EOSC secretariat.eu

Setup and management of the EOSC Secretariat supporting the EOSC Governance

Report on the Workshop "Co-creating the EOSC: Needs and requirements for future research environments"

Lead Partner:	**TU Wien**
Contributors to the report:	Akgun, Omer Can (Delft University of Technology) Bazilinskyy, Pavlo (Delft University of Technology) Budroni, Paolo (TU Wien) Demoulin, Baptiste (Université Catholique de Louvain) Dias, Frederic (University College Dublin) Flicker, Katharina (TU Wien) Gazzola, Valeria (Netherlands Institute for Neuroscience) Giroletti, Juliana (TU Wien) Grossi, Valerio (ISTI - CNR Pisa) Günes, Murat (Université Paris-Sud) Leidel, Sebastian (University of Bern) Leppert, Linn (University of Bayreuth) Mangeney, Juliette (École normale supérieure de Paris) Nogues Gonzalez, Isabel (CNR - Research Institute of Terrestrial Ecosystems) O'Neill, Gareth (Leiden University) Ramadan, Husam A. (University of Sussex) Rauber, Andreas (TU Wien) Rezaei, Mohammad (Johannes Kepler University Linz) Rychnovská, Dagmar (Institute for Advanced Studies Vienna) Sanchéz Solís, Barbara (TU Wien) Saurugger, Bernd (TU Wien) Steiner, Ullrich (Adolphe Merkle Institute) Stocker, Markus (TIB - Leibniz Information Centre for Science and Technology University Library) Susi, Toma (University of Vienna) Verde, Licia (Institució Catalana de Recerca i Estudis Avancats)
Version:	1.1
Date:	9th March 2020
Status:	Version agreed with workshop participants
Dissemination Level:	Use for publication and distribution

Abstract

The report summarizes the workshop "Co-creating the EOSC: Needs and requirements for future research environments" that was held on 13th January 2020 in Feldkirch, Vorarlberg, Austria. TU Wien invited top-level researchers to discuss their needs and requirements for their research environments and identified several services that EOSC shall offer researchers as key stakeholders. Thus, this report lists these requested services.

TABLE OF CONTENT

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

Report on the Workshop "Co-creating the EOSC: Needs and requirements for future research environments"

1. Is the EOSC to researchers, what the elephant is to blind men?

There is a parable that, in connection with EOSC, has been used a lot. It is the story of a group of blind men who get to touch and explore an elephant for the very first time in their lives. All of them get to feel different parts of the elephant and start to "picture" the animal from their very own specific perspectives. Each of the men describes his experiences and, of course, these experiences differ greatly. If the elephant stands for the EOSC, do the blind men stand for research communities? Is the EOSC to researchers, what the elephant is to the group of blind men? Yes, because there still is no answer to what the EOSC really is, as it is still in co-creation process. Thus, there are still many directions, in which this initiative can develop. No, because, we as representatives of research communities already have explored some parts of the elephant in greater depth: We are experts in our scientific domains, we know our current research environments and we know what we need to enhance science. Even though we might not know what the EOSC is exactly, we can envision what we want it to be.

In addition, others came before us and before the EOSC initiative was launched. Projects to support research and innovation or building trusted scientific data infrastructures are not new developments. Thus, we can build on already existing researcher engagement initiatives[1] focussing on future research environments[2] and reports, such as the final report of the High Level Expert Group on Scientific Data "Riding the wave. How Europe can gain from the rising tide of scientific data"[3]. Its contributors formulated recommendations in order to see their "Vision 2030 for scientific data infrastructures" come true. Many of those recommendations strongly relate to the needs and requirements that participants of the workshop "Co-creating the EOSC: Needs and requirements for future research environments" voiced. Among them are the need to work out legal issues to encourage (global) data sharing, create reliable and sustainable data frameworks, work closely with the real users, or make room for a new field of jobs as data experts[4] and data scientists.

