



Italian Computing and Data Infrastructure

TURNING OPEN SCIENCE AND OPEN INNOVATION INTO REALITY

ICDI Position paper on EOSC Partnership
Strategic Research and Innovation Agenda

September 2020

Contents

4	Introduction
5	General observations
7	EOSC Guiding Principles
8	Action Areas
8	General remarks of the Action Areas
8	Missing perspectives in the different Areas
8	Boundary conditions have practical implementation aspects
9	Comments of specific Action Areas listed in the document
9	AA4 - Authentication and Authorisation Infrastructure
9	AA6 - Resource Provider Environments
9	AA7 - EOSC Interoperability Framework
10	AA10 - Business Models
10	AA11 - Skills and Training
10	AA12 - Rewards and Recognition
11	AA13 - Communication
11	AA14 - Widening to the Public and Private Sectors
11	Missing Action Areas:
11	AA+1: Automation and FAIR-by-Design
12	AA+2: Data protection
12	AA+3: IPR
12	AA+4: harmonisation of national/regional infrastructures
12	AA+5 Multi-inter and trans-disciplinary research activities
12	AA+6: transparency, reproducibility, quality check and accountability
14	Priorities
14	General remarks on Priorities
14	Comments of specific Action Areas listed in the document
14	AA4 P6: Establish and implement a common framework for managing user identity and access in a highly distributed ecosystem
14	AA5 P7: Ensure a feedback mechanism to engage with users and further develop the EOSC environment to meet their needs

14	AA7 P9: Promote the use of open specifications, where available, to ensure technical interoperability when establishing EOSC services and P10: Agree and implement a common set of rules to ensure data and services within EOSC support interoperability
14	AA11 P17: Develop Open Science training and professionalise associated roles
14	AA12 P18: Create a Europe-wide framework for rewards and recognition that includes Open Science
15	Missing Priorities:
15	AA3+P: basic Open Science skills
15	AA4+P: reusing and adapting data analysis software for new communities
15	AA+6+P: Quality check on FAIR data
16	KPIs
17	Synergies
17	Missing Synergies
18	Additional comments
18	Implementation stages
19	About

Introduction

This document summarises the views expressed by the Italian Computing and Data Initiative (ICDI) in response to the open consultation for the EOSC Strategic Research and Innovation Agenda (SRIA), closed on the 31st of August. It provides insightful input and suggestions about the current draft of the SRIA document shared with the wider EOSC community, with the aim of helping to shape the future vision of the European Open Science Cloud.

 This work is licensed under a Creative Commons Attribution 4.0 International License

Editors: Donatella Castelli (CNR), Stefano Cozzini, Sara Di Giorgio (GARR), Federica Tanlongo (GARR)

Authors: Sanzio Bassini (CINECA), Tommaso Boccali (INFN), Stefano Cacciaguerra (INGV), Donatella Castell (CNR) Massimo Celino (ENEA), Massimo Cocco (INGV), Sara Di Giorgio (GARR), Alessandra Giorgetti (OGS), George Kourousias (Elettra Sincrotrone), Mario Locati (INGV), Donatella Lucchesi (INFN), Silvio Migliori (ENEA), Gelsomina Pappalardo (CNR), Laura Perini (INFN), Caterina Petrillo (Università degli Studi di Perugia), Roberto Pugliese (Elettra Sincrotrone), Giorgio Rossi (Università Statale di Milano), Federico Ruggieri (GARR), Riccardo Smareglia (INAF), Federica Tanlongo (GARR).

This position paper is the result of a collective effort. We thank the ICDI representatives who provided useful inputs and feedback.

General observations

In this section we list some general observations regarding the Strategic Research and Innovation Agenda, its broader principles and scope as seen by ICDI members. Hence, these observations are not necessarily addressing a specific point or section of the document, but the underlying vision and strategy.

EOSC scope should be wider than just implementing the “web of FAIR data and Related services”, and support a new approach to the processes of data-driven science.

EOSC has been introduced as a platform to support Open Science and Open Innovation 2.0: “Open science refers to a new approach to the scientific process based on cooperative work and new ways of disseminating knowledge, improving accessibility to and reusability of research outputs by using digital technologies and new collaborative tools” (COMMISSION RECOMMENDATION (EU) 2018-790 of 25 April 2018 on access to and preservation of scientific information).

