

# A collection of services to perform cutting-edge research in Europe and beyond

The [EOSC](#) initiative aims at supporting more than 1.7 million researchers and boosting interdisciplinary research in Europe. To understand better what the research community needs, the [EOSC Secretariat](#) partner [TUWien](#) is organizing a series of [workshops](#) and an [exploration series](#). The objective is to collect visions, needs and requirements for (future) research environments as well as services that support cutting-edge research. This document contains a list of all services collected so far<sup>1</sup>. It will be shared with the community, and feed directly into the work of the [EOSC governance bodies](#) and [EOSC Working Groups](#) (WGs), providing input crucial for the development of the EOSC.

## 1 Metrics and Evaluation

Service	Description
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, we need services to promote science and increase its impact in society (e.g. services such as "Google Docs for EOSC" or "GoToMeeting for EOSC")
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 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 identify all research output produced by one institution, researcher, network...	Such services exist already. However, 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 to 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.

<sup>1</sup> The full reports as well as key takeaway messages are available at <https://zenodo.org/record/3701194#.X1de7HkzaUk>, <https://zenodo.org/record/4015121#.X1deLHkzaUk> and <https://zenodo.org/record/3701269#.X1dfWXkzaUk>

## 2 Communication and Language Barriers

Service	Description
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 & communication adaptation	<p>Translation services to communicate research outputs to policy makers, to the public and across disciplines, e.g. services for automatic metaphor translation.</p> <p>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.</p>
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.
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.
Services for 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.
Services that enable 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.
Services for knowledge brokering	Such services facilitate two-way or multiway exchange of information, bridge gaps between knowledge producers and knowledge users and build connections between different audiences. Thus, they also promote research output, increase research impact in society and steer research towards a more comprehensive understanding of a problem.

### 3 (Research) Support Services

Service	Description
Services for checking the forms and formal criteria	All of these tasks are time-consuming. Thus, providing (automated support for) such services saves time for actual research.
Services to observe ethical guidelines	
Services to support human resource management	
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.
Services to support human peer review processes	AI approaches have value in supplementing and reinforcing human peer review in terms of e.g. running data/software checks, checking biases/reporting standards and potentially some elements of reproducibility. Thus, they could assist in quality control, give access to primary data and support trust in the conclusions derived. Eventually, it is crucial to ask if machines could help to overcome human biases and thus would be better at peer reviewing as algorithms can be inspected and evaluated, offering - in principle - higher transparency than human decision making processes.
Services to support robust research by providing access to primary data	Access to primary data is crucial in order to reproduce research results. Thus, EOSC should provide researchers with the means to access primary data and provide services that support reproducibility.
“Human expert infrastructure” services	Scientific expert programs support researchers, who can contact experts in e.g. statistics, programming, data stewardship, and many more, if questions and requirements outside their core domain expertise should arise.
Incentives for and recognition of highly focused studies	Drawing conclusions from big data does not necessarily enhance science or scientific output, because big data also comes with higher complexity. That might lead to a higher false discovery rate. Thus, while having access to big data is useful it is also crucial to have support in EOSC to conduct highly focused studies.
Micro services that support researcher’s day-to-day routine	Such micro services include data repositories, pseudonymization and anonymization services and semantic processing services for research data (interoperability), cloud-services for data analysis and Open Science Schools.

## 4 Supporting trust in EOSC and data quality

Service	Description
Services to regulate access and to enable open access, when possible	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 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)
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 help with checking and verifying data quality	<p>The trustworthiness of data has to be ensured by all means. However, trust requires understanding where data comes from. Consequently, there needs to be a transparent framework to understand the data. In general, provenance and data quality, has to be checked or, at the very least, checkable.</p> <p>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 the many different formats. Examples are services to enable the automatic recording of provenance metadata for data, computation, and processes, services for data capturing (that can even include cryptographic verification of data like cryptographic signatures on the metadata on the data being processed), and services for trusted validation processes<sup>2</sup>.</p> <p>The latter is crucial because researchers will refrain from using datasets that they do not trust. Thus, EOSC must find ways to guarantee the quality of data.</p>
Services for secured and monitored data visiting	Such services need to support the negotiation of access to data automatically. Trust analysis algorithms need to be established and leakage of information needs to be prevented.