[1] For example, the Deutsche Forschungsgemeinschaft (DFG) launched a series of workshops to discuss research in 20 years with young researchers (https://www.dfg.de/dfg_magazin/aus_der_dfg/aktuelles/190912_young_researchers_workshop/index.html). Topics that were discussed include amongst others research without boundaries, research without disciplines, science communication and scientific quality. (https://www.dfg.de/dfg_magazin/aus_der_dfg/aktuelles/191001_next_generation_workshop_digitaler_wandel_berlin/index.html). The workshop's findings will feed into a project called "The Digital Turn in the Sciences and the Humanities". The project aims at establishing policy positions focusing on scientific perspectives (https://www.dfg.de/en/research_funding/principles_dfg_funding/digital_turn/index.html).

[2] A BOF-Session on "Visions for future research environments: Services, Processes and Policies" took place at the 14th RDA Plenary in Helsinki. Objectives included e.g. the identification of needs that researchers and RIs are facing with an evolving scientific landscape, identifying RDA outputs that may help to address these needs and providing a basis for decision makers, industry and research environments: https://www.rd-alliance.org/visions-future-research-environments-services-processes-and-policies

[3] Los, Wouter. (2010). Riding the wave How Europe can gain from the rising tide of scientific data Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission: https://www.fosteropenscience.eu/content/riding-wave-how-europe-can-gain-rising-tide-scientific-data

[4] The University of Cambridge, for instance, launched a "Data Champion Programme". Data champions advise members of research communities on proper handling of research data. The programme is open to all university members interested in the issue: https://www.data.cam.ac.uk/intro-data-champions. Another example is the TU Delft, which provides disciplinary support for research data management and sharing via data stewards. They

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

Against this background, we are well aware that the EOSC is a process – it has to evolve and adapt to future research environments. Even though the building of the EOSC starts from existing infrastructures and services, we as the research community and anticipated main users of the EOSC services see a need to address visions that are more ambitious. Right from the onset, we need to ensure that design decisions taken now do not preclude other approaches that may follow. We need to combine research and development (R&D) at higher Technology Readiness Levels (TRLs) to roll out services with R&D streams at lower TRL levels to feed the EOSC pipeline with more advanced services. Moreover, we need to ensure that the EOSC provides services that support innovative research and a competitive advantage for R&D in Europe 5 to 10 years from now and beyond that timespan. Thus, collecting input from researchers in only a short period is not enough. We need solid development tasks for the EOSC and advanced research activities to prepare the next generation of the EOSC. All of this takes time, especially if you are blind men exploring an elephant for the first time.

2. The Workshop on needs and requirements for future research environments

The EOSCSecretariat[5] partner TU Wien[6] is organizing a series of workshops, interviews and consultations to better understand what research communities need. The goal of these workshops is to elaborate visions on how research will be conducted in 5 to 15 years and what the effect and impact on research infrastructures will be. The findings will feed directly into the work of the EOSC Executive Board (EB) and Working Groups (WGs), thus provide crucial input for the development of the EOSC. As part of this initiative, 20 European researchers were invited to brainstorm not only about the actual needs of their current research but also about future needs and requirements concerning research infrastructures, and services. Eventually 19 researchers participated.

The by-invitation-only workshop "Co-Creating the EOSC. Needs and requirements for future research environments", took place on 13 January 2020 in Feldkirch in Vorarlberg, Austria. It targeted a selected group of top-level researchers coming from different domains with a focus on natural and technical sciences[7] as they are more data driven. High-profile researchers (ERC[8] grant holders, members of the Marie Curie Alumni Association[9] and members of the Young Academy of Europe[10]) were selected, considering geographic, gender and disciplinary spread (c.f. Fig. 1).

help with e.g. setting up secure data storage, give advice on good data management practice, share information about data archiving and are contact points for any data questions: https://www.tudelft.nl/en/library/current-topics/research-data-management/r/support/data-stewardship/

[5] see also https://www.eoscsecretariat.eu/node

[6] see also https://www.tuwien.at/en/

[7] As part of TU Wien's actions concerning researcher engagement, other domains will be approached. For example, TU Wien organizes another workshop in Vienna that focuses on field specific needs in the Health Care domain. This by-invitation-only workshop takes place in the beginning of May 2020.

