10.5281/zenodo.4533784

Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration

[30-09-2020]

Advancing Open Scholarship

D6.1– Report on the general interoperability requirements

Version 1.0 – Final
PUBLIC

H2020-INFRAEOSC-2019
Grant Agreement 863420

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 863420.

Report on the general interoperability requirements

Project Acronym:	**TRIPLE**
Project Name:	**Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration**
Grant Agreement No:	**863420**
Start Date:	**1/10/2019**
End Date:	**31/03/2023**
Contributing WP	**WP6**
WP Leader:	**CNR-ILC**
Deliverable identifier	
Contractual Delivery Date: 02/2020	**Actual Delivery Date: 09/2020**
Nature: Report	**Version: 1.0 Final**
Dissemination level	**PU**

Revision History

Version	Created/Modifier	Comments
0.0	Francesca Di Donato, Tiziana Lombardo (Net7)	
0.1	Francesca Di Donato, Tiziana Lombardo (Net7)	Initial structure
0.2	Francesca Di Donato, Monica Monachini (CNR), Tiziana Lombardo (Net7), Suzanne	Initial inputs from partners

	Dumouchel (CNRS -Huma-Num) Arnaud Gingold (CNRS - OE), Yin Chen (EGI), Carsten Thiel, Joshua Tetteh Ocansey (CESSDA), Maria Eskevich, Elisa Gorgaini (CLARIN), Yoann Moranville, Erzsébet Tóth-Czifra (DARIAH), Gert Breitfuss (KC)	
0.3	Francesca Di Donato (CNR)	ToC revision and new content
0.4	Peter Kraker (OKM)	First internal review
0.5	James Moir (AU)	Second internal review
1.0	Francesca Di Donato (CNR-ILC), Suzanne Dumouchel (CNRS -Huma-Num), Tiziana Lombardo (Net7), Arnaud Gingold (CNRS - OE), Yin Chen (EGI), Carsten Thiel, Joshua Tetteh Ocansey (CESSDA), Maria Eskevich, Elisa Gorgaini (CLARIN), Erzsébet Tóth-Czifra(DARIAH)	Final version

Table of Contents

Table of Figures

Acronyms

EOSC	European Open Science Cloud
RI	Research Infrastructure
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable Accessible Interoperable Reusable
ERIC	European Research Infrastructure Consortium
PID	Persistent Identifier
AAI	Authentication and Authorisation Infrastructure
API	Application Programming Interface
MVP	Minimum Viable Prototype
OPERAS	Open Scholarly Communication in the European Research Area for Social Science and Humanities
SSHOC	Social Sciences & Humanities Open Cloud
SSH	Social Sciences and Humanities
WG	Working Group
TRL	Technology Readiness Level
VO	Virtual Organisation
EOSC IF	EOSC Interoperability framework
ESFRI	European Strategy Forum on Research Infrastructures
MVE	Minimum Viable EOSC
SLA	Service Level Agreement
OAI-PMH	Open Archives Initiative - Protocol for Metadata Harvesting
TEI	Text Encoding Initiative
RoP	Rules of Participation

Publishable Summary

TRIPLE - Transforming Research Through Innovative Practices for Linked Interdisciplinary Exploration is a EU funded project under the INFRAEOSC-02-2019 - Prototyping new innovative services topic, which started in October 2019 and will end in March 2023. Its main objective is to design and develop a discovery platform for SSH, called GOTRIPLE.

This deliverable is the main outcome of Task 6.1 which started at M4 at ends at M36, whose aim is to deal with the definition and the set-up of general TRIPLE's interoperability requirements, considering all the components which are composing the TRIPLE ecosystem (data, resources and tools). As preliminary results, we thus present here a general overview of the main EOSC interoperability requirements and specifications, both coming from a mapping of the EOSC Working Groups outputs, and of the most relevant results of EOSC related projects related to TRIPLE. We also attempt to provide TRIPLE's answers, proposals and solutions to the above mentioned requirements. The final picture presents different levels of precision, which depends on the fluidity of the EOSC definition on the one hand, and on the consequent fact that some implications are still unclear, and a discussion on the measures to address EOSC requirements is still on-going.

While tackling interoperability, we introduce TRIPLE in its context in order to locate the GOTRIPLE platform in the EOSC frame, and more specifically in the SSH cluster of the EOSC (section 1).

Section 2 defines the general interoperability requirements, starting with the software (2.2) and then presents an analysis of the main outputs released by the EOSC Working groups (2.3), taking into consideration as a general reference, the EOSC Interoperability Framework, and, more specifically, the FAIR and Architecture WGs documents (2.3.2, 2.3.4). These are the main guiding references for the design and realization of the EOSC, considering specifically interoperability. Section 2.3.3 illustrates how TRIPLE is translating into practice the FAIR requirements, while section 2.3.5 is focused on TRIPLE current decisions regarding the integration of the TRIPLE solution into the EOSC. To present an enriched scenario, the section includes as well a brief overview of other relevant outputs released by the EOSC WGs (Landscape, RoP, Sustainability and Skills and Training) (2.3.6).

With the aim to provide a comprehensive analysis of the EOSC interoperability requirements, the WP6 partners have analyzed relevant deliverables produced by the main EOSC related projects as preparatory activity. The analysis was useful to understand the EOSC environment and its evolution in terms of interoperability and at the same time to understand which external deliverables have to be taken into consideration for the overall project development in TRIPLE. Section 3 includes a synthesis of this work, which is fully presented in Annex I. Section 4 - Conclusions and Outlook, outlines TRIPLE's the next steps to achieve interoperability and the strategies that will be adopted.

1 | Introduction

TRIPLE is a funded project under the INFRAEOSC-02-2019 - Prototyping new innovative services call, and aims at creating the European open science discovery platform for SSH (named GOTRIPLE). Envisaged to be a service of the European Open Science Cloud (EOSC), GOTRIPLE will be a major component of the SSH marketplace started to be developed under the SSHOC project[1].

In addition, TRIPLE is not a standalone project, but it is embedded into the development of the OPERAS - Open Scholarly Communication in the European Research Area for Social Science and Humanities RI and will be one of its core services[2]. The development of a protocol for the inclusion of OPERAS RI in the EOSC is currently under definition in the OPERAS-P project - funded by the European Commission (2019-2021). OPERAS-P outputs are thus important references in defining TRIPLE's interoperability requirements.

While enabling researchers to discover and reuse SSH data across disciplinary and language boundaries, TRIPLE aims at facing one of the main problems of SSH tackling interoperability issues: its fragmentation. In SSH, use and reuse of resources is suboptimal, and interdisciplinary collaboration possibilities are often missed - as underlined in the ESFRI Roadmap 2018 [1].

This deliverable identifies and analyses the main EOSC interoperability requirements, in the context of the EOSC General Interoperability Framework v.1.0 released in May 2020 and currently out for comments [2], and then analyses the main outputs of the EOSC WGs concerning interoperability; at the same time it aims at describing how TRIPLE is translating them in its solution, taking into account TRIPLE's architectural relations with and dependencies to OPERAS. The main point here is describing how TRIPLE will be integrated into the EOSC (section 2). The provisionality of our decisions depends on the evolving scenario of the EOSC definition.

In the identification of the EOSC interoperability requirements we had to face some difficulties, due to the fact that the EOSC definition process is still on-going, and the context is continuously evolving. The release of EOSC-related outputs, which mainly follows the established roadmaps, is the result of a participatory process to which the TRIPLE's consortium - and especially WP6

[1] Specifically targeted to SSH, the INFRAEOSC-04 project Social Sciences & Humanities Open Cloud (SSHOC, https://sshopencloud.eu/) is the cluster project which aims at building the SSH component of the EOSC. SSHOC is realising the transition from the current landscape with disciplinary silos and separated e-infrastructure facilities into a cloud-based infrastructure of FAIR data, where tools, facilities and training are available for SSH scholars to enable them to adopt a data-driven scientific approach.

[2] A list of the OPERAS services include: the OPERAS certification service; the OPERAS Metrics service; the OPERAS Publishing Service Portal; the OPERAS check-in service; and the OPERAS XML toolbox service. See https://www.operas.unito.it/projects/triple/ and https://www.operas.unito.it/services/discovery-service-triple/.

team - are asked to provide comments and feedback very frequently. In order to be able to monitor the production of the main documentation relevant to our task, and more in general to the TRIPLE implementation, a reporting methodology has been established, by assigning a responsible partner for each working group. This process allowed the team to have a constant update on relevant outcomes and progress of different EOSC Working Groups. Moreover, the TRIPLE Scientific coordinator is member the EOSC Architecture and the Skills and training WGs, which ensures TRIPLE active participation to the EOSC definition process, and at the same time that TRIPLE's experience, issues and needs are shared and and taken into account in the deputated place.

A second difficulty we had to deal with comes from the richness and variety of the EOSC ecosystem, which is composed of several projects, RIs, and initiatives and actions, where a hierarchy of authoritative sources is not always clear. We are addressing this problem thanks to the involvement in our consortium of partners such as EGI, CLARIN, CESSDA and DARIAH, whose different expertise and perspectives, as well as their participation in relevant related projects, helps in understanding the complexity of the scenario which compose the EOSC ecosystem. As a complementary activity, WP6 team performed a mapping exercise, in order to have a complete vision of relevant deliverables produced by the main EOSC related projects. A synthesis of the most relevant outputs is described in section 3 (while the full analysis can be consulted in Annex I) .

2 | General Interoperability Requirements

2.1 Introduction

EOSC governance organizes the definition of the EOSC through thematic Working Groups (WGs), which have been constituted around the following 6 priorities:

- ☐ *Landscape*: The objective is to map the existing research infrastructures which are candidates to be part of the EOSC federation;
- ☐ *FAIR*: Implementing the FAIR data principles by defining the corresponding requirements for the development of EOSC services, in order to foster cross-disciplinary interoperability;
- ☐ *Architecture*: Defining the technical framework required to enable and sustain an evolving EOSC federation of systems;
- ☐ *Rules of Participation*: Designing the Rules of Participation that shall define the rights and obligations governing EOSC transactions between EOSC users, providers and operators;
- ☐ *Sustainability*: Providing a set of recommendations concerning the implementation of an operational, scalable and sustainable EOSC federation after 2020;
- ☐ *Skills and training*: to work on building competence (skills) and capabilities (training) for EOSC.

The EOSC WGs outputs have the status of recommendations for standards, providing the guiding references to the EOSC definition and implementation.

In the perspective of delivering a discovery platform for SSH which will be part of the EOSC marketplace, TRIPLE project has to deal with the work done in the different EOSC WGs, in terms of interoperability, especially to fulfill with the requirements to become an EOSC service, but also in terms of sustainability. For what concerns the purpose of this deliverable, the most relevant output are those coming from the FAIR and Architecture working groups. In this section, after a description of the general software requirements (2.2), an overview of the main outcomes of these two WGs (respectively, 2.3.2 and 2.3.4) is provided. Section 2.3.3 is focusing on TRIPLE's answers to data FAIRness general requirements while section 2.3.5 is dedicated to the analysis of two interoperability levels, to which GOTRIPLE provides its solutions. Last section 2.3.6 presents a short update from the other WGs (Landscape, Rules of Participation, Skills and Training and Sustainability), focusing on the relevant outputs which are not strictly relevant for interoperability issues; for this reason, the practical translation of the requirements for TRIPLE is not even sketched there.

