron Radiation Facility. "We will have to learn how to talk to the European Commission and the various e-infrastructure. It will hit limitations. We cannot talk to everyone."

Layered on top of this is the issue of achieving a greater standardisation of data. Without that, the whole effort will be meaningless, scientists say.

Experts propose using the 'FAIR Data Principles' of findability, accessibility, interoperability, and reusability. The data management and stewardship principles enhances the ability of machines to automatically store, find and use data, in addition to supporting its reuse by individuals, they say.

"If we don't have FAIR data, you can forget about the EOSC," said Dimper. If today, no publicly-funded science infrastructure in Europe completely lives up to FAIR principles in its data handling, that is because, "implementing FAIR data is a really difficult job," he said.

The terms of how EOSC will be accessed, whether free, paid, or with embargoes on with other usage restrictions are yet to be resolved. Although the EOSC is intended to make research data free at the point of use for scientists, commercial entities could be required to pay for access.

The proposal will also have to contend with political hurdles. "Ministers know they have to do something on EOSC, but they're not clear on sustainability. It's the work of the EOSC to put this [plan] into the hands of the decision makers," said Natalia Manola, the managing director of OpenAIRE, an organisation that researches and advises on open science policy.

This is one of the many moving pieces. "We're not in the business of solving sustainability this afternoon," said Philippe Froissard, deputy head of the research infrastructures unit in the Commission's research directorate.

Enlisting researchers

Perhaps the hardest task for the people driving the EU initiative is to convince scientists to open up access to repositories and share their data with others.

"EOSC is hardware, but it's also about people, and getting them organised," said Dekker.

Resistance to change is visible in science as in any other field. "We've spent 20 years corralling our researchers to standardise their data and share it," said Massimo Cocco, director of research at the National Institute of Geophysics and Volcanology in Rome.

Culture, not technology, is the greatest hurdle to EOSC, agreed Bicarregui. "I think the hardest thing is the change of attitude. Researchers work for curiosity and recognition. Today we value the paper more than the data. If data sharing were more recognised, then I would want others to use my data. It should be a good thing, I shouldn't feel scooped, I should feel complimented."