[8] See also https://erc.europa.eu/

[9] See also https://www.mariecuriealumni.eu/

[10] See also http://yacadeuro.org/

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

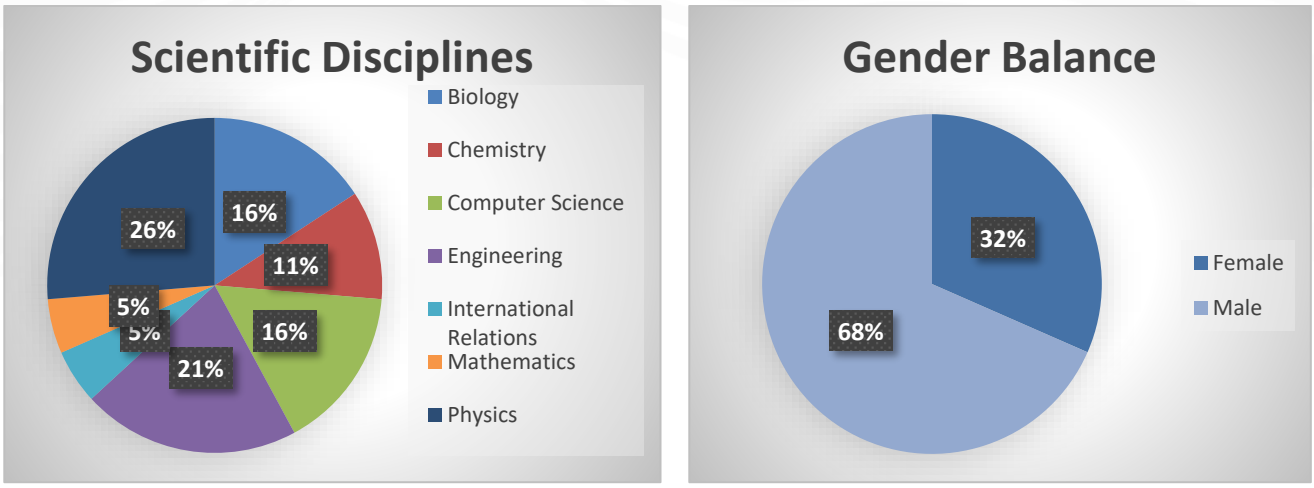

Fig. 1: Participants' profile: (a) disciplines and (b) gender balance

During three breakout sessions, the participants covered topics including the impact of research in society as well as means of measuring it, communication and language barriers, the roles of AI and researchers in research processes, sustainability of the EOSC infrastructure, the building of trusted and sustainable data frameworks and legal frameworks, guidelines and Rules of Participation. The researchers voiced not only visions of what the EOSC could be. In addition, they identified challenges, named prerequisites that are crucial for the success of EOSC and determined services that they need. The following chapters discuss all of these issues in detail. Tables listing the aforementioned services summarize each chapter.

3. A "wish list": Identified Key Services

3.1 Impact of Research in Society

Research has to gain public awareness and the public needs to trust scientific facts. Against this background, empathy, transparency and the mediation of research ethics might have a bigger impact on the public status of research than data quality or quantitative ways of measuring impact. Both concepts – increasing the impact of research by qualitative means and measuring impact – have to be explored.

3.1.1 Qualitative means of increasing the impact of research in society

If the impact of research in society is to be increased, proper storytelling is of importance. Instead of only communicating facts and results, researchers need to explain why research is conducted as well as how and why society benefits from science. However, the challenge is to simplify research in a way that makes it understandable for the public without losing the essence of it. That is why storytelling is crucial: It transports knowledge and it is the basis for teaching and educating. In this regard, **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, organizing events, collaborative document authoring** (a "Google Docs for EOSC") **and virtual conference facilities** (a "GoToMeeting for EOSC") are essential collaborative research activities to be supported by the EOSC. All of these are qualitative means of increasing the impact of research in society. A challenge that comes with embracing storytelling and the promotion of science is the need to work out e.g. legal and ethical issues. What can we communicate and share with the public? Who is the public: is it every

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

single human being with access to EOSC services? Is it all human beings who contribute to EOSC? Or is it every human being living on this planet?