2.2 Software requirements

The EOSC Strategic implementation plan [3] clearly states that, in order to be EOSC compliant, software has to be open source by default. Exceptions will have to be justified and approved.

More specifically, as an EOSC service, the GOTRIPLE platform will have to satisfy the service maturity criteria defined for services listed in the EOSC Portal which defines the Technology Readiness level of the service.

As of this writing, the EOSC-hub service maturity classification [4] applies which might be changed or improved with the EOSC Enhance project. The classification focuses on the operational aspects of the service and includes user support through documentation and helpdesk as well as privacy policy, terms of use and an SLA. On the technical side, evidence that the services and its functionalities meets user expectations and requirements as well as evidence for real use by end users (use cases) have to be provided.

For the future TRIPLE platform to be technically mature, software development must follow common industry standards and best practices. The infrastructures CESSDA, CLARIN, DARIAH and OPERAS have joined forces under the umbrella of the European Research Infrastructure Engineers' Network (EURISE Network) to exchange knowledge and experiences on the topic. Therefore, the EURISE Technical Reference [5], a collection of best practices collaboratively maintained by the infrastructures, will be followed. The Reference outlines guidelines for developers, operational, policy recommendation and as well as some approaches to software quality.

A few overall questions on the technical development have been clarified prior to the actual development of TRIPLE. Such questions, and the decisions made, try to follow the fact that GOTRIPLE will be an Open Infrastructure. As the development of the core platform will be mostly done by Huma-Num, their current workflow mostly applies for GOTRIPLE, such as the use of their own Gitlab instance (for Source Code repository), as well as CI (Continuous Integration done with Gitlab-CI). Development will be in general, and when possible, done in an agile manner with a team which will also periodically code review each other's work and make use of git's merge requests to try to achieve a 2 weeks sprint rotation. As for all projects this size, testing is important in the development, and will include unit tests, functional testing, in different separated environments (such as development, test, acceptance and production), and will be documented for future developers/maintainers' take over, but as well for users and managers' uses. GOTRIPLE Core will be Open Source, even though, until now, no definite answer on a specific license has been made, we are looking at EURISE Network's propositions for Apache-2.0 [6], EUPL-1.2 [7], or outside such as Etalab-2.0 [8], MIT [9], etc. However, some of its pre-existing innovative services might not be Open Source and therefore not follow this license scheme.

In addition to the recommendations for software development, the use of automation should be considered. The EOSC-Synergy project is developing a Software Quality as a Service[3] solution. This solution will be based on the common software quality assurance baseline criteria[4] developed by the INDIGO-DataCloud, DEEP-Hybrid-DataCloud, eXtreme-DataCloud and

[3] https://www.eosc-synergy.eu/home/software-services/

[4] https://indigo-dc.github.io/sqa-baseline/

EOSC-Synergy projects. These criteria should therefore be applied to the GOTRIPLE platform and followed from the initial development stage.

2.3 EOSC WGs main recommendations

2.3.1 Introduction

The Interoperability Task Force of the EOSC FAIR Working Group, with participation from the Architecture WG, released the EOSC Interoperability Framework [2], currently out for comments.

The report, to which the TRIPLE provided comments and feedback[5], is the main general reference on interoperability.

The document introduces the term Digital Object "to refer to the kind of objects that allow binding all critical information about any entity" which include research data, software, scientific workflows, hardware designs, protocols, provenance logs, publications, presentations, etc., as well as all their metadata (page 7).

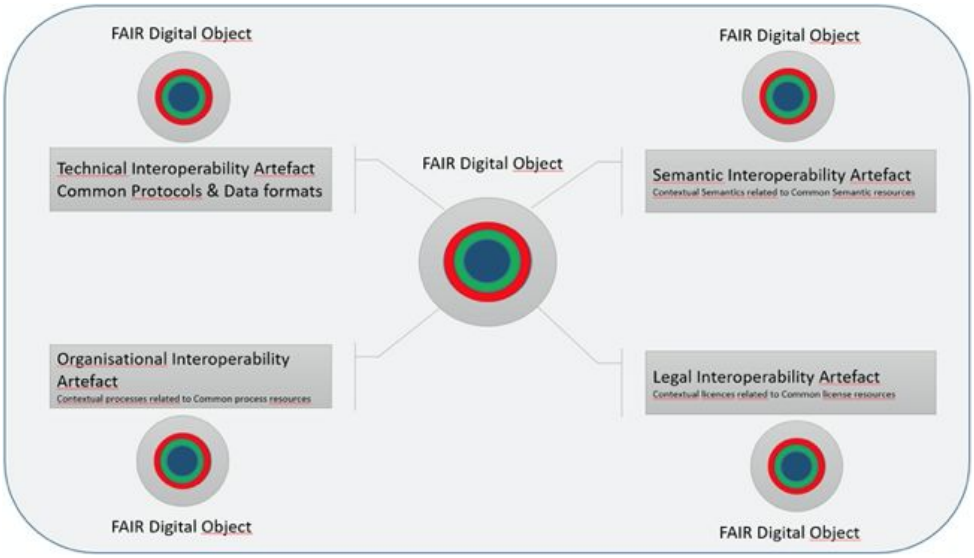

FIGURE 1. The EOSC Interoperability Framework

Shown in Figure 1, the framework is based on FAIR digital objects with PID links to common artefacts, and the key components include: 1) The FAIR digital object and metadata references provisioned by the PID infrastructure (kernel metadata); 2) An ecosystem of services resolving

[5] TRIPLE project gathered and submitted their comments to the EOSC IF (v.1.0). The full list of comments is available at the following link:
https://docs.google.com/document/d/12Cn7_vKv3QW7yr6KUKUbhRlvcZ7oPXUvXWZIG0g_HtY/edit

the references into interoperability artefacts; 3) Interlinkage between interoperability artefacts (FAIR Digital objects) to manage granularity etc.; 4)Principles/conceptual framework(-s) on the [abstract] model(-s) of the interoperability artefacts

Meant to be a generic framework that can be used by all the entities which take part in the design and development of the EOSC, the Interoperability Framework provides a common definition of the requirements, challenges and recommendations to be taken into account, and a set of principles to be followed on how to address these recommendations.

The document identifies the general principles that should drive the creation of the EOSC Interoperability Framework, which, according to the European Interoperability Framework [10], are organised into four layers: technical, semantic, organisational and legal interoperability (the last point is not addressed in the current version).

Technical interoperability includes the following needs:

- ☐ A common **Authentication and Authorisation Infrastructure (AAI)** (see: Architecture WG AAI principles)
- ☐ The availability of research data in **multiple general-purpose formats or community based models** stresses the need of standard alignment, with a trust (and sustainability) framework (see: Architecture WG AAI principles).
- ☐ Need to **find/query coarse-grained or fine-grained research data** from other communities, without knowledge about how to query specific domain repositories.
- ☐ Need to have a common and well-understood **PID policy** (see: PID Technical Architecture charter).

Semantic interoperability includes the following needs:

- ☐ Need for **principled approaches and tools for ontology and metadata schema** creation, maintenance, governance and use.
- ☐ Need for **harmonisation across disciplines**.
- ☐ Need to **harmonise the same type of data**.
- ☐ Need for **federated access** over existing research data repositories (both inside a discipline and across disciplines).

Organizational interoperability includes the following needs:

- ☐ A **governance framework** that includes clear instructions on how the other levels of interoperability will be handled across organisations and user communities (data formats, AAI services, metadata schemas, ontologies, etc.).

- ☐ **Documents explaining terms and conditions** and acceptable use policies for services providing interoperability.

- ☐ **Interoperability certification mechanisms** for service providers.

TRIPLE consortium analysed the report and provided some general and specific comments to it. Comments have been articulated as answers to the following questions:

Q: Is this what you expected to see or are some things missing?

A: Overall, it's a good overview of the main interoperability issues and possible solutions, although the different chapters are often of unequal level of detail.

The 4 layers identification seems efficient to distinguish the actions to take; at the same time it is not always very clear who should take care of which (legal: data stewards? semantics: infra and communities? technical: infra and EOSC?)

Positive aspects include:

1. Very positive that Semantic Artefacts are mentioned as first-class citizens, but not all problems that relate to their creation and maintenance are mentioned (such as scalability).
2. It brings a list of problems that have been gathered from different actors of EOSC - always interesting to know and think about.
3. Positive that publications are mentioned as a research output; this is essential for SSH and especially for the Humanities.

Negative aspects are:

1. Some drawbacks include: the document is in high abstraction, without any guidelines how to design and specify a FAIR digital objects, neither discussions about the feasibility for implementation. It relies on a common referent model/metadata standard to describe and interpret EOSC data and services, however, it would be very hard to create such a model/standard that addresses the requirements of various EOSC stakeholders who come from different domains of science and beyond. It would be even harder to reach the consensus for all EOSC service providers and users to agree on such a model/standard, not even mention the arduous efforts needed for enforcement of adoption and implementation. The concern is although it is a sound approach but infeasible for realisation.
2. There's a general lack of detail. Since some interviews have been performed, one could expect a more elaborated approach to various metadata schemes applied by different

"research communities". The document notes multiple standards but doesn't mention possible mappings. Perhaps a way to go here would be to work on minimal requirements for metadata interoperability.

3. The document lacks solutions in most chapters. And when there are solutions provided, there is no explanation on how to achieve them (e.g. share concepts' definition in EOSC). A more detailed "hands-on" description was expected. Looks like a collection of best practices, not really a (technical) "framework" description.

Q: **Are the concepts clear or do some aspects need further clarification?**

A:

1. Organizational interoperability is focused on as the need for alignment of public organizations dealing with standards and implementation, but for research 'social' interoperability seems more important, which is the need to organize communities to contribute to for instance Semantic Artefacts.
2. Very good that Digital Objects (DO) are mentioned, it's very welcome that not everything is a service. This terminology is more close to what the SSH community is familiar with. Nevertheless, it should include a clear interpretation of the concept of FAIR Digital Object in the context of the proposed EOSC Interoperability Framework. Does EOSC implies/equals FAIRfulness? If not, what is missing/constrained?
3. Legal interoperability is not defined. If this is part of the overall interoperability model, it is difficult to evaluate the completeness of the current proposal.
4. Maybe this is EOSC lingo but we find difficulties in grasping the difference between research communities and disciplines in the context of interoperability.The language should be more precise with regards to the target groups and their problems, or, at least, link to another document with a more precise definition of the target communities.
5. The "Scientific workflows" described in 2.1 are not clear, and haven't found a correct explanation of it in 3.1.

Q: **Are the minimum requirements and recommendations appropriate?**

A:

1. We repeat here that they're quite broadly sketched, thus the document is too generic. It is understandable that research communities would use different metadata formats, but there are no minimum requirements linked to them explained in the document.
2. Recommandations exist, yes and seem appropriate.

Q: **Is it clear who is responsible for what and how this should be followed?**

A: Not always: what is concretely intended by « PID infrastructure » and who is responsible for it? Who will take in charge the definition of a "simple vocabulary ... allowing discovery over

existing federated research data and metadata (extension of DCAT-AP, DDI 4 Core, or DataCite core schema)", knowing that, for instance, the SSHOC project is working on the same topic?

Q: As a service provider, could you conform with / implement the framework?