“Open innovation 2.0 is a paradigm based on principles of integrated collaboration, co-created shared value, cultivated innovation ecosystems, unleashed exponential technologies, and extraordinarily rapid adoption.” (...)

Turning Open Science and Open Innovation into a reality requires going beyond the mere provision of facilities to find and access existing reusable data.

Finding data is just the first step towards a data-driven approach to science, and it is not enough to differentiate EOSC and its value pitch in comparison to other, more resource-rich, players.

- **EOSC aims to support and facilitate Open Science and Open Innovation 2.0. Its scope is thus wider** than just implementing a “web of FAIR data and related services”. It should:
 - Cover not only the FAIRification of existing data, but most importantly the **cooperative execution of data-driven research** activities and the **production of new “FAIR-by-design” data** products;
 - Cover the **whole data lifecycle**, from production, to analysis, reuse, and extraction of new knowledge through the connection with infrastructures and facilities;
 - Offer **services and a common policy framework** to support the management of the organisational and IPR problems connected to access, use and reuse of (open) data (including publications);
 - Support the **different scientific scenarios** where data products are generated (e.g. creation of datasets ex novo, extraction of new knowledge from third-party datasets, integration of heterogeneous datasets);
 - **Bridge research and educational activities**, by timely making available to young researchers and other data science professionals training and hands-on environments giving access to newly developed scientific outcomes.
- In the expression “web of FAIR data and related services”, **“services” is the operative word**: to empower researchers in their daily work, **services should be emphasised**. This also includes customisation with respect to the single researcher’s and research group’s experience, as well as its enrichment that can be obtained via **AI and machine learning** techniques that are already available, or will be in the next few years.
- **Automation is a key point** for the viability of “FAIR-by-design” data, and hence to achieve EOSC objecti-



ves. Besides the overhead required out of researchers, a manual management of dataset access does not allow for timeliness and scalability, nor for much-needed flexibility and dynamicity of access policies (e.g. strict / dynamic embargo periods, access limitations etc.).

- To guarantee openness of research data at the highest possible level, according to the motto 'as open as possible, as closed as necessary', **policies and practices for the management of IPR of data, data products and literature** should be clarified and translated into practical provisions **in a cross-national and multidisciplinary framework**.
- **Multi-, Inter-, Trans-disciplinarity aspects should have more weight**, as they will be key to tackle the big challenges of this century, from climate change to food and water security, to the fight against cancer and other diseases, and healthy ageing. Besides ensuring interoperability among data from different domains (through appropriate ontologies, etc.) this should also include the opportunity of having one's frontier research adequately recognised, and therefore should cover not only technological aspects but also **research evaluation, scientific communication, careers and IPRs ones**.
- It is not entirely clear whether the work in the three implementation stages foreseen in the strategy will be completely sequential or there is a possibility for some activity foreseen in stages 2 and 3 to start at the same time as those in stage 1. A completely **sequential approach in the implementation stages appears inadvisable**. Some of the activities that are needed to achieve stage 2 and 3 outputs should start ASAP as the technology is available and they would offer a clear added value. It is of the utmost importance in order to motivate researchers that EOSC move beyond something that is already old/commodity. It must offer both **leading edge services** and **extend the exploitability of existing services to multiple domains** by overcoming the fragmentation of current (e-)infrastructures.
- **Harnessing and coordinating national/regional levels is one of the drivers for a successful EOSC**. While the strategy correctly foresees to monitor the situation at the level of MS, and to stimulate the funding governing of national activities and infrastructures, there's no mention of an active work to coordinate, interoperate and federate these national initiatives, nor can it be expected to be achieved at the end of 2020 or happen automatically. As for other points above, alongside the technological aspects (interoperability framework) it will be of paramount importance to cover **organisational and policy aspects** to really enable researchers to work with their peers in Europe as if they were in their context (e.g. access policies for scientific facilities and tools, workflows, but aspects like multilingualism should be also tackled).
- **The distinction between implementation challenges and boundary conditions is not entirely convincing**, as many points that are listed in the latter are actually things that need to be implemented (e.g. the Competence Centres) or imply some further implementations (e.g. the definition of RoPs implies the implementation of mechanisms to verify their enforcement).