3.1.2 Measuring the impact of research: Metrics and Evaluation

Metrics and evaluations are key to illustrate the scientific performance in order to measure the impact of research and to indicate how to distribute resources (e.g. funding). Currently the evaluation of research is based on publications and their impact factor. Obviously, this is not an effective way of measuring research impact, as it does not consider **innovation as a factor of the researchers' output**. There are many other tasks that researchers have to perform that have nothing to do with innovation per se, but should be considered invaluable contributions to the broader scientific development of a discipline, such as the building and maintenance of scientific instruments, or software code, which is rarely valued with other traditional measures of "impact".

Highly cited papers may include useful data but this does not mean that the research performed is innovative. Thus, **EOSC shall offer services that enable the computation of more advanced metrics**. A possible solution includes the association of persistent identifiers (PIDs), e.g. DOIs[11] to support computation of metrics. Currently, DOIs are predominantly applied to papers. However, they should be used more intensively to identify different parts of analysis, or even different aspects of the data. If datasets are being considered in evaluation processes, then **EOSC needs to offer services to track (meta) data (contribution)**: In order to show the value and usefulness of a dataset, it is essential to know how many times it has been downloaded, how often access has been granted for computing processes, or if and how it has been re-used. In addition, the sizes of specific research communities are relevant for understanding both relative and absolute impact.

A challenge that comes with measuring the impact and relevance of research by the number of papers is the inability of these criteria (numbers, quantity) to **measure the quality of the researchers' output. Moreover,** such metrics are no guarantee for future research outputs. Unfortunately, funding institutions rely on the number of published and cited papers in order to distribute resources. Furthermore, **research proposals and other forms of scientific output (code, data, mentoring…) should gain more weight** in the evaluation of researchers.

Another challenge concerns **collaboration.** Data is often not produced by individual researchers. Thus, evaluation processes have to be designed to not just measure a researcher's performance but that of a network of collaborators. This holds true for curating papers, but is more pronounced in the case of data and its processing pipelines, including machinery, software and services. This is specifically relevant for interdisciplinary research.

[11] To get an example for the use of DOIs for linking datasets with publications, check out: Palanisamy, Giri & Shrestha, Biva & Younkin, Katarina & Jundt, Rolanda & Martin, Mark & Elliott, Jannean. (2016). Data Always Getting Bigger—A Scalable DOI Architecture for Big and Expanding Scientific Data. Data. 1. 11. 10.3390/data1020011

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

Tab. 1: Service Suggestions: 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.

3.2 Communication and Language Barriers

Hurdles of current research are language barriers. Very good English language skills are key to perform research. **EOSC can help to get rid of language barriers** by providing researchers with advanced translation services. Researchers can then write their proposals and papers in their native language and accelerate scientific output production as it enables them to focus on their actual research.

Thus, **translation services need to be provided by the EOSC**. It is also possible to develop these as automatic services. Implementing these services requires also some **rules on the usage of the data** curated by them. These services shall offer machine to human translation as well as voice-to-voice translation (big asset for virtual meetings).

Language barriers are not limited to spoken language: getting familiar with some mathematical concepts for specific research questions may be time-consuming. Services as **math translation** can be very useful, specifically in interdisciplinary research processes. Thus, another category of translation services to be provided by the EOSC include **horizontal translation services** in order to translate domain specific technical terms of research outputs across disciplines or to policy makers, and **vertical translation services**. The latter cover personal communication adaption services and the translation of scientific concepts as well as explanations for different career stages, i.e. within a discipline but across different levels of expertise, from specialists to early career researchers, students, and the interested public at large.