A:

1. At least in some aspects, but the framework at this stage is still too generic.
2. It's not clear what role to play in each case (e.g. on legal interoperability between various sources of various countries: should we provide a policy, recommendations, share information?).
3. A "use cases" companion document should be produced. To ensure interoperability you cannot remain at such a high-level description of the technical solutions to apply. There should be at least a reference implementation of this framework (maybe it's just me that got confused with the term "framework"...).

Q: Is the model for FAIR Digital Objects sound?

A:

1. TRIPLE's partners perspectives are different to this respect: CLARIN ones welcome the DFOs model, while for other it is a bit confusing that it's not totally similar to the model developed within GOFAIR; various FDOs flavors are possible, however the GOFAIR model only envisioned ID+resolvable link+type, while the model here described, with all semantics, is much more complex and therefore difficult to reach.
2. In general, within the various SSH communities there are different levels of readiness regarding FDOs (sometimes far from using more than a generic metadata standard, to not talk about semantics).

Q: What other feedback and comments would you like to offer?

A:

1. We'd suggest working on a reference implementation and/or a use cases document for this framework, at least for the technical and semantic interoperability aspects, with a specific and detailed description of the technical solutions to apply.
2. It would be great to see slightly more information about the interviews with stakeholders - the disciplines are listed but countries, career stage and the number of interviews would be helpful in understanding who guided the recommendations.

In the following sections, we analyse the main outputs of the EOSC WGs that complement the main points of the EOSC IF, namely AAI service, the PID policy, the metadata and ontology schema, and other relevant decisions in terms of interoperability. Section 2.4 presents a synthesis of TRIPLE (provisional) decisions in terms of EOSC integration.

2.3.2 FAIR requirements

The EOSC FAIR Working group provides "recommendations on the implementation of Open and FAIR practices within the EOSC", outlining the connection between FAIR and Open Science principles, although it is frequently reminded that "FAIR" doesn't mean "open".

Divided in four task forces, focusing on PID Policy, FAIR Practice, Interoperability and Metrics & Certification, the FAIR WG works closely with the Architecture WG with this distribution of roles: "the former addresses cultural aspects such as semantic and legal interoperability, certification and community data standards, while the latter focuses on the related technical specifications that address FAIR requirements."

Fundamentally based on the EC experts 2018 report "Turning FAIR into reality" [11], the FAIR WG relies also on the activities of the FAIRsFAIR project, the FREYA project and RDA (especially: GEDE-RDA, IG Data fabric, FAIR data maturity model).

The FAIR WG recommendations are focused on the following topics:

- ☐ The development and adoption of data standards and sharing agreements
- ☐ Best practices that are already applied in specific scientific domains or countries and can be adopted at the multi-disciplinary and European levels
- ☐ An EOSC Interoperability Framework that overarches disciplinary approaches and encourages research infrastructures to be interoperable-by-design
- ☐ Service requirements for FAIR implementation, relevant to the Architecture WG
- ☐ A Persistent Identifier (PID) policy for the EOSC
- ☐ Frameworks to assess FAIR data and certify services that enable FAIR, including the collation of results e.g. catalogues of certified repositories
- ☐ The international dimension of FAIR principles, converging towards globally-accepted frameworks.

According to the defined **roadmap,** results from testing, an updated PID policy and an updated scheme for accrediting FAIR data and certifying repositories are planned to be released by the end of 2020.

The main **outputs** of the WG include the following documents:

1. **FAIR WG work plan, Dec. 2019** [12];

2. **FAIR metrics for EOSC (Provisional), Feb. 2020** [13]:

The document reports on the activities of the RDA WG on the FAIR data maturity model, the FAIRsFAIR project, and more focused works (e.g. FAIR software). The FAIR metrics and the FAIR assessment tools are intended to guide progression towards FAIRness - which partly contradicts the fact that the FAIR metrics will also be part of the FAIR certification: are the FAIR metrics an auto-assessment tool or a technical requirement to be part of the EOSC? The report contains a list of FAIR data indicators which will be detailed by the WG in a future work.

3. **EOSC service certification for FAIR outputs (Provisional), Feb. 2020** [14]:

The draft report mainly suggests using the CoreTrustSeal certification for repositories as a tool to build a FAIR ecosystem. The certification could then be used to establish a «European Network of trustworthy repositories». It is planned to combine the certification with the FAIR metrics. The document contains reports on various workshops and surveys organized by the FAIR WG and the project FAIRsFAIR which all seem to have a very provisional nature.

4. **PID policy for EOSC (Second version), May 2020** [15]:

The draft report (final version planned for October 2020) provides definitions and recommendations for a sustainable PID infrastructure. It contains details on technical requirements, distribution of roles, and governance. The link with FAIR principles, and more precisely FDOs, is explicit. Not very precise is the nature of the « PID infrastructure » itself, partly because the actual target audience of the policy is unclear, partly because the responsibility of EOSC as a legal entity in this context is mentioned but not defined.

The TRIPLE project follows the FAIR principles throughout the building process of the platform and for each type of data.

5. **Recommendations for Services in a FAIR Data Ecosystem, Aug. 2020** [16]:

The document reports on workshops co-organized by FAIRsFAIR, RDA Europe, OpenAIRE, EOSC-hub, and FREYA. The recommendations address the FAIR principles from an ecosystemic point of view, considering that there is a lot of activity around the concept of FAIR data « but it is much less clear what should be expected from a data service in the FAIR data ecosystem». The report thus analyses the gaps, both for each actor of the ecosystem and between each of them (researchers, data stewards, service providers, etc.). A first workshop was held for «service providers and research infrastructures», a second one with «research support staff and

researchers», each group defining its own recommendations. A third workshop established a prioritization process of the recommendations. Interestingly, the report notes that different skills have to be combined to realize a FAIR ecosystem (technical/domain expertise), and that there are some discrepancies between the «Turning FAIR into reality» report and the communities priorities, thus suggesting that the official roadmap for FAIRification could be reshaped through their insights.

2.3.3 TRIPLE's implementation of the FAIR requirements

Findability

Findability is obviously at the heart of the building of a discovery platform like GOTRIPLE. Technically, findability will be supported by the use of PIDs, either harvested from the providers' repositories or generated by the platform, for each searchable element; the PIDs will be registered in the metadata record. Rich minimal metadata will facilitate data discovery thanks to the establishment of a TRIPLE model using schema.org: the model will allow it to gather any useful information from the data providers, whatever the model they internally use, and offer it to the user in a seamless way. Findability will be further increased thanks to automatic enrichments: the contents will be classified according to 27 categories in 9 languages; recognized named entities will be automatically linked to widespread controlled vocabularies in various languages (English with LCSH, French with RAMEAU, Spanish with BNE subjects).

Accessibility

All the previous findability features will be part of the search interface, but data and metadata will be also accessible through free, open, and documented protocols, namely: OAI-PMH, SPARQL endpoint, and APIs.

Interoperability

GOTRIPLE will tag variable-level information in the most relevant open standards for SSH i.e. in Data Documentation Initiative (DDI), Text Encoding Initiative (TEI), Metadata Encoding and Transmission Standard (METS), Metadata Object Description Schema (MODS).

Metadata records produced by GOTRIPLE will be published using the following standard vocabularies: Component MetaData Infrastructure, Dublin Core Metadata Element Set and DCMI Metadata Terms. Moreover, metadata records published in RDF will use the following linked open data vocabularies: Data Catalog Vocabulary (DCAT), Open Digital Rights Language (ODRL), DDI-RDF Discovery Vocabulary (Disco).

GOTRIPLE will also use and manage a domain-relevant thesaurus (Frascati terms for SSH and their correspondence in Library of Congress descriptors) which respects also the FAIR principles:

use of PIDs, expression in a widespread open standard (SKOS), qualified metadata for alternate labels (nine languages).

REUSABILITY

GOTRIPLE will ensure the reusability of all the content that the project will create: the project grants Open Access to all project results, which will be published in Open Access Journals (Gold road) and, when relevant, deposited in Open Access repositories (Green road). All data and metadata (with the exclusion of the User Research Data) will be available in Open Access with open licenses allowing reuse.

Furthermore, TRIPLE will work closely with the data providers in order to have a consistent licensing policy either for data or for metadata.

2.3.4 Architecture requirements

The **goal** of the EOSC Architecture Working Group is to define the technical framework required to enable and sustain an evolving EOSC federation of systems. Such a technical framework includes standards, APIs and protocols that will facilitate interoperable services delivered by diverse providers. An interoperability layer is going to be defined (has to be agreed upon by relevant stakeholders) that allows to build the EOSC federation of systems.

It is rational for the Working Group to give emphasis on the interoperability layer, that is essential for communities' services, such as TRIPLE, to be interacting with EOSC. However, as already noticed 2.2.1, it is very challenging to have all relevant stakeholders to agree on interoperability specifications.

The working group aims to provide an in-depth independent review of the current offering and the required evolution of the EOSC technical architecture, its standards and best practices.

The planned activities in the **roadmap** include describing and/or defining:

- ☐ The EOSC core services and their interfaces
- ☐ The EOSC open source APIs for reuse by thematic services
- ☐ The EOSC portal components and federated catalogues of service offerings
- ☐ The EOSC data description standards
- ☐ Any other standards and best practices necessary to ensure the evolution of EOSC and the widening of its user base to the industry and the public sectors.

Among this list, the EOSC open source APIs and the EOSC data description standards are more relevant to TRIPLE development, and will be closely followed up on.

Initially, five main types of services are identified as EOSC core, and they are:

- ☐ A unique identification and authentication service (AAI) and an access point and routing system towards the resources of the EOSC
- ☐ A protected and personalised work environment/space (e.g. logbook, settings, compliance record and pending issues)
- ☐ Access to relevant service information (status of the EOSC, list of federated data infrastructures, policy-related information, description of the compliance framework) and to specific guidelines (how to make data FAIR, to certify a repository or service, to procure joint services)
- ☐ Services to find, access, re-use and analyse research data generated by others, accessible through appropriate catalogues of datasets and data services (e.g. analytics, fusion, mining, processing)
- ☐ Services to make their own data FAIR, to store them and ensure long-term preservation.

By the time of writing this report, several task forces have been established to investigate some chosen topics, including PID, AAI, and EOSC service core. Some preliminary **outputs** include:

1. **EOSC AAI First Principles, Apr 2020** [17]

This report identifies three principles for EOSC AAI: 1)User experience is the only touchstone; 2) All trust flows from communities; 3) There is no center in a distributed system.

These principles clearly state that the design of the EOSC AAI will be community-driven, and the implementation will be a distributed architecture.

2. EOSC AAI Architecture 2019[6], Jun 2020

This report captures the current status of the EOSC AAI architecture discussions that are based on the AARC Blueprint Architecture 2019 [18]. It also identifies the challenges and the areas that require further work.

[6] EOSC AAI Architecture 2019. This is a draft report, currently shared internally among the Working Group Members.

The potential benefits are: Being a GOTRIPLE user, s/he can also access EOSC services. On the other hand, EOSC SSH researchers and other Science communities users by default become GOTRIPLE users and are able to use the GOTRIPLE platform -- this will enlarge TRIPLE user-base and make TRIPLE more visible to European science communities.

TRIPLE already includes the integration of EGI Check-in service in its technical roadmap. Check-in is an implementation of AARC Blueprint Architecture 2019, it is also one of the three EOSC AAI solutions, and adopted by EOSC Portal. Integration with EGI Check-in keeps TRIPLE AAI well interoperable with EOSC AAI. A description of this process is presented in the following section.