EOSC Guiding Principles

EOSC should cover the full lifecycle from data production/ discovery to curation, long-term preservation, archiving, and reuse. It should provide **services and virtual environments where researchers can collaboratively and dynamically perform experiments by using each other's resources** (including infrastructures and facilities) and by sharing intermediate results along the full research process.

It should also **facilitate the openness and transparency** of the research methodologies, as the **verifiability and reproducibility** of the obtained results.

This is not just about building another technological platform, it is a paradigm shift, and it is technically feasible today. Many technologies can be harnessed to make it happen and the current trends indicate that more will be available in the near future.

In the EOSC vision there is an opportunity to bring forward a true cultural change. It is about realising an equal-opportunity access to substantial analytical, computational, archiving and network resources for scientific data. In order to make it an impactful initiative, all large new datasets must be FAIR by design, also using AI tools at the data acquisition step, and the data quality must be guaranteed/endorsed by the research operators (RIs, RPOs, e-Is).

When the FAIR data-sets will be abundant, then AI instruments can effectively support the researcher's work and enable an all-new dimension of interdisciplinarity in research. All of the above requires increasing resources along the whole scientific data-lifecycle, including the **professionalisation of data stewardship, the integration, upgrade and expansion of the existing data repositories and data services, the provision of resources to analyse and further develop the data-sets, the clear link of all new analysed data-sets** with the original FAIR data sets.

Action Areas

General remarks of the Action Areas

For the majority of Action Areas described in the current version of the SRIA there is no **definition of what the area is intended to cover** ("what"). The meaning is simply conveyed by the name of the Area, e.g. "Interoperability framework". The description reported is mainly about what WGs have done so far to address them. In the lack of such a definition, it is difficult to understand the motivation for the selected priorities and for the actions required to implement them. Each of us can comment and propose something different.

Recommendation:

Better define what the AA is expected to cover

Missing perspectives in the different Areas

The **granularity** of the Areas is very different. User Environments, Resource Provider Environments, and Interoperability Frameworks imply much more than what is described in the document. For example, they have all to deal with the setting and /or the satisfaction of policies that may be very different. EOSC should definitely provide support for their enforcement and monitoring. Moreover, these environments should facilitate the user in a customized way, i.e. according to the needs and research practices of users. These, and other requirements, demand for well structured actions.

The current list of implementation priorities is very partial.

Recommendation:

Define all the AA with an eye to the overall context and objectives and not a separate silos. This applies in particular to the User Environments, Resource Provider Environments and Interoperability framework

Boundary conditions have practical implementation aspects

It should be noted that all **the challenges listed under "boundary conditions" also require the provision of appropriate services**. For example, the change of the rewarding mechanism implies service supporting the new approaches and metrics that will be defined. There is no mention of this class of services anywhere in this document.

The whole strategic approach presented in this part of the document assumes that abundant FAIR datasets will be available to operate open access services on them. The premise is not necessarily true, given the data volume increase and the very limited amount of FAIR-conversion if this were left to the good-will of individual researchers, whose work resources are limited and concentrate on research. Too much emphasis on the downstream processes (which are the original EOSC contribution) without consideration that either the whole data lifecycle is supported, from acquisition to analysis and re-use, or the purely "downstream" activities will be limited by insufficient supply of FAIR data sets.

Recommendation:

Consider including appropriate actions to implement the services needed to answer the challenges underpinned among the “boundary conditions”

Comments of specific Action Areas listed in the document

AA4 - Authentication and Authorisation Infrastructure

AAI is one of the key components for a working EOSC. However, the goals of the “Authentication and Authorisation Infrastructure” Action area are not clearly explained in the open consultation document.

Recommendation:

Better explain the goals relating to the AAI Action Area

AA6 - Resource Provider Environments

No service nor policy action is foreseen to support service providers, developers, small and medium-size enterprises, albeit they are regarded as key actors for the success of EOSC. For example, nothing is specified regarding the need of defining and harmonizing service catalogue onboarding policies and procedures, especially where different catalogues (national, thematic catalogues) are in place.