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

Automatic metaphor translation and ontology mapping are future services that ESOC shall offer to support the researchers' work. AI based copy editing as a service could improve the accuracy, readability, and fitness of papers and ensure that it is free of errors, omissions, inconsistency, and repetitions.

Another translation service eventually to be supported by EOSC is the conversion of the description of steps in an analytical process to corresponding code and computational processes, starting from support for graphical programming and workflow (re-) composition services. Advanced technologies shall support the fully automatic or guided development of such processes.

Virtual environment services or **virtual reality solutions** for meetings, supporting sophisticated interaction between researchers shall be developed to overcome the current barriers of video conferencing and remote interaction. **Infrastructures to enable such virtual (team) meetings** shall allow for coffee break like chats, because lots of important input relating to research is being discussed this way. Besides, virtual reality solutions can lower barriers (to speak one's mind in front of others).

Scholarly communication in the digitalization era means, among other things, ensuring that scholarly knowledge communicated in the literature is machine actionable. Scholarly knowledge is buried in millions of PDF documents, inaccessible for machines. The representation of scholarly knowledge in form of articles is designed for human experts but this form severely hinders finding and reusing scholarly knowledge.

Tab. 2: Service Suggestions: 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 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.
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

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

	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.

3.3 Re-defining the roles of researchers and AI in research processes

Artificial Intelligence will have increasing impact in research. Machines will definitely automate many activities in the near and distant future. Examples, suggesting this future may not be distant, include Springer publishing its first machine-generated book[12] and a Chinese court ruling that an AI-written article is protected by copyright. Thus, future research environments need to take into account the role of machines and AI in research processes as well as the role of researchers in such processes.

For one thing, researchers might be "replaced" by machines and AI, as they are out-performing human beings on a number of until recently unimaginable tasks, with further improvements to be expected. Then again, research is largely creative and explorative, and researchers often get inspired about important (e.g. societally relevant) research question in non-research environments, e.g. while reading a novel. In addition, emotional intelligence (EI) is a human trait. In research, empathy and EI can be a valuable asset. Moreover, researchers act as educators and play crucial roles in the (personal) development of their PhD students.

AI driven research promises huge advantages as it speeds up research processes. However, there are boundaries and limits. Machines and AI might not be able to identify relevant research topics. It is difficult to imagine an AI that is capable to distil the SDGs for the 22nd century. Moreover, **rules need to be established** to regulate what we want machines and AI to do. For instance, we need to be clear about the ownership of scientific output (e.g. the creator of an algorithm or who operates the execution environment?). There also need to be restrictions of research done by machines as well as by human beings when it comes to body experiments.

Semantic interoperability between humans and machines will be achieved in the next 5 to 10 years. Thus, co-working between researchers and machines will be possible. Machines can create first drafts, which are than revised and corrected by humans. **Services for checking the forms and formal criteria** of proposals and the **observance of ethical guidelines** shall help in the research process. Another time-consuming work of researchers that machines can support is **human resource**

[12] See also https://www.springer.com/gp/about-springer/media/press-releases/corporate/springer-nature-machine-generated-book/16590126

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

management. AI based services can evaluate the CVs of students, or researchers and suggest candidates, building upon advanced metrics for research impact assessment as outlined in section "3.1 Impact of Research in Society". However, transparency of the resulting evaluations is essential and its lack in current AI systems is the biggest barrier to immediate deployment of such systems.

Tab. 3. Service Suggestions: AI supporting research(ers)

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.

3.4 Building trusted and sustainable data frameworks

3.4.1 Trusted data frameworks

Trust in the quality of data and the quality of services is crucial for the success of EOSC. Data needs to come with context. Thus, metadata is highly important as well as challenges that come with the differences in quantitative and qualitative data, or the role of AI in data gathering.

Access to data is essential in order to trust EOSC services. The freedom of use of data is extremely important for EOSC. However, researchers cannot be forced to open all of their research data: For one, sensitive data and data that cannot be anonymized cannot be open. Ensuring data privacy is an extremely complex process for researchers.