3. PID Architecture (draft)[7], Jun 2020

This report describes the main components of a global PIC architecture, and the PID registration and resolution framework. It discusses some existing technology for implementing such a PID framework, and examples of the PID services.

In GOTRIPLE, ORCID identifier is adopted for data registration and processing, which is interoperable with the proposed EOSC PID Architecture. TRIPLE also closely interacts with relevant EOSC projects such as FREYA, a 3-year project collaborating with OpenAIRE Advance and EOSC-hub and focusing on developing a PID infrastructure for EOSC.

2.3.5 TRIPLE's implementation: different interoperability levels

While planning GOTRIPLE integration into the EOSC from an architectural point of view, we envisage then two levels of integration.

First level integration: Integration of services in EOSC Marketplace

For the sake of harmonization with other OPERAS services and also to reuse already established working solutions, in this section we are going to follow the protocol established in OPERAS-P D4.5, Protocol for the integration of OPERAS RI services into EOSC[8]. The summary of the protocol can be found below and is complemented by extra contextual information largely gathered from two online training sessions of the EOSC Hub Week (19. and 20.05.2020): EOSC-hub Week 2020 - 4.1 Service Onboarding & Catalogue of services; EOSC hub Week 2020 - 7.1 EOSC portal service onboarding: Training for EOSC projects. At the time of writing, these are considered as the latest updates about onboarding services to the EOSC Marketplace, however, the EOSC-Enhance project will soon provide a novel way to onboard services to the EOSC via a web form. The exact timeline of delivering this is unknown but is supposed to be operational within the next few months.

[7] PID Architecture (draft). This is a draft report, currently shared internally among the Working Group Members.
[8] See OPERAS-P D4.5, since TRIPLE is an OPERAS service, reuse the same protocol/workflow.

What is onboarding?

In the context of EOSC, onboarding resources that are relevant to the scholarly communities in Europe means that providers connect them with the EOSC and through the EOSC, with a wide range of communities interested in knowledge production, following a protocol for integration. This protocol consists of a standardized administrative process aiming at collecting information about the resource in question and a validation process assessing the compliance of the provided information from the Service Providers thanks to a set of inclusivity criteria (see below).

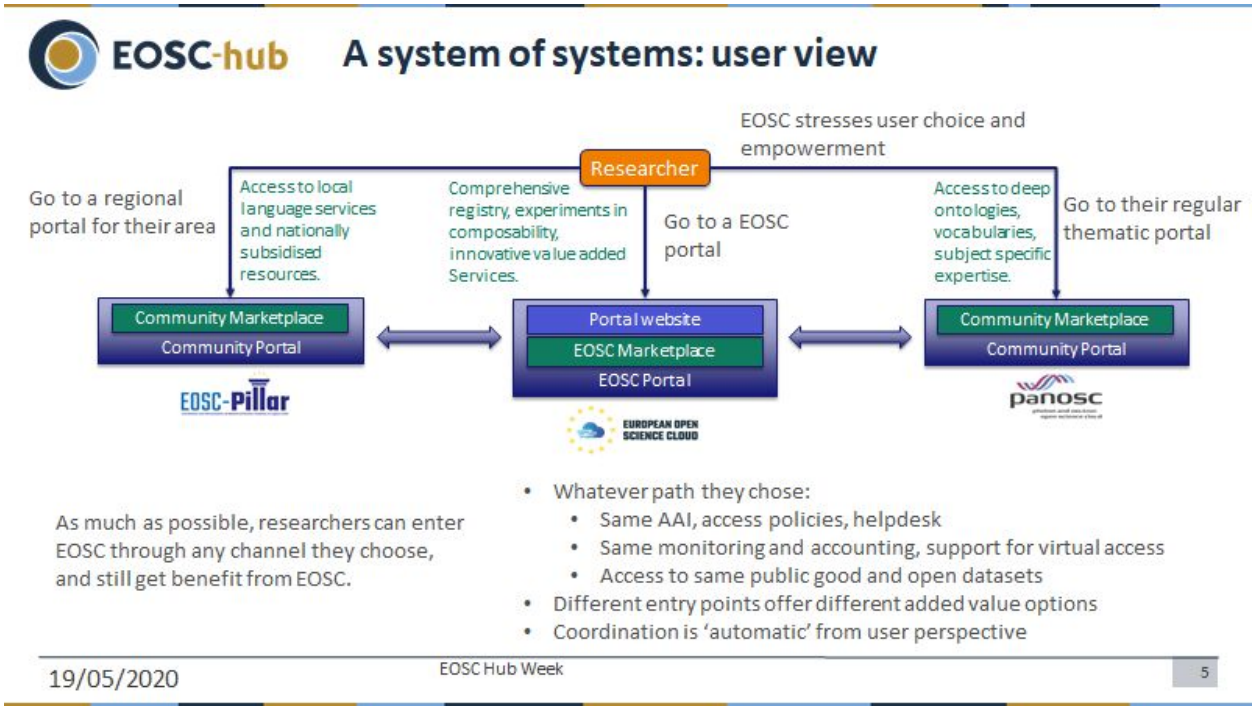

FIGURE 2. Multiple routes of onboarding services to the EOSC [21]

Onboarding services to the EOSC will also be possible through regional or thematic portals (e.g. EOSC Nordic, see these onboarding mechanisms explained on Figure 3) and their marketplaces as well, but current and established work around onboarding is centered around the central route, namely, integration through the EOSC Portal Marketplace, currently under heavy development in the EOSC Enhance project. At a later stage, the TRIPLE service is also expected to be integrated within the SSHOC Marketplace, once it reaches production (until December 2020, the SSHOC Marketplace is a non-public Alpha release, at which point it will enter its public beta phase).

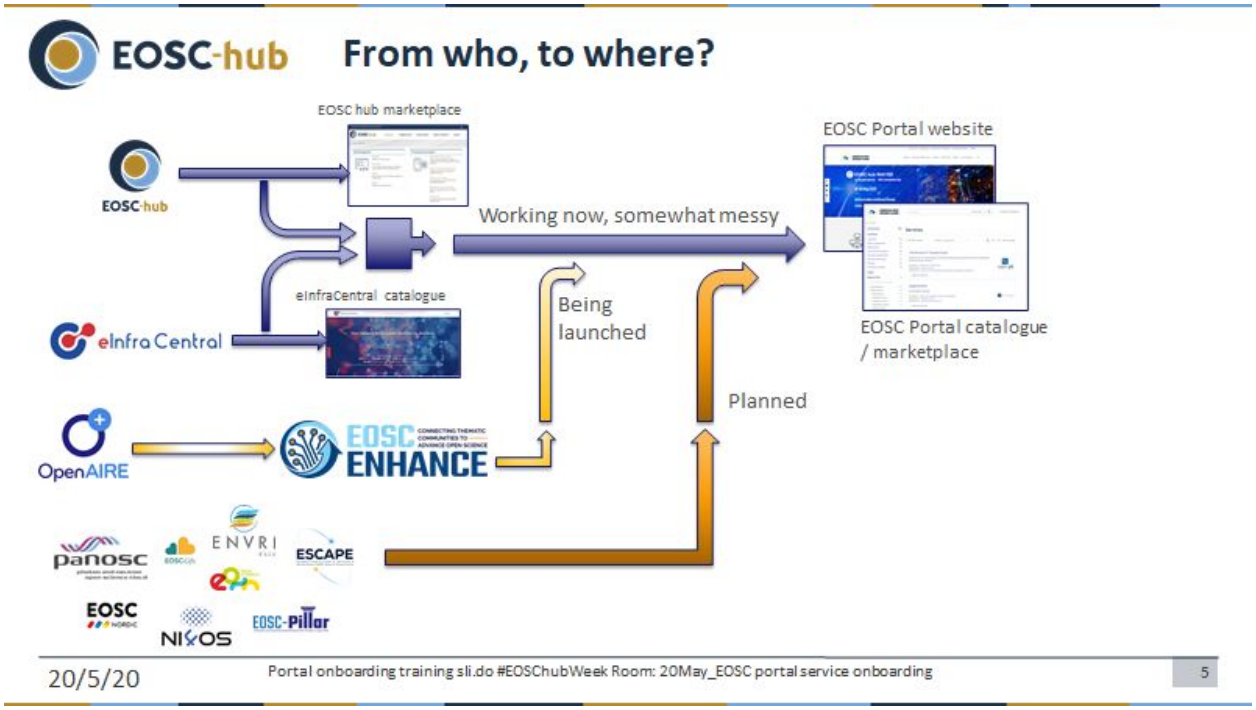

FIGURE 3. Onboarding workflows and enablers now and in the future[22]

This workflow is defined by preliminary work done in the EOSC Hub and the eInfra Central projects and is currently optimized for the integration of services. Within the framework of the EOSC Enhance project, this protocol will be extended to datasets and other scholarly resource types (e.g. training resources) as well, while integration through the regional and thematic clouds will enable giving access to specialised, context-dependent knowledge structures, for instance, e.g. implementing language services in regional portals or detailed, domain-specific ontologies and vocabularies in the thematic portals. In the EOSC, data repository and data discovery services, like GOTRIPLE, are onboarded as services within the EOSC Marketplace for the benefit of SSH Researchers.

The current onboarding process for service providers

Before providing a summary of the current protocol, it is important to emphasize that at this point of maturity of the EOSC Portal, the onboarding workflow is under continuous development which means that although drastic changes are not expected between the time of writing this deliverable and the time of integrating GOTRIPLE service to the EOSC Portal marketplace, certain aspects of the workflow (see examples below) and the currently used Service Description Template (Currently in production: EOSC-HUB Service Description Template v1.3, soon to be updated to v3.0.0 already available as beta by EOSC Enhance project) are expected to change.

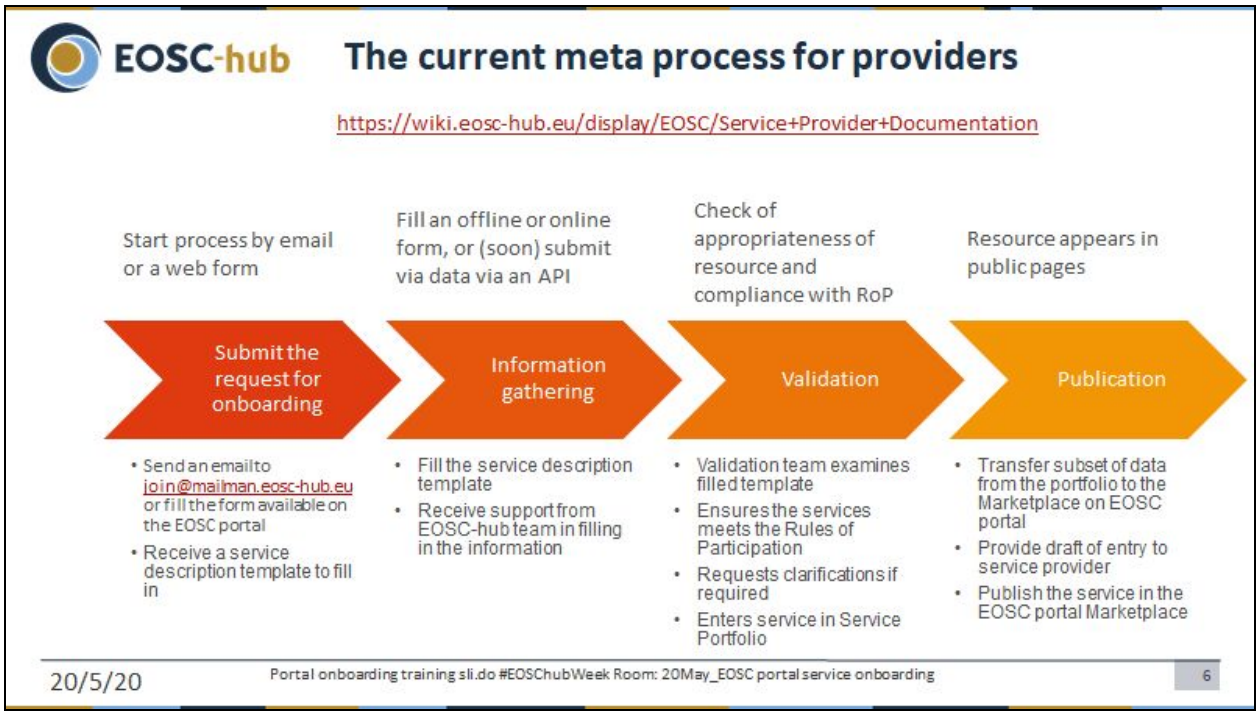

FIGURE 4. The current onboarding workflow [23]

Figure 4 summarizes the current onboarding process for resource providers[9] from submission to publication. Current limitations and future directions of improvement are also made explicit. In the future, opportunities for replacing the first step of initiating contacts involving written communications (e-mail or web form) will be complemented by API submission possibilities as well. This will be implemented under the EOSC Enhance project.