Recommendation:

Consider adding specific services/policy actions to support service providers, developers, small and medium-size enterprises

AA7 - EOSC Interoperability Framework

The interoperability framework is the key novelty of EOSC with respect to FAIR data and the potentially disruptive transformation of research practices. It does not rely on openness of data collections only. On the contrary, it largely implies the existence of common services capable of making FAIR data-based research a concrete opportunity for all researchers, regardless of their storage, software or computing resources

Interoperability is essential to seamlessly integrate the existing infrastructures, but cannot be limited to data (finding, moving, accessing). A lower-level common interface to access resources (computing, storage, algorithms, software) needed to process the data, is also needed to avoid situations where Petabytes of data are “found”, but there is no practical way to process them. A situation where a researcher of a resource-rich institution can do something others can't do would actually be very unFAIR and question, to all practical ends, the openness of the system: to truly advance data science, EOSC should not only provide technology and common standards, but also ensure open participation, inclusiveness and transparency.

Interoperability should not be regarded as a quality of datasets, but as a programme, whose implementation requires strong support by skilled people in distributed teams which may be supported/coordinated by the Association.

Recommendation:

The implementation of common services that can be used by researchers cross-discipline to perform data driven research, regardless their access to storage, computing and software resources should be emphasised in this Action Area

AA10 - Business Models

In the current document, the section dedicated to the Business Models Action Area mainly reports about the Work done by the Sustainability WG till now, but it does not identify the challenges to be addressed in the period 2021-2027. In particular, the challenges related to the definition of the contribution to the Sustainability of EOSC and its elements (e-Infra, RIs, and resource providers in general) are missing.

Recommendation:

The challenges to be addressed in the period 2021-2027 need to be identified in detail, with a special focus on the contribution to the Sustainability of EOSC and its elements (e-Infra, RIs, resource providers)

AA11 - Skills and Training

Skills and training should cover the whole research data cycle, from FAIR-by-design implementation to access, reuse, publication, curation and long term preservation.

Common/comparable curricula across Europe should be a priority as they would facilitate career exchanges and mobility, and have an impact on AA12.

Recommendation:

Consider widening the scope of this Action Area in order to cover all aspects of the FAIR data cycle. Actions to harmonise Open Science curricula across Europe should be also included in this AA

AA12 - Rewards and Recognition

This action area is particularly relevant to get started. The implementation of FAIR principles and openness requires a significant overhead, and this effort can't be put down to the good will of single researchers and research organisations. The long term goal should be to **make contributing to EOSC effortless for the data producers** and to share the benefits among all researchers. The **evaluation and recognition of cross-disciplinary research** should also be taken into account: currently, research evaluation and careers are strongly domain-specific, which provides little incentive and rewards for joint and collaborative activities.

Recommendation:

This Action Area is pivotal to the uptake of FAIR data principles and Open Science and should therefore have a high priority. The evaluation/reward of cross-disciplinary research should be specifically addressed

AA13 - Communication

This Action Area may require to establish effective **Competence Centers** or services to enable the prompt exploitation of the EOSC features. A widespread network of national, regional or thematic competence centres would effectively contribute to create awareness on Open Science and FAIR data principles, provide guidance to those approaching them for the first time and act as a EOSC one-stop-shop.

Recommendation:

A network of Competence Centres should be put in place early in the implementation strategy to facilitate the prompt exploitation of the EOSC

AA14 - Widening to the Public and Private Sectors

This Action Area suddenly became extremely urgent in the context of the COVID-19 outbreak, but its importance goes well beyond the challenge at hand: the pandemic has dramatically demonstrated the value of including in EOSC the data coming from areas (e.g. healthcare, civil protection, utility management) that are not per se research but can help research to address key societal challenges.

Hence, its objectives must be pursued with clear actions and not be left to spontaneous fallout effects.

Widening to the Public and Private Sectors also require the clarification of **usage policies**, especially when services are publicly funded. The list of stakeholders addressed by this Action Area should be further detailed, in particular, elements such as Education and industry (which were identified elsewhere, e.g. Tinman document) should be explicitly mentioned.

Recommendation:

This Action Area should be prioritised in view of the challenges posed by the COVID-19 situation. A clear action plan should be set out to address it, which should include the clarification of usage policies, especially when services are publicly funded. The list of stakeholders addressed by this AA should be further detailed

Missing Action Areas:

We have identified some key aspects that either are not taken into account in the current presentation of the Action Areas, or are tackled together with other topics but would be better addressed, in our members' view, in a dedicated Action Area.