With that in mind, **EOSC will contribute by giving open access and enable automatic recording of metadata**. Thus, research can be cross-examined and datasets can be verified. Measurement devices, algorithms and computation processes shall automatically record provenance metadata and other metadata and make it accessible to the researcher, to the co-authors and to the collaborators (both humans and machines).

As working with data becomes more complex, researchers, their students and personnel have **to gain data expertise**. Beyond trainings, **hiring data (management) experts, data wranglers and data scientists** as part of broadly deployed "data wrangling on demand" services will increase the efficiency of researchers and the quality of research outputs.

An observable trend is the move from data sharing to data visiting both because of data privacy reasons but also due to the massive volumes of data involved in analyses. Data visiting allows keeping the data close and private, so that some privacy issues might not come into play.

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

In this light, **it is crucial that EOSC finds ways of dealing with open access and automated access guided by clear and transparent Rules of Participation**, because there is no way to separate data from the services that operate on the data.

Tab. 4: Service Suggestions: 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 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 the 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, 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.

3.4.2 Sustainable data frameworks

The climate crisis, the extinction of species, global poverty and social inequality are only a few of the challenges that humankind has to face in the 21st century. The Sustainable Development Goals[13] (SDGs) serve as a guideline to bring about a better and sustainable future for all. Research and researchers are crucial in finding ways to make this real. Thus, participants of the workshop identified several (data) challenges in research and expressed ideas of what the EOSC should offer to support them in these endeavors.

First, researchers have to have a say in defining societal grand challenges. 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. In that sense, **EOSC has to be a vehicle to join forces in front of**

[13] See also https://www.un.org/sustainabledevelopment/sustainable-development-goals/

grand challenges. It shall help to transform individual research efforts to collective efforts by changing day-to-day habits of individuals, by altering the ways researchers use services and by eliminating the barriers between different fields. That will also change the structure of universities and re-define the ways to conduct research. Against this background, EOSC can offer, for example, a database to collect pressing research questions.

In addition, **Sustainability** touches two other crucial topics: **Data storage capacities** and **cost/benefit ratios**. **Eliminating redundancies or the duplication of efforts** is crucial to avoid stressing our data storage and computing capacities too much. In order to do so, we need to be able to track data. Furthermore, data has to be as findable and accessible as possible. Researchers need to know what exists already. Besides such efforts, we need to come to terms with the **deletion of data**.

Data Management Plans (DMPs) help with getting to terms with the deletion of data as well as with the handling of data. As such, they are perceived as **useful instruments**. However, researchers shall not necessarily be the ones to write DMPs. Researchers **are experts on their data, and they should be consulted as such, but they shall not be responsible for setting up a DMP**. Rather, they shall be involved in the process of its creation.

Tab. 5: Service Suggestions: 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.
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.

3.5 Legal Frameworks, Guidelines and Rules of Participation

Participants discussed the relevance of legal frameworks, guidelines and Rules of Participation across different topics. They are considered crucial when it comes to regulating AI (see chapter “3.3 Communication and Language Barriers”), or the handling of data (e.g., the deletion of data as mentioned in chapter “3.4 Re-defining the roles of researchers and AI in research processes”, or

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

regulating access to data in chapter "3.5 Building trusted and sustainable data frameworks"). **EOSC has to work out transparent legal frameworks and establish concrete guidelines and Rules of Participation that support their implementation in the EOSC.**

3.5.1 Open Data

Positions on how to use data or how to contribute data need to be explicit. These 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.

EOSC needs to impede market dominance for any stakeholders or any participants in the EOSC framework. 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).

3.5.2 Citizen Science

In order to include citizens properly, there need to be guidelines that inform citizens on what data is valuable and on how to collect data the right way so that it can be used for research. Thus, we need **trusted validation processes**. Researchers will refrain from using datasets that they do not trust, because basing results on poor datasets comes with a loss of reputation that may end careers.

Tab. 6. Service Suggestions: Legal Frameworks, Guidelines 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).
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.

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