The validation process, which is currently a semi-automated process, is based on the Criteria for possible inclusion in the EOSC Service Portfolio which is a practical implementation of the EOSC Rules of Participation, and the Service maturity classification (Technology Readiness Level - TRL) descriptions. Following this classification, the TRIPLE service, at the end of the project, is expected to start from TRL8 (production ready) with perspectives to reach TRL9 later.

In addition to the above cited documents and training materials, the EOSC-hub Integration handbook for service providers [24] is available as support.

Protocol for the integration of OPERAS RI services into EOSC

For the sake of harmonization with other OPERAS services and also to reuse already established, working solutions, we are going to follow the protocol established in OPERAS-P D4.5, Protocol for the integration of OPERAS RI services into EOSC. The main steps of this

[9] see also: https://wiki.eosc-hub.eu/display/EOSC/Service+Provider+Documentation

protocol include: A) the description of the service, followed by B) the application to become an EOSC service provider (via the EOSC website). A key component of the application is C)-D) filling in and submitting the latest version of the EOSC Service Description Template where all the information of the service and its provider are described. E) The validation of the phase will be realized in iterated interactions with the EOSC-Hub onboarding team. F) After a final validation and proofreading phase, the EOSC-Hub onboarding team will transport (meta)data from the Service Description Template to the EOSC Marketplace. G) As a final step, once the online draft is approved by the service provider, the onboarding team publishes the service in the EOSC Marketplace so that it becomes available to researchers.[10]

Issues, challenges, open questions relevant to the TRIPLE integration

In this scenario, open questions concern the service integration workflow. Currently, we can consider the onboarding protocol as a moving target that has been changed considerably even during the writing phase of this deliverable. As a solution, we decided to keep OPERAS-P D 4.5 as a living document aiming to be updated with e.g. developments within the EOSC Enhance project.

Second level integration: Integration of EOSC Core services (e.g. AAI)[11]

A second level integration of the EOSC can be achieved by implementing some of the EOSC Federation Services[12] into our Services, such as the GOTRIPLE core platform and innovative services. Such EOSC Federation Services include, but are not limited to, the AAI, the Helpdesk, the Accounting Service, the Monitoring Service, etc. Some, such as the Accounting Service, have probably little to do with TRIPLE's aim, but others, especially the Helpdesk or the AAI can be important additions to our platform.

Following OPERAS Research Infrastructure decision to use the EGI Check-in, GOTRIPLE will follow to be compliant with the other OPERAS Core Services. EGI Check-in is, at the time of writing, also one of the implementations that is used for the EOSC AAI, even if the full EOSC AAI will only be available in 2023[13]. The implementation of the EGI Check-in AAI solution within the platform will allow the OPERAS Virtual Organisation to be used for managing the users on the platform.

At the time of writing this deliverable, OPERAS-P is preparing a public deliverable (D4.2 "AAI Service Specifications") for the use of EGI Check-in and how to connect Services (existent or in development) to this AAI solution and how to be able to make use of the user management thanks to the OPERAS Virtual Organisation (OPERAS VO). This VO, even though dedicated and

[10] Steps A)-G) follow the exact structure of OPERAS-P Deliverable 4.5.

[11] See OPERAS-P D4.2; AAI is being implemented for OPERAS services (Certification Service, Publication Service Portal, ...) including the GOTRIPLE platform and its innovative services.

[12] https://wiki.eosc-hub.eu/display/EOSC/Documentation+for+Federation+Services

[13] https://www.eoscsecretariat.eu/sites/default/files/open_consultation_booklet_sria-eosc_20-july-2020.pdf

managed by OPERAS directly, allows the OPERAS managers to have different user subgroups managed directly by trusted persons. It enables the TRIPLE team to manage a certain subgroup of users specific to the TRIPLE service autonomously, without any input from the OPERAS VO managers.

Another interesting EOSC Federation Service that could be integrated in the TRIPLE platform could be the Helpdesk. The current implementation, done in the EOSC-Hub project, is GGUS which is developed and provided by Karlsruhe Institute of Technology (KIT). However, there are plans, mentioned during the recent EOSC-Hub week talk by Pavel Weber (KIT)[14], to move to an OTRS[15] implementation for the EOSC Helpdesk. The presentation mentions 3 different levels of integration of the Helpdesk with external Services, where TRIPLE could be positioned: 1. Direct usage of the EOSC Helpdesk by the TRIPLE team (answers and follow-up happen on the EOSC Helpdesk), 2. A Ticket redirection from the EOSC Helpdesk towards the TRIPLE Helpdesk (or other Helpdesk) is performed, likely via an automatic email notification, and 3. Full integration thanks to the use of OTRS APIs between EOSC Helpdesk and the Service helpdesk where the issue is then taken care of. However, the EOSC Helpdesk would only be available starting 2021 at best (possibly even after 2023), after testing and validation is done by the various stakeholders.

There are other levels of integration, for example by applying the ordering facilities of the EOSC Marketplace to our services. However, such ordering facilities does not yet make sense for the GOTRIPLE platform.

[14] https://www.eosc-hub.eu/eosc-hub-week-2020/agenda/eosc-hub-contribution-eosc-architecture

[15] See also: https://community.otrs.com/

2.3.6 Other requirements

The monitoring of all EOSC Working groups is an activity whose outcomes are important for the TRIPLE project as a whole. As already mentioned in the previous section, the WP6 put in place a methodology to ensure the continuous update and progress on this activity. In this chapter we summarise briefly the most relevant outputs coming from the working groups whose activity is less relevant in relation to the interoperability itself, but important for the TRIPLE project in general. For this reason, in this section requirements are presented but not explicitly addressed to TRIPLE's context and choices.

LANDSCAPE

EOSC Landscape Working Group is an activity task of providing alignment and convergence of major European Structure and Initiatives. Structures and initiatives involved in the landscaping exercise include national research infrastructures and e-infrastructures, national open science policies, ESFRI RIs and cluster projects, thematic initiatives and clouds, EOSC-relevant H2020 projects and other international working groups.

The main **objectives** of the Landscape WG are as follows: 1) Deliver mapping of EOSC-relevant and current level spending on research data infrastructure; 2) Take account of federated challenges and opportunities at various national and regional structures and initiatives; 3) Conduct an analysis of Member states' readiness to provide both financial and political willingness towards infrastructure planning of EOSC; 4) and Finally propose standards and best practices to facilitate the alignment and convergence.

The final report of the Landscape WG is yet to be released in Q3 2020, and a Validation workshop of the WG is happening in the days of writing the present deliverable. However, the project drafted a comprehensive list of all European infrastructures and what they do, received inputs from several surveys including survey of EOSC 5b (national/regional) projects as well as engaged in-depth discussions and collaborations with other working groups.

The Landscape WG also made meaningful contributions in different ways towards other specific EOSC related projects and/or activities:

- ☐ EOSC Nordic – Contributing to the project to promote openness of research data in the Nordic and Baltic countries
- ☐ EOSC Synergy – Landscape WG made contributions in terms of understanding national plans/roadmaps for e-infrastructures, how these national plans feed the needs of EOSC. Also contributed best practices from national plans to overcome transnational barriers for collaboration
- ☐ E-IRG – EOSC PILLAR drafted national surveys (which include the horizontal and thematic e-Infrastructures per country). Survey had target audiences such as e-infrastructure

providers, RI, universities and funding bodies in Europe. The Landscape WG contributed to the survey by providing aggregated summaries of the landscape analysis, good practice examples and country sheets.

Rules of Participation

In the EOSC secretariat a specific Working Group is designed to take care of the so-called EOSC Rules of Participation. A request for comments, to which TRIPLE consortium replied to[16], was launched in January 2020.

This Working group is in charge of recommending a minimal set of clear Rules of Participation i.e. the rights, obligations and accountability which regulate all EOSC transactions by the various EOSC users, providers and operators.

The Rules of Participation [19] are defined so as to act as a common denominator across the heterogeneous requirements of European Research Infrastructures and service providers; and to respond to the principles of openness, transparency and inclusiveness.

They can be differentiated on the basis of: different users and different services (e.g. data producers, service providers, data/service users); different scientific disciplines; Different level of readiness of infrastructures (and their differently established rules); different types of providers and users involved in the EOSC (e.g. public versus private, horizontal versus specialized); different existing legal frameworks in Europe (GDPR, etc.); and continuously growing users' audience.

The same Rules of Participation will apply differently to different EOSC participants, depending on the maturity and role (providers vs. users, scientists or innovators), and location (European vs. global research partners).

The base for the design of such rules has been laid by the EOSCpilot project, while the EOSC Declaration set the general principles.

The Rules should address:

- ☐ The use of the tools, specifications, catalogues and standards (EOSC shared resources) and applicable framework (FAIR research data)
- ☐ The principles for regulating transactions (e.g. financial mechanisms and procedures, agreements/bylaws established by the EOSC governance framework)
- ☐ The applicable legal frameworks (e.g. GDPR, copyright, Data Security and Cybercrime).

[16] TRIPLE consortium Answers to the open consultation on the EOSC Rules of Participation are available here:
https://docs.google.com/document/d/1krF-ddn7Gn5AXyzzOmqG-B4EwhuEQZY_bhZB_IDsWCQ/edit

Sustainability

The **overall mission** of the Sustainability WG is to provide a set of recommendations concerning the implementation of an operational, scalable and sustainable EOSC federation after 2020. Interoperability issues are not prominent in the WG agenda, though in the outputs there are some references to interoperability issues at various layers.

Among the current results from the Working Group, the most relevant **output** for TRIPLE are the iterations of the internal documents "Solutions for a Sustainable EOSC"[17]. As one of the stages of development for EOSC the document presents a schematic representation of key elements of the 'Minimum Viable EOSC'. The MVE is supposed to be composed of EOSC-Core which will provide the means to discover, share, access and re-use data and services. Some of the service categories listed mention interoperability (e.g. in relation to authentication & authorization, architecture and metadata), but for details the description of MVE is referring to the work of WGs Architecture and FAIR.