AA+1: Automation and FAIR-by-Design

EOSC should elaborate and provide examples of technological options to make "FAIR-by-design" data in all domains in order to stimulate the instrumentation industry to **produce "EOSC-ready" data outputs from instruments**. First basic instructions to collect automatic metadata should be defined. Co-creation between laboratories and industry may occur in this domain to accelerate the production of FAIR data sets.

AA+2: Data protection

The EOSC and FAIR data-based protocols should be made suitable for health, population, climate, mobility data for obvious reasons. The GDPR must be fully analyzed and **application guidelines** must be produced as it is not a real obstacle to circulation of useful scientific or clinical information, but it is often perceived as an obstacle by the data owners.

AA+3: IPR

The **balance between openness and recognition** (and in some cases remuneration) of intellectual property is one of the keys to making Open Science viable: Openness can't be at the expense of the careers of researchers nor of the legitimate profit of individuals/organizations who invest in producing and making available quality data. **Appropriate licensing and IPR protection mechanisms** must be identified and solutions to manage the "relative openness" of access policies at a given time and **allow for their flexibility and dynamicity** should be part of a successful strategy for the uptake of Open Science.

AA+4: harmonisation of national/regional infrastructures

Harnessing and coordinating national/regional levels is one of the drivers for a successful EOSC. While the strategy correctly pinpoints the need for monitoring the situation at the level of Member State, and to stimulate the funding governing of national activities and infrastructures, active work should be also foreseen to coordinate, interoperate and federate these national initiatives, which is only just started. This should include technological aspects (interoperability framework) as well as **organisational, context and policy aspects** (e.g. access policies for scientific facilities and tools, workflows, but also aspects like multilingualism).

AA+5 Multi-inter and trans-disciplinary research activities

The big scientific challenges of this century (eg. the UN Sustainable Goals, the Horizon Europe Mission programme, as well as those raised by the recent COVID-19 Pandemic) have in common the need to focus on integrated complex problem solving. This demands cooperation between specialists with diverse backgrounds in both the natural and social sciences. In the next 7 years, EOSC is expected to **support the necessary multi-inter and trans-disciplinary research activities by removing fundamental barriers** that are hampering an effective and efficient addressing of these challenges. This not only includes basic lack of interdisciplinary infrastructures and collaboration supporting environments, but also **organisational, training and career aspects** (e.g. the lack of researchers trained on integrated research, of quality journals for publishing these types of research activities, of rewards and career evaluation criteria for those working across-domains, etc).

AA+6: transparency, reproducibility, quality check and accountability

One of the big challenges of Open Science is to make data and data products more transparent, reliable and trusted. This not only affects science and scientists and their careers, but is an important societal challenge, which impacts on the Public's understanding and trust in science and scientific results, as well as in the decision



making by policy makers on sensitive matters (from vaccines and public health decisions, to climate change and the control of extreme climate events, etc). However, the fact that data are FAIR does not per se imply they are higher quality than others, but **strategies can be put in place to favour the process of verification and reproducibility, while incentivizing the debate and peer review of data and data products**. At the same time, a massive alphabetization re OS themes should be carried out at the level of society (education, decision makers, media, citizens...).

Priorities

General remarks on Priorities

Most of the questions (like PID, AAI, APIs) need a central concertation, and cannot be left to single institutions. For most categories, **EU, National and Institutional levels should all be competent and part of this concertation.** To be as effective and impactful as possible, the actions should be carried out at all the three levels and the **European, national and institutional strategic agendas should be reasonably harmonised.** However, the EOSC partnership is European in scope and while the necessary definitions should be achieved via consensus between the parties, **central guidance is needed** to make sure they will work.

In terms of timing, we believe that **most of the priorities should be addressed as soon as possible** because there is a high level of interdependencies among them, however this does not exclude that problems can't be tackled in phases and initial solutions can be progressively perfected.

Comments of specific Action Areas listed in the document

AA4 P6: Establish and implement a common framework for managing user identity and access in a highly distributed ecosystem

Exploit the existing identifiers (ORCID) before establishing the need of creating new identifiers.