## 5 Supporting Sustainability

Service	Description
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.

<sup>2</sup> A more detailed description of these services can be found here: <https://zenodo.org/record/3701194#.Xu0fHWgzaUk> (DOI 10.5281/zenodo.3701194)

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.).
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.
Initiatives to optimize data uses that are supported by citizens	In order to derive increasing societal benefits from the use of data, data has to be taken away from those with an immediate (economic) interest in data (such as research institutions and health systems) and put into a neutral trusted body that only benefits from data being used. Thus, citizens are needed to act as their own hub and as a trusted third party. Such civic data cooperatives are not about radical changes but about developing a culture that starts to think more as a group or as a system.
Services and funding mechanisms to facilitate research collaborations within small(er) groups	Research collaborations and research consortia facilitate pooling resources in order to achieve common goals, but there are limits to this kind of joint efforts: involving and synchronizing all key players efficiently is not possible, if consortia are too big. Thus, research collaborations within smaller groups must be facilitated and funded.
Knowledge Management Systems	<p>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.</p> <p>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.</p> <p>Knowledge Management Systems may include services that enable experts and already existing data quickly, or search catalogues for device inventories.</p>
Services and organizational incentives to connect various (health) systems	<p>Connecting various (health) systems with each other is key in order to understand data-intensive systems better because it allows bringing in different communities and considering socio-economic deprivation. This is difficult to achieve because it requires organized efforts of society.</p> <p>National trusted nodes, however, are a good starting point to connect different health systems by working on specific topics, aligning with each other, flexibly agreeing on e.g. protocols on the technical, legal and trust level and – by doing so – enabling an organic growth of research institutions, health provider networks.</p>

## 6 Legal Frameworks and Rules of Participation

Service	Description
Services need to conform to RoP that are transparent	Rules of Participation 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 need 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).
Rules to avoid quasi monopolies and dominance	Control of market share and ensuring that no contributor of services becomes dominant in size or too-big-too-fail is crucial. In combination with mechanisms countering lock-in into specific services or providers, the EOSC should ensure resilience and flexibility by guaranteeing that a diverse set of actors are providing services in a transparent and cooperative manner.
Rules, protocols and verification mechanisms to avoid vendor lock-in	<p>It is crucial to avoid vendor lock-in and vendor dominance, because data intensive systems must be able to interact. Thus, interoperability must be enforced: Tool chaining that allows for flexible processing workflows, or interoperability sessions can counteract vendor lock-in. Such Interoperability sessions can be scheduled on a regular basis (e.g. biennially) for three to five working days and oblige computing vendors to get together to demonstrate tool chaining the lossless exchange of data and prove their interoperability.</p> <p>The growth from a bottom-up network in combination with basic GDPR rights to data portability and the duty of transparency and cooperation for each participant support the interaction of systems</p>
Rules and services to counteract data colonialism	Counteracting data colonialism is a political imperative. Potential solutions must involve society and need to be based on trust and mutual aid. Thus, it is crucial to have the data work for residents and do better for society with what is already there. Implementing consortia of actions and protocols, impeding vendor dominance, establishing basic GDPR rights to data portability, duties of transparency and cooperation for each participant, or relying on civic data cooperatives are ways of building up trust and increase access to health care data with an active consent of citizens in the long term.
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.

## 7 Trainings

Service	Description
EOSC@School / Data Literacy	People need to have competences in reading data, evaluating data, classify information correctly, or assess whether data is suitable for addressing a specific task. Thus, data literacy basics should be anchored in curricula early on and form an integral part of EOSC training at all levels of education.
Guidelines, trainings on how to use (EOSC) services and manuals	Such trainings and services need to be offered not only to researchers, but also to students. Working efficiently is impossible when one lacks the know-how of the tools one is supposed to work with.

*TUWien representatives of EOSC Secretariat: Paolo Budroni, Katharina Flicker, Juliana Giroletti, Andreas Rauber (PI), Barbara Sánchez, Bernd Saurugger*

*Date of Release: 21 September 2020*