Skills and Training

The Skills and Training WG had been established in February, 2020 due to the recognition that in addition to data, services, and their documentations, providing a framework for a sustainable training infrastructure to support EOSC in all its phases to ensure its broad community uptake will be a key success criterium of the EOSC. To realize this, the WG's mission is to build competence (skills) and capabilities (training) for EOSC. The WG consults and converges with existing initiatives and H2020 EOSC projects to agree upon key components for skills development and training, determine how they can be embedded in different levels of EOSC (institutional, national, EU), and identify what structures are needed to make EOSC a viable success (sustainability).

The Group is currently working on the Rules of Participation for EOSC training service providers. During May, the WG delivered a draft of the SRIA highlighting the gaps, the types of skills needed and the key priority areas in terms of skills development, training and education.

On the 3rd of August 2020 the "Minimum EOSC Skill Set" Task force of the Skills & Training Working Group released a diagram to describe the EOSC actors profiles and their interactions with the EOSC ecosystem, required skills and training needs for consultation. One of the main questions and organizational challenges the WG is facing, and this is certainly not a unique issue across the EOSC WGs, is to meaningfully harmonize the national vs. the thematic (or domain-specific) angles of organization and content, i.e. keeping the national angle when enumerating competence centres but at the same time also making sure that researchers from every discipline find relevant training to interact with the EOSC. In the context of TRIPLE, positioning discipline-specific training in the EOSC might be an especially relevant question.

[17] The so-called tinman report (Solutions for a Sustainable EOSC), which is the successor of the strawman report of 2019, and after a consultation round is to be followed by the Iron Lady version, is inexplicably not accessible, and thus can't be properly cited.

The workshop report 'Training in the EOSC' [20] highlights the main directions towards the EOSC Rules of Participation adapted for training materials.

3| FURTHER EOSC-RELATED PROJECTS OUTPUTS

WP6 partners have carried out a mapping exercise of relevant deliverables produced by the main EOSC related projects in order to provide a comprehensive analysis of the EOSC interoperability requirements. The analysis was aimed to understand the developments of the EOSC environment in terms of interoperability and at the same time to understand which public deliverables have to be taken into consideration for the overall project development in TRIPLE. A total of 11 projects have been investigated and 38 relevant deliverables have been thoroughly analysed. The table below offers the highlights of the mapping exercise that is extensively described in Annex 1. We envisage that TRIPLE has to continue monitoring the appearance of further deliverables from the listed set of projects, as it helps to develop to contextualise TRIPLE technology, and to ensure the compliance with the common standards. Explicit references to the relevance of specific deliverables are meant here to be shared with other TRIPLE Work packages (especially WPs 2, 3, 4 and 5), which may find them useful for their work.

Project	**Analysed Deliverables**
EOSC-hub brings together multiple service providers to create the Hub: a single contact point for European researchers and innovators to discover, access, use and reuse a broad spectrum of resources for advanced data-driven research.	*Deliverables related to Architecture WG* D4.2 Operational Infrastructure Roadmap - *relevant as it describes the guidelines for the actions that are to be taken in order to ensure interoperability at the level of EOSC-hub service catalogue which can be taken as lessons learned for the work in TRIPLE* D5.3 1st Report on maintenance and integration of federation and collaboration services D6.2 First report on the maintenance and integration of common services D7.2 First report on Thematic Service architecture and software integration

	D10.3 Technical Architecture and standards roadmap v1 - *relevant as it gives examples how Research Enablings services benefit from diverse features of Access Enabling services when being incorporated within a unified Hub.* D10.4 EOSC Hub Technical Architecture and standards roadmap v2 - *relevant for the TRIPLE plans for managing researchers' identity (WP4)* D10.5 Requirements and gap analysis report v1
FREYA is a 3-year project funded by the European Commission under the Horizon 2020 programme. The project aims to extend the infrastructure for persistent identifiers (PIDs) as a core component of open research, in the EU and globally. FREYA will improve discovery, navigation, retrieval, and access to research resources.	D2.1 PID Resolution Services Best Practices - *relevant for WP2 and WP4* D3.1 Survey of Current PID Services Landscape - *relevant for WP2, especially to discuss the needs of a TRIPLE ID* D3.2 Requirements for Selected New PID Services - *relevant for TRIPLE WP2 and WP5, especially for the links to innovative services*
OpenAIRE-Advance continues the mission of OpenAIRE to support the Open Access/Open Data mandates in Europe. By sustaining the current successful infrastructure, comprising a human network and robust technical	D 4.2 – A multi-module Open science kit - *relevant for Task 6.3 as a preliminary work on Open Science training*

services, it consolidates its achievements while working to shift the momentum among its communities to Open Science, aiming to be a trusted e-Infrastructure within the realms of the European Open Science Cloud.	D 7.3. Interoperability with Research Infrastructures - relevant as it highlights how the work that focuses on services built on the basis of Open Science publishing practices support cross-communities communication and collaboration. Moreover, this deliverable allows to draw distinction between the OpenAIRE services and the implemented and envisaged ones of the TRIPLE project. These are: 1. Scope: The OpenAire Research Graph and its discovery service, OpenAire Explore and the Research Community Gateways aim to cover the full scholarly landscape while TRIPLE focuses on SSH only. 2. Differences in data model: the OpenAire data model distinguishes Publications, Research Data, Software and Other research outputs in their data model and there is also a possibility to search along Institutions, Projects, or Funders. By contrast, TRIPLE focuses on People, Projects and Publications. 3. Differences in functionalities: OpenAire offers the Research Community Gateways to RIs as a monitoring tool, this is something TRIPLE is not aiming for, but it will have a range of other functionalities where community curation plays an important role.
EOSC Enhance project is committed to improve the EOSC Portal by making it the added value one-stop shop/entry point for the EOSC users and stakeholders, by enabling easy access to EOSC resources such as services, data, scientific products and other resources to European scientists.	D 2.2 EOSC Processes Development and Consensus D 2.4: EOSC Service Catalogue Analysis - *relevant for TRIPLE because it facilitates the discoverability of EOSC resources across disciplines*

	D 3.1: EOSC Portal Functional and Non-Functional Specifications D 3. 2 : EOSC Portal Open APIs Specifications of Service and Resources Providers. - *Relevant for TRIPLE as it shows the requirements needed to be integrated in the EOSC portal*

Table 1. Mapping projects and deliverables

4| Conclusions and Outlook

In order to be integrated to and fully compliant with the EOSC, the GOTRIPLE platform has to follow different kinds of technical requirements, concerning data, software, the service(s), and users. The above mentioned requirements are evolving quite fastly due to the different agendas of the EOSC and the related projects. Moreover, TRIPLE has to be developed from SSH usages being the entry door for local repositories, which normally are not designed taking into account interoperability.

In sections 2 and 3, we presented the complex scenario of the EOSC definition, and the main EOSC WGs and EOSC related projects outputs which are relevant for GOTRIPLE interoperability - as well as our preliminary solutions in terms of integration of GOTRIPLE within the EOSC.

For a proper development of GOTRIPLE platform we will then focus on two aspects, or sides, of interoperability.

Firstly, TRIPLE work will be conducted taking into account internal aspects. This practically implies to check interoperability

1. Between the repositories and GOTRIPLE discovery system
2. Among the vocabularies in each of the SSH disciplines and in each of the nine languages of the platform, and
3. Between GOTRIPLE platform and the different innovative services.

This will contribute to empower interdisciplinarity among SSH researchers. Here, interdisciplinarity is a guiding principle to be conceived as a specific part of interoperability, being at the same time the aim of interoperability and one of the ways to achieve it.

Secondly, we will deal with external interoperability, which implies being fully integrated and compliant with the SSHOC/EOSC Marketplace and, more in general, the EOSC ecosystem. In this perspective, the TRIPLE project has not only the mission to develop a suite service interoperable with other EOSC services, but also to contribute to this general interoperability objective. To reach this objective, TRIPLE contribution will include the production and reuse of Open Science guidelines and best practices for a better service interoperability in the Social Sciences and Humanities, which is a specific objective of Task 6.3.

This deliverable highlighted mainly the external aspect of interoperability, presenting TRIPLE consortium contributes to the definition of global interoperability. Further TRIPLE deliverables, such as D2.1, D4.1, are more focused on the internal one.

As a general final consideration, WP6 team shared the idea to continue both the activity of monitoring the EOSC WGs work, and the mapping exercise for the whole duration of Task 6.1, in order to have an updated scenario on the EOSC interoperability framework, specifications and requirements to that date. An informal report on this will be released by M36, when the end of Task 6.1 is planned.

REFERENCES

[1] ESFRI Roadmap & Strategy Report, http://roadmap2018.esfri.eu/media/1066/esfri-roadmap-2018.pdf, 2018

[2] EOSC General Interoperability Framework v.1.0, May 2020, Draft for consultation, https://www.eoscsecretariat.eu/sites/default/files/eosc-interoperability-framework-v1.0.pdf

[3] Executive Board of the European Open Science Cloud (EOSC)-Strategic Implementation Plan, Section 5.6, https://dx.doi.org/10.2777/202370

[4] EOSC Hub Service Maturity Classification, https://wiki.eosc-hub.eu/display/EOSC/Service+Maturity+Classification

[5] EURISE Technical Reference, https://technical-reference.readthedocs.io/en/v0.2/

[6] Apache License 2.0, http://www.apache.org/licenses/LICENSE-2.0

[7] European Union Public License 1.2, https://joinup.ec.europa.eu/collection/eupl/eupl-text-eupl-12

[8] Etalab Open License 2.0, https://github.com/DISIC/politique-de-contribution-open-source/blob/master/LICENSE.pdf

[9] MIT License, https://opensource.org/licenses/MIT

[10] New European Interoperability Framework. Promoting seamless services and data flows for European public administrations, https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf

[11] EC experts 2018 report "Turning FAIR into reality", https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf

[12] FAIR WG work plan, Dec. 2019, https://op.europa.eu/en/publication-detail/-/publication/cf2b519d-1b69-11ea-8c1f-01aa75ed71a1/language-en

[13] FAIR metrics for EOSC (Provisional), Feb. 2020, https://repository.eoscsecretariat.eu/index.php/s/C3a5WkpsFHL6GD3#pdfviewer

[14] EOSC service certification for FAIR outputs (Provisional), Feb. 2020, https://repository.eoscsecretariat.eu/index.php/s/zCnHTcytBHaLjRp#pdfviewer

[15] PID policy for EOSC (Second version), May 2020, https://doi.org/10.5281/zenodo.3574202

[16] Recommendations for Services in a FAIR Data Ecosystem, Aug. 2020, https://doi.org/10.1016/j.patter.2020.100058

[17] EOSC AAI First Principles, https://docs.google.com/document/d/1BG9oq6nHEaUVQrQYNpexA2k9EN3vdjU3itxGWhZrYn0/edit#

[18] AARC Blueprint Architecture 2019: https://zenodo.org/record/3672785#.X1ZPQ2czai4

[19] European Open Science Cloud, Rules of Participation - Version 0.2 - 29 January 2020, https://repository.eoscsecretariat.eu/index.php/s/QWd7tZ7xSWJsesn#pdfviewer

[20] Marjan Grootveld, Ellen Leenarts, Ilona von Stein, Cees Hof, Giuseppe La Rocca, Iryna Kuchma, ... Katerina Zourou. (2020, April 3). Workshop report "Training in the EOSC" (Version 01.00). Zenodo. http://doi.org/10.5281/zenodo.3739055.

