



EOOSCsecretariat.eu

Setup and management of the EOOSC Secretariat supporting the EOOSC Governance

EOOSC T3.6 Researcher Engagement Collection of Services

TU Wien: Paolo Budroni, Katharina Flicker, Juliana Giroletti, Andreas Rauber (PI), Barbara Sánchez, Bernd Saurugger



EOOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOOSC-2018-4, grant Agreement number 831644

Architecture: Collection of Services

- **Services to enable the automatic recording of provenance metadata for data, computation, and processes:** Researchers need the context of data / metadata in order to have trust in its quality. If there is no trust in the data, it won't be used for the research.
- **Services for data capturing:** EOSC shall offer services that allow for the documentation of how data is gathered. It should make (meta) data more searchable, findable and trackable and it should support many different formats. Such services can even include cryptographic verification of the value of data (cryptographic signatures on the metadata on the data being processed).
- **Services that control the upload of data:** Algorithms can help finding out, which data has to be uploaded (e.g. automated duplicate finders).
- **Services to track provenance:** In order to determine the value and usefulness of code and other research output, it is essential to know how it has been created, modified / manipulated, and (re-) used.
- **Services for data visiting:** Such services need to support the negotiation of access of data automatically. Trust analysis algorithms need to be established so that they are trustworthy, and the leak of information can be prevented.

Architecture: Collection of Services

- **Services for machine translation of research outputs:** These services shall offer human to machine and machine to human translations as well as voice-to-voice translations but require rules on the usage of the data.
- **Services for horizontal translation and communication adaption:**
 - Translation services to **communicate research outputs** to policy makers, to the public and across disciplines, e.g. services for automatic metaphor translation
 - Services to **translate scientific concepts** and explanations for different levels of expertise within a discipline, e.g. services for math translation explaining some mathematical concepts for specific research questions
- **Services for advanced automated programming:** From graphical programming, workflow re-composition to fully, or semi-automatically process definitions by machines.
- **Services for ontology mapping:** (Semi-) Automated services that provide mappings between and integration of ontologies across domains.
- **Services for automated copy editing:** Automated copy editing as a service can improve the accuracy, readability, and fitness of papers and ensure that it is free of errors, omissions, inconsistency, and repetitions.

Architecture: Collection of Services

- **Services for Trusted Validation processes:** Researchers will refrain from using datasets that they do not trust. Thus, EOSC must find ways to guarantee the quality of data.
- **Micro services** such as **data repositories, pseudonymization, anonymization services** and **semantic processing services** for research data (interoperability), **cloud-services for data analysis** and **Open Science Schools**. In addition, the need to reproduce data in order to verify research results was mentioned
- **Services transparently measuring the impact of research:** EOSC needs to provide services that capture and aggregate a range of advanced metrics for research output beyond papers (data, code, re-use, ideas, proposals, mentoring...). These need to be transparent and verifiable.
- **Services for trusted collaborative editing, collaborative writing, coding, collaborative developments services:** Such services already exist. Yet, all such data is collected centrally with little control over its use. Trusted services that ensure data is encrypted and only accessible to identified participants are essential. Even the use of metadata (such as collaboration networks or timing of use) has to be limited to quality of service improvement.

FAIR: Collection of Services

- Establish **reward mechanisms** as incentives for researchers to provide access to their data, code and other research outputs. Open science as a key enabler for high quality research (Note: access includes specifically access for / by machines, not just humans).
- **Services to identify components of research**, from data, processing tools, workflows, interim results, to final results and outputs: PIDs (e.g. DOIs) can be used to identify different parts of analysis, or even different aspects of the data.
- **Services to regulate access and enable open access:** EOSC shall enable the freedom to use data whenever possible. However, EOSC shall also enable the regulation of access to data, because not all data can be open (e. g. researchers, their students and others who need certain datasets may ask to get access. Thus, data is not open in general, but is accessible when it comes with a research request)

RoP: Collection of Services

- **Services to conform to Rules of Participation (RoP) that are transparent:** RoP need to be clear on how to get access in terms of money and give a definition on who is actually allowed to use specific datasets
- **Services to conform to rules to impede market dominance:** e.g. rules are required to prevent quasi-monopolies in service provisioning across the entire EOSC ecosystem (i.e. ranging from infrastructure services such as cloud storage or high-performance computing via generic or domain specific services such as collaborative editing tools or genome sequencing, to higher-level services such as impact assessment of research or certification).