AA5 P7: Ensure a feedback mechanism to engage with users and further develop the EOSC environment to meet their needs

Feedback and feedforward as it is the users communities that, while exploiting data and services to perform research, will identify new combinations and gaps in the EOSC offer.

AA7 P9: Promote the use of open specifications, where available, to ensure technical interoperability when establishing EOSC services and P10: Agree and implement a common set of rules to ensure data and services within EOSC support interoperability

Interoperability will prompt emerging of more dimensions than foreseen now. A learning process must be set-up according to EOSC usage that leads to robust results.

AA11 P17: Develop Open Science training and professionalise associated roles

These professional figures will have a key role in the development of Open Science. Hence, these professional figures must be created as soon as possible.

AA12 P18: Create a Europe-wide framework for rewards and recognition that includes Open Science

As said before, this is a crucial point. However, in order to reward the openness of science, criteria and standards must be clearly assessed.

COVID-19 has shown that FAIR data are needed also in domains that are not per se of research nature, as



clinical or social data, but GDPR is considered a barrier to sharing them. However individual privacy protection can be guaranteed whilst making urgently available important data to develop research across the scientific dimension (structural, omics, clinical trials) to the population evidence data (e.g. viral charge distribution in time and geography and phenotype and social environment -school, work, entertainment events...). The COVID-19 example shows that research needs also these data to identify patterns and develop models. **A careful analysis of real and apparent barriers to data sharing that result from non-research institutions must be carried out** and proper framework of communication and incentives/obligations must be developed.

Missing Priorities:

AA3+P: basic Open Science skills

Parallel to the training of new data scientists, technologists and stewards, a massive alphabetization of users to EOSC/Open Science opportunities from the education system, professional update training and citizen awareness must be put in place.

AA4+P: reusing and adapting data analysis software for new communities

Engage the science community to expand and adapt successful data analysis software (proper statistics, simulations, models) to the broadest community.

AA+6+P: Quality check on FAIR data

It is essential that the EOSC is, and is perceived, as a most trustable space of access to highly reliable, reproducible when it applies, and state-of-the-art quality of the findable datasets. Much emphasis is put on the quality of services, but the whole success of EOSC enables research relies first on the fact that the **FAIR data sets are all scientifically trustable**, not affected by detectable artifacts, not polluted by malicious -albeit FAIR- datasets. This point is of paramount importance and cannot be left unnoticed, or given for granted as the whole system may collapse if a mixture of good data and untrustable or malicious datasets are mixed together. Clear rules and procedures are needed to ensure that data are managed and quality is ensured by data providers sharing their datasets through EOSC. All of this dimension is overlooked, and this is a dangerous point.

KPIs

The indicated KPIs reflect the partial view of EOSC as expressed in the large part of the document. e.g. “EOSC is populated with a valuable corpus of interoperable data” – nothing similar is claimed for services or, more generally, for resources.

Most importantly, **many of the proposed KPIs are not well quantified/quantifiable**. So while they may pinpoint relevant aspects to be measured, in the proposed formulation they fail at measuring them.

Quantifiable elements should be identified for each of the KPIs, and percentages of success should be indicated. Eg: while KPI3 is highly desirable, it is nevertheless meaningless as a KPI.

Scientific evidence of EOSC's impact on research should be the ultimate KPI: how many studies could not be done without EOSC services? How many users conducted original research with EOSC? How many researchers/innovators were trained to exploit the EOSC?

Publicly funded research must include, as eligible and necessary, the cost of implementation of **FAIR-by-design methodologies and technologies** in order to guarantee that FAIR data are effectively produced, not limited by researcher manpower to be distracted from the true research activity.

RIs and RPOs must adopt and demonstrate FAIR-by-design sustainable practices.

RIs and RPOs must implement quality-check/quality-assurance of the released FAIR datasets.

It is worth mentioning that at this very time a consultation, ideated by the INFRAEOSC5 projects from an idea discussed with the landscaping WG, is ongoing on the indicators to be used to monitor and assess the EOSC development at the level of MS. The idea is to move from a static landscape to **a set of living indicators**, and select a set of relevant aspects to be monitored to accomplish this objective.