[21] Multiple routes of onboarding services to the EOSC, Owen (2020).

[22] Onboarding workflows and enablers now and in the future, Owen (2020).

[23] The current onboarding workflow. Source: Owen (2020) CC-BY 4.0.

[24] EOSC-hub Integration handbook for service providers, https://zenodo.org/record/3826907#.Xx6oPedS-uU

ANNEX I - EOSC-RELATED PROJECTS MAPPING

This section includes the full analysis of the mapping exercise which has been performed by WP6 partners to have a complete vision of relevant deliverables produced by the main EOSC related projects.

For each project, a list of relevant deliverables have been provided. Each deliverable was evaluated according to its relevance to EOSC WGs, relevance to the purpose of this deliverable, and main standards mentioned. Release date was also taken into account as some statements in deliverables might no longer be considered valid due to a natural evolution of the EOSC landscape over time.

It has been considered a very useful *exercise* to analyse not only the deliverables individually, but also to evaluate the results in an aggregated manner, and to have a simultaneous overview on the results.

The aim of this section is from one side to provide an overview of the main outcomes of this mapping exercise, and at the same time to highlight the meticulous deepening that has been carried out.

1. Analysed projects

A total of 38 deliverables from 11 different projects have been analysed. The analysis comprised the evaluation of deliverables both produced by ongoing and by finished projects. Each deliverable has been analysed taking into consideration the relevance not only for the current document, but also in relation to EOSC WGs and under the light of possible interaction with other TRIPLE WPs.

A list of **ongoing projects** includes:

1.1 EOSC-hub

EOSC-hub brings together multiple service providers to create the Hub: a single contact point for European researchers and innovators to discover, access, use and reuse a broad spectrum of resources for advanced data-driven research. For researchers, this will mean a broader access to services supporting their scientific discovery and collaboration across disciplinary and geographical boundaries. The project mobilises providers from the EGI Federation, EUDAT CDI, INDIGO-DataCloud and other major European research infrastructures to deliver a common catalogue of research data, services and software for research.

LIST OF ANALYSED DELIVERABLES

1. D4.2 Operational Infrastructure Roadmap

2. D5.3 1st Report on maintenance and integration of federation and collaboration services (currently, under review at EC)
3. D6.2 First report on the maintenance and integration of common services
4. D7.2 First report on Thematic Service architecture and software integration
5. D10.3 Technical Architecture and standards roadmap v1
6. D10.4 EOSC Hub Technical Architecture and standards roadmap v2
7. D10.5 Requirements and gap analysis report v1

MAIN OUTCOMES RELEVANT FOR THIS DELIVERABLE AND THE TRIPLE PROJECT

The analysed deliverables are all related to the Architecture WG.

D10.3 - Technical Architecture and standards roadmap v1 describes the Service Architecture of the EOSC-Hub, detailing the different service types, and defining their functions and their relationships with other components of the architecture and end users. Services in the Hub have been categorised in Access Enabling services and Research Enabling services. Access Enabling services includes Federation services (authentication and authorisation, monitoring, accounting, etc.) and Open Collaborative services (open science platforms for discovering and sharing research digital objects). Research Enabling services includes Common services (e.g. EGI Cloud Compute or EUDAT B2SAFE), and Thematic services (e.g. scientific applications offered by the Research Infrastructures).

D10.4 EOSC Hub Technical Architecture and standards roadmap v2 defines the EOSC technical architecture. The EOSC AAI follows the architectural and policy recommendations defined in the AARC project. As such, it enables interoperability across different Service Provider- and Identity Provider-Proxy services, each of which acts as a bridge between the community-managed proxies (termed Community AAIs) managing the researchers' identity and the generic services offered by Research Infrastructures and e-Infrastructures (termed R/e-Infrastructures or Infrastructures).

1.2 The Social Sciences and Humanities Open Cloud (SSHOC)

SSHOC is one of the thematic cluster projects of the EOSC, aiming to realize and develop the social sciences and humanities area of the EOSC. The project runs from January 2019 to April 2022, and provides both the infrastructural components of the SSH thematic cluster (such as the SSHOC Marketplace or software packages to establish local Dataverse instances, standardization frameworks etc.) and training materials to open up scholarly workflows in the SSH. Besides, SSHOC continuously monitors ongoing developments in the EOSC so as to conform to the necessary technical and other requirements for making the SSHOC services

sustainable beyond the duration of the project. Once it reaches its production-level maturity, the TRIPLE service will be harvested by the Marketplace.

LIST OF ANALYSED DELIVERABLES

1. D7.1 System Specification - SSH Open Marketplace
2. D3.1 Report on SSHOC (meta)data interoperability problems
3. D3.2 Inventory of SSH citation practices, and choice for SSHOC citation formats and implementation planning
4. D8.2 Certification plan for SSHOC repositories

MAIN OUTCOMES RELEVANT FOR THIS DELIVERABLE AND TRIPLE PROJECT

The EOSC interoperability solutions in the SSHOC project, which are considered to be the most relevant areas of the project for this deliverable will be, at the time of writing, future outputs of the WP3: *Lifting Technologies and Services into the SSH Cloud*. In the broader context of TRIPLE, the SSHOC *D 3.1. Report on SSHOC (meta)data interoperability problems* is especially relevant for the SSH data domain as it provides a timely inventory of data formats and metadata standards that are currently used and relevant for the research infrastructures currently managed by the SSHOC main stakeholders and makes recommendations of specific formats and standards for increasing interoperability, and prioritisations for providing conversion services and planning solutions. Besides, D7.1 System Specification - SSH Open Marketplace can be taken as a starting point for outlining the workflow for harvesting GOTRIPLE by the SSHOC Marketplace.

1.3 FAIRsFAIR

FAIRsFAIR - Fostering Fair Data Practices in Europe - is a 36 months H2020 project (March 1, 2019 - March 2021) aiming to supply practical solutions for the uptake of the FAIR data principles throughout European researcher and research support communities. FAIRsFAIR[18] plays a key role in the development of global standards for FAIR certification of repositories and the data within them contributing to those policies and practices that will turn the EOSC programme into a functioning infrastructure. FAIRsFAIR is also involved in delivering the FAIR dimensions of the Rules of Participation (RoP) and regulatory compliance for participation in the EOSC.

LIST OF ANALYSED DELIVERABLES

1. D2.1 Report on FAIR requirements for persistence and interoperability

[18] https://www.fairsfair.eu/fairsfair-nutshell

2. D4.5 Evaluation of procedures and processes of certification mechanism

MAIN OUTCOMES RELEVANT FOR THIS DELIVERABLE AND THE TRIPLE PROJECT

D 2.1. Report on FAIR requirements for persistence and interoperability 2019 reviews and documents commonalities and possible gaps regarding semantic interoperability, and the use of metadata and persistent identifiers across infrastructures. As such, it is a highly relevant document in the context of TRIPLE and can underpin decisions made in the development of TRIPLE in questions related to the choice of PID systems and services, semantic interoperability frameworks, an how to support FAIR data in the data life cycle. Furthermore, it gives an interview of how different components of FAIRness translate into infrastructure development. The study highlights the rapidity of change in technical solutions and wide variation across scientific domains in the uptake and among others, also outlines SSH-specific challenges such as multilingualism as a semantic interoperability challenge. That said, trans-language interoperability requires multilingual semantic artefacts (e. g., vocabularies, ontologies and concepts schemes in different EU languages).

D 4.5. Evaluation of procedures and processes of certification mechanism is a milestone report that aims to provide an evaluation of processes and procedures for FAIR-aligned repository evaluation and assessment in the EOSC. The report considers how repositories can enable the FAIR data principles and how FAIR data characteristics affect core certification of repositories. Such mechanism of evaluation of FAIR level can be relevant when assessing the quality and FAIRness of the different repositories and other data sources that GOTRIPLE will harvest.

1.4 FREYA

FREYA is a 3-year project funded by the European Commission under the Horizon 2020 programme. The project aims to extend the infrastructure for persistent identifiers (PIDs) as a core component of open research, in the EU and globally. FREYA will improve discovery, navigation, retrieval, and access to research resources. New provenance services will enable researchers to better evaluate data and make the scientific record more complete, reliable, and traceable. By engaging with the global community through the Research Data Alliance (RDA) and other research infrastructures, FREYA works to realise the vision of fully accessible data. FREYA follows on from the successful THOR project.

LIST OF ANALYSED DELIVERABLES

1. D2.1 PID Resolution Services Best Practices
2. D3.1Survey of Current PID Services Landscape
3. D3.2 Requirements for Selected New PID Services

MAIN OUTCOMES RELEVANT FOR THIS DELIVERABLE AND TRIPLE PROJECT

The deliverables are not particularly relevant for the purpose of the definition of TRIPLE's general interoperability requirements, however, they are relevant for TRIPLE WP2 and WP5 - especially for the links to innovative service, and in particular regarding the crowdfunding tool with the research campaigns (D3.2), while D3.1 has a relevance with WP2, especially to discuss the needs of a TRIPLE ID. It can be relevant as well for the publications part, while D2.1 has relevance for WP2 and WP4.

1.5 OpenAIRE Advance

OpenAIRE-Advance continues the mission of OpenAIRE to support the Open Access/Open Data mandates in Europe. By sustaining the current successful infrastructure, comprising a human network and robust technical services, it consolidates its achievements while working to shift the momentum among its communities to Open Science, aiming to be a trusted e-Infrastructure within the realms of the European Open Science Cloud. In this next phase, OpenAIRE-Advance strives to empower its National Open Access Desks (NOADs) so they become a pivotal part within their own national data infrastructures, positioning Open Access and Open Science onto national agendas.

LIST OF ANALYSED DELIVERABLES

1. D 4.2 – A multi-module Open science kit

2. D 7.3. Interoperability with Research Infrastructures

MAIN OUTCOMES RELEVANT FOR THIS DELIVERABLE AND TRIPLE PROJECT

This report provides an overview of a multi-module Open science kit – a diverse set of training materials on different open science aspects for life sciences, social sciences, humanities, computing, engineering and other disciplines being developed in collaboration with FOSTER project.

Targeted towards researchers, content providers, research managers, funders, research communities and innovators the kit aims to progress researchers from being aware of open science to being able to put open science into practice in their daily workflows.

The document is relevant for the EOSC Skills and Training WG. For what concerns the TRIPLE project, it is relevant for Task 6.3 as a preliminary work on Open Science training.