Skills & Training: Collection of Services

- **Guidelines, trainings** (how to use certain services) and **manuals** (Git, Repos, VM) should be offered to researchers as well as to their students. Working efficiently is impossible when one lacks the know-how of the tools one is supposed to work with.
- IT-support and **services for specific teaching challenges**

Sustainability: Collection of Services

- **Knowledge Management Systems** have to be established in order to make better use of what is there already as well as in order to avoid redundancies. When involved in projects some effort should be put in their evaluation and summarizing lessons learned. Working open access on a daily basis should also be part of efficient knowledge management.
 - **Services that enable them finding experts and already existing data quickly** – not only in research, but on basically everything they need on a day-to-day basis to get the nothing-to-do-with-research-part-of-their-jobs done efficiently (“**yellow pages**” on a university level) as well as to find out about potential collaborations
 - **A search catalogue for device inventory:** in many cases devices and equipment are acquired for (student) projects and are stored and put away soon after the project ends. With a catalogue such as this researchers and other university staff would have a means of finding out what’s there already and to re-use the devices and the equipment in an efficient way
- **Services for machine actionable scholarly knowledge sharing:** machines need to be able to search for, find and act upon scholarly knowledge across all types of research outputs. Examples include concepts such as Nano-publications (nanopub.org) and infrastructures such as Open Research Knowledge Graph (orkg.org), or services to find data by properties (e.g. distributions) or code segments.

Sustainability: Collection of Services

- **Services for research promotion to increase the impact of science in society:** EOSC is not only about data, or data infrastructures. It is also about laying the foundations for the production of knowledge and knowledge transfer (teaching, mentoring, and organizing events). Thus, services to promote science and increase its impact in society are needed (e.g. services such as “Google Docs for EOSC” or “GoToMeeting for EOSC”)
- **Services to identify all research output produced:** such services exist already, but EOSC must enable higher transparency. Business models must allow these services to be consumed free of charge at the point of use, i.e. by researchers
- **Services for automated deletion processes:** as it is impossible to store all data, automated deletion processes need to be decided, considering aspects such as (expected) value of data, feasibility of-re-creation and abstraction.

Others: Collection of Services

- **Services to collect questions that need scientific answering:** researchers have to have a say in defining grand challenges of society. Inclusive definition processes have to lead to a list of urgent research topics that need to be addressed within the next decades at local and global level. Against this background, a database to collect questions that need scientific answering is crucial
- **Services to collect a wish list of datasets that researchers consider essential to face the grand challenges of society:** together with a database to collect questions that need scientific answering, there should be another one to collect a wish list of datasets that researchers consider essential to face the grand challenges of society. This strategy would also help to pass from individual research efforts to collective efforts, as it would potentiate collaboration among research groups. Within this framework, EOSC should also work to incorporate datasets of interest for researchers (for instance, data from satellites, climatic data, clinical datasets, etc.)

Others: Collection of Services

- **Checking forms and formal criteria, observe ethical guidelines, support human resource management:** All of these tasks are time-consuming. Thus, providing (automated support for) such services saves time for actual research
- **Services for detecting, evaluating and monitoring upcoming AI solutions:** AI is in rapid development. Technical and human-in-the-loop services need to be established within EOSC to monitor these and identify promising solutions for fast but controlled and tested inclusion in EOSC.
- **Virtual team meetings:** Virtual environment services or virtual reality solutions for meetings, supporting sophisticated interaction between researchers shall be developed, in order to overcome the current barriers of video conferencing and remote interaction.