The working proposal that is the object of the consultation builds on the results of the first version of the Landscape WG report and of the INFRAEOSC5 projects surveys on National OS Initiatives. All the indicators in the proposal are intended to be quantitative. Further info on the consultation can be found at:

<https://www.eoscsecretariat.eu/questions-validation-readiness-indicators>

Synergies

EOSC support to Horizon Europe partnerships and Missions: of course, essential partners are existing RDIs, and the high performance and high throughput infrastructure (e.g. EuroHPC and EGI). **They are left out of the current version of the SRIA document**, which is a considerable shortage and makes potentially the EOSC an easy to predict failure.

The EOSC Partnership should formulate its strategic agenda with periodic checkpoints and updates, and this should become the reference for novel INFRA-EOSC type projects, which should be as synergistic as possible among themselves. **Support to "bottom-up" conceptual designs in the domain of data science should nevertheless remain as a key part of EC support programmes** in order not to miss potentially disruptive developments that will most certainly feed-forward to the EOSC process and further implementations.

At the same time, attention should be paid not to duplicate efforts and running parallel projects that do the same things.

Some considerations on synergies with other funding programmes:

- EOSC is expected to **reduce difference in opportunities**, and this matches the structural funds scope.
- **Funding activities** which help nations to consolidate their computing infrastructure would have a significant impact, since it can be "one shot" funding whose effects are lasting.
- Funding centrally, from the EU, the overhead that institutions will see when needing to move to open data (the curation / **stewardship costs** for example) would be a good incentive for those institutions which do not go that way for budget reasons.
- **ESIF could support EU-coordinated broadband implementation initiatives** to truly overcome the digital divide. Part of COVID-19 recovery fund should be dedicated to this, in manner as concerted as possible. A clear foresight of EOSC related data traffic, archive and computing resources for research should be developed and investments quantified and designed to meet ESIF criteria.
- The availability of **computing resources** is a key to data-driven science. At the same time, allowing institutions to sell excess computing cycles at more than the cost in a marketplace (in the EOSC-exchange, or elsewhere as long as it is central to the EOSC environment) would act as an incentive to deploying large computing resources. If the cost is not compatible with what Amazon costs, funding from the EU could cover the excess part.

Missing Synergies

Among the synergies with other initiatives, the complete absence of mentions of GAIA-X stands out, which is peculiar as it is partly funded by the EC. While for some aspects GAIA-X can be regarded as a competing initiative, complementarity and convergence would be far more desirable in this case.

Additional comments

Implementation stages

The stages identified do not include support to open scientific workflows nor to user environments.

It is not clear whether they correspond to priorities for funding in the indicated periods. If so, they should be carefully reviewed. Many of the activities for the first period have already been addressed and funded in the past, so while it is conceivable the results can be improved or brought further, they shouldn't be the only ones planned for this stage. Many of the activities indicated in the further stages should start ASAP and, as they are of different nature and will be addressed by different actors, they could well proceed in parallel.

Seven years is a very long time for Information Technologies: it is important that the EOSC vision be open to new developments (e.g. the dramatic evolution of AI, that could have a far wider impact on OS than is now envisaged in the strategy). The SRIA should be a living document and its periodic updates should incorporate new findings and ideas, as well as relevant feedback from users and stakeholders. EOSC should demonstrate to have a robust strategy, including quick learning by doing, to provide timely results and to cope with competition from the private sector.

About

ICDI is a bottom-up initiative aiming at promoting synergies at national level and optimize Italian participation to European and global challenges (EOSC, EDI, HPC,...) that brings together under the coordination of the GARR, major Italian national research and digital infrastructures and Italian public research institutions, National chapters of ERICs and ESFRI Projects. ICDI is one of the four founding bodies of the EOSC Association: the founding act was signed on the 29 July 2020 in Brussels by representatives from GARR, the CSIC (with a similar national representation role of scientific data activities for Spain) and by two European organisations: CESAER and the GÉANT network.

ICDI can therefore be considered as a contact point for the coordination of Italian participation in EOSC development initiatives and for this reason it has collected the comments and recommendations related to the SRIA document. This position paper may be updated as the process to build EOSC proceeds and further opportunities for consultation arise.

If you would like to learn more on ICDI, or get in contact with us, please visit our initiative's website.

www.icdi.it