D 7.3. presents the activities carried out in the OpenAIRE-Advance project to promote and ease the adoption of Open Science publishing practices in the context of three research infrastructures: ELIXIR-GR, DARIAH-EU and the Italian node of EPOS. This is done by building Open Research gateways thanks to the OpenAIRE Research Community Dashboard (RCD), added value services on the top of the OpenAire research Graph. The results of this OpenAire

Advance task force (T 7.3.) are especially relevant for TRIPLE as the outputs are similar: both projects build discovery environments that go beyond the scope of publications in the strict sense and apply semantic interlinking techniques of different sorts to enable research communities in different disciplines to explore connections between research articles, data sets, the underlying software and research projects that would not be otherwise visible. One difference between the two discovery platforms, the TRIPLE platform vs. OpenAire Explore and its community gateways lie in the main components of their data models. In TRIPLE, the focus will be on data and publications, people and projects, while the OpenAire discovery environments are centered around Research Outcomes (including Publications, Research data, software and Other research outputs), Projects, Content providers and Organization. Since the OpenAire Advance project started years earlier than TRIPLE, there is a possibility to learn lessons from the OpenAire team's experiences and apply working practices in TRIPLE. As a concrete manifestation of this latter, the TRIPLE metadata standards will follow the same DataCite standards applied in the OpenAire discovery services. Since the OpenAire research graph is becoming one of the most exhaustive, living databases of digital scholarly objects, achieving bi-directional interoperability with it is a crucial component of any European level scholarly discovery services. Finally, the experiences and outcomes of T 7.3. That concerns DARIAH and building a community gateway to humanities scholars around DARIAH as especially relevant for TRIPLE as they outline the major domain-specific discovery challenges (no global discovery platform for the SSH, the multilingual and highly-fragmented publication landscape, low awareness of publication types other than legacy formats, varying metadata quality, PIDs still being an entry barrier for many content providers etc.) and address some of them. As a result of the project, two DARIAH services, TextGrid and NAKALA become integrated with the OpenAire discovery platform. The latter, NAKALA, plays a role as a data source in GOTRIPLE as well.

1.6 CO-OPERAS

CO-OPERAS – open access in the European research area through scholarly communication – IN aims to build a bridge between SSH data and the EOSC, widening the concept of "research data" to include all of the types of digital research output linked to scholarly communication that are, in SSH, part of the research process. The goal is to contribute to a better integration of SSH research objects into the EOSC, as a major component of the IFDS.

List of analysed deliverables

1. CO-OPERAS workshop - FAIR data Turin 10 September 2019 Report

Main outcomes relevant for this deliverable and TRIPLE project

The document is relevant for the FAIR EOSC WG. It has few technical specifics (standards, repositories, etc.) and focuses mainly on theoretical discussions and expression of needs. It

might be relevant to get the GOTRIPLE potential end-users and information on their FAIR practices to be integrated. As main outcomes from the workshop, it is worth to mention:

☐ Definitions of data

- dynamic/diachronic entity which represents a “process”
- data as “access and participation”
- data as anything that can be formalized in a language (including
- bits)
- the relation between data (mere registrations of a fact) and documents which lay upon deliberate interventions

☐ Huge lack of interoperability

☐ Urgent need of training on data issues since University curricula

1.7 RDA

The Research Data Alliance (RDA) was launched as a community-driven initiative in 2013 by the European Commission, the United States Government's National Science Foundation and National Institute of Standards and Technology, and the Australian Government’s Department of Innovation with the goal of building the social and technical infrastructure to enable open sharing and re-use of data.

LIST OF ANALYSED DELIVERABLES

1. RDA Metadata Standards Directory Working Group: Final Report
2. Legal Interoperability of Research Data: Principles and Implementation Guidelines
3. RDA Data Fabric Interest Group (2018): Summary of Virtual Layer Recommendations
4. Data Discovery Paradigms: User Requirements and Recommendations for Data Repositories
5. RDA/WDS Publishing Data Workflows WG Recommendations

MAIN OUTCOMES RELEVANT FOR THIS DELIVERABLE AND TRIPLE PROJECT

The reports provide useful inputs that might have relevance for the FAIR, Architecture and Skills and Training EOSC WG. For what concerns the TRIPLE project, the outcomes of this report have

been considered with a low relevance for the purpose of the publication of this deliverable, also because they have been published between 2016 and 2018 and the information might be obsolete or updated by other documents.

1.8 EOSC Enhance

EOSC Enhance project is committed to improve the EOSC Portal by making it the added value one-stop shop/entry point for the EOSC users and stakeholders, by enabling easy access to EOSC resources such as services, data, scientific products and other resources to European scientists.

EOSC Enhance seek to leverage the collective intelligence, experience and outcomes of key EOSC coordinating and contributing projects (EOSCpilot, EOSC-Hub,OpenAIRE-Advance, CatRIS, eInfraCentral, OCRE, RDA, ESFRI clusters and other thematic clouds and EOSC national and regional/thematic initiatives) to fulfil this objective by improving key functionalities of the EOSC Portal. The following are the objectives for the EOSC Enhance project:

- ☐ Enhance the EOSC provider interface and incorporate new EOSC resources into the EOSC Catalogue
- ☐ Accelerate the deployment and uptake of EOSC resources
- ☐ Increase user demand for EOSC resources via portal improvements and development
- ☐ Enabling easier access to EOSC Resources of thematic clouds

The revamped and improved EOSC Portal will facilitate interoperability and discoverability of EOSC resources across scientific disciplines by linking up thematic and regional/national providers (public and commercial), gateways, aggregators and marketplaces. The EOSC Portal will focus on processes, specifications, guidelines, tools and APIs to support providers in making them interoperable.

LIST OF ANALYSED DELIVERABLES

1. D 2.2 EOSC Processes Development and Consensus
2. D 2.4: EOSC Service Catalogue Analysis
3. D 3.1: EOSC Portal Functional and Non-Functional Specifications
4. D 3. 2 : EOSC Portal Open APIs Specifications of Service and Resources Providers.

MAIN OUTCOMES RELEVANT FOR THIS DELIVERABLE AND TRIPLE PROJECT

The deliverables can be related to the Architecture working group and FAIR. EOSC Enhance will facilitate interoperability and discoverability of EOSC Resources across scientific disciplines through EOSC onboarding process, EOSC quality assurance APIs and other specifications and guidelines. Also it facilitates FAIR by consolidating the Portal as the point of access for researchers through an improved User Interface for facilitation and continuous enhancement of findability and discoverability.

Past projects include:

1.9 EOSC-Pilot

The EOSCpilot project supported the first phase in the development of the European Open Science Cloud (EOSC) by:

- ☐ proposing and trialing the governance framework for the EOSC and contribute to the development of European open science policy and best practice;
- ☐ developing a number of demonstrators functioning as high-profile pilots that integrate services and infrastructures to show interoperability and its benefits in a number of scientific domains;
- ☐ engaging with a broad range of stakeholders, crossing borders and communities, to build the trust and skills required for adoption of an open approach to scientific research.

LIST OF ANALYSED DELIVERABLES

Three deliverables have been analysed from the EOSC-Pilot project, because they are cited in other relevant deliverables produced by EOSC Hub:

1. D6.9: Final report on Data Interoperability
2. D6.7: Revised Requirements of the Interoperability Testbeds
3. D10.5 Requirements and gap analysis report v1

MAIN OUTCOMES RELEVANT FOR THIS DELIVERABLE AND TRIPLE PROJECT

The deliverables can be related to the Architecture working group. In terms of relevance for the interoperability topic, the most important document has been considered the *D6.9 report on data interoperability*, where the main standard reported is EDMI (EOSC Datasets Minimum Information), as metadata guideline.

1.10 PARTHENOS

PARTHENOS aimed at strengthening the cohesion of research in the broad sector of Linguistic Studies, Humanities, Cultural Heritage, History, Archaeology and related fields through a thematic cluster of European Research Infrastructures, integrating initiatives, e-Infrastructures and other world-class infrastructures, and building bridges between different, although tightly interrelated fields. In order to achieve this objective, PARTHENOS focused on the definition and support of common standards, the coordination of joint activities, the harmonization of policy definition and implementation, and the development of pooled services and of shared solutions to the same problems.

LIST OF ANALYSED DELIVERABLES

1. D4.4 Report on Standardization
2. D5.1 Report on Common Semantic Framework
3. D5.6 Report on Mappings (Final)
4. D5.7 Report on integration of Reference Resources
5. D5.8 Common Design Requirements
6. D6.2 Report on services and tools

MAIN OUTCOMES RELEVANT FOR THIS DELIVERABLE AND TRIPLE PROJECT

The deliverables can be related to the Landscape working group. In terms of relevance for the interoperability topic, the most important document is the *D6.2 report on services and tools* released in April 2017, where the main standards reported are the X3ML toolkit, D-NET toolkit and Metadata cleaner.

1.11 DARIAH-HAS

The HAS - Humanities at Scale project furthered the DARIAH ERIC's aim to integrate digitally enabled research in the arts and humanities in Europe and beyond and to operate a platform to enable trans-national arts and humanities research. The project helped DARIAH sustain existing knowledge in digital arts and humanities in Europe and enable new one. DARIAH connects various hubs of excellence in the domain, and helps them share their results and innovations. By sharing knowledge, DARIAH works proactively to enhance the reach of digital arts and humanities within the European Research Area (ERA). This proposals aimed to address some critical limitations of the current model of sharing knowledge in DARIAH and of connecting the national services in digital arts and humanities initiatives in Europe

1. DARIAH Report on researchers' service needs

MAIN OUTCOMES RELEVANT FOR THIS DELIVERABLE AND TRIPLE PROJECT

The report on researchers' needs analysed the results of a survey conducted by the Humanities at Scale project to determine the key basic services needed by researchers in the field of Digital Humanities and those affiliated with DARIAH and its community in particular. The document has been published in 2016, therefore certain parts of the report became somewhat obsolete but the generic conclusions are still relevant. This study has no direct relevance to the EOSC but was an important blueprint study shaping the preparation of the SSH EOSC cluster project, the SSHOC project.

Tools or services for data management in a content management system (such as data repositories) were the most highly rated tools and services for collaboration. People mentioned institutional and national repositories as well as wiki systems and Gitlab/ Github. The research data management system EASY provided by DANS-KNAW was brought up as well as the web-publishing platform for libraries, museums, archives, and scholarly collections and exhibitions - OMEKA. The META-NET network, the research infrastructure CLARIN, MS Access Database and the website "ArcheoData" were mentioned as well. Most often respondents referred to Dropbox, Google Drive, Typo3 and Zotero as currently used tools and services for data management (in descending order of the frequency of mentions). The diversity of replies indicated that the respondents have very different needs and requirements with regard to data management and have very heterogeneous ideas of the performance. The spectrum from Zotero as a data management tool to the dedicated data management system EASY and the whole research infrastructure CLARIN shows the variety. (...) Interestingly, people have little ideas on what tool to use in the future, even if they do not seem to be satisfied with the current situation. (p. 12-13)

Repositories already exist and are actively used by researchers, but they need to become more user friendly and intuitive. A focus should be on services for the citability of data. The awareness of existing services needs to be extended and training programs should be offered on how to use the services. The need for research specific services like optical character recognition is relatively low. Regarding data manipulation tools, they are available in several forms of maturity, quality and sustainability and therefore no special need has been articulated. (p.18)

The last question block about software development revealed that the hosting of source code is by far the most needed service. It is followed by issue tracking and services for authentication and authorisation. (p. 19)

As a conclusion, it becomes clear from the survey that a lot of commercial tools from the private sector are frequently used in DH research projects. This situation does not only apply to the software development question but to all question blocks. It is irrespective of the existence of open and freely available tools and services or a real lack of them. People tend to prefer using commercial products like Google Docs, GitHub or Microsoft applications, even if tools like Etherpad exist. Due to its limited functionality in comparison with Google Docs it is not sufficient. The need for improvement and further development of Etherpad for example (with regard to formatting, versioning, general user friendliness) is one result of the survey.