Project Name	**FREYA**
Project Title	**Connected Open Identifiers for Discovery, Access and Use of Research Resources**
EC Grant Agreement No	**777523**

D6.1 First Annual Report on PID Commons and Sustainability

Deliverable type	Report
Dissemination level	Public
Due date	30 November 2018
Authors	Simon Lambert (STFC) with contributions from other partners
Abstract	The first report on sustainability looks at the “what”, “who” and “how”, discusses the background and proposes a plan for the next cycle leading to emergence of the PID Commons.
Status	Submitted to EC 27 May 2019

The FREYA project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 777523.

FREYA project summary

The FREYA project iteratively extends a robust environment for Persistent Identifiers (PIDs) into a core component of European and global research e-infrastructures. The resulting FREYA services will cover a wide range of resources in the research and innovation landscape and enhance the links between them so that they can be exploited in many disciplines and research processes. This will provide an essential building block of the European Open Science Cloud (EOSC). Moreover, the FREYA project will establish an open, sustainable, and trusted framework for collaborative self-governance of PIDs and services built on them.

The vision of FREYA is built on three key ideas: the **PID Graph**, **PID Forum** and **PID Commons**. The PID Graph connects and integrates PID systems to create an information map of relationships across PIDs that provides a basis for new services. The PID Forum is a stakeholder community, whose members collectively oversee the development and deployment of new PID types; it will be strongly linked to the Research Data Alliance (RDA). The sustainability of the PID infrastructure resulting from FREYA beyond the lifetime of the project itself is the concern of the PID Commons, defining the roles, responsibilities and structures for good self-governance based on consensual decision-making.

The FREYA project builds on the success of the preceding THOR project and involves twelve partner organisations from across the globe, representing PID infrastructure providers and developers, users of PIDs in a wide range of research fields, and publishers.

For more information, visit www.project-freya.eu or email info@project-freya.eu.

Executive summary

The first report on FREYA's sustainability takes a wide view of the meaning and requirements on sustainability that the context of the project imply. The "what", "who" and "how" of sustainability are analysed in turn. Facets of sustainability are identified, and the EOSC understanding of its components is discussed. The role of stakeholders is clarified, and the use of background context is discussed (whether the evolving forms of the EOSC, or what can be learnt from previous endeavours and interpretations of sustainability). A concrete plan is put forward for the next cycle of work, which will lead to emergence of the structures and mechanisms of sustainability and their relation to the PID Commons.

Contents

1 Introduction

The FREYA project is committed to sustainability, not only of its own outputs such as services, but of the PID infrastructure as a whole, and has a Work Package dedicated to that goal: "WP6 will ensure that the FREYA PID infrastructure is established as a foundation service within the European Open Science Cloud, accessible and usable across Europe and globally, assuring the sustainability and open participative nature of the infrastructure."

The sustainability envisaged by FREYA is conceived at several levels: the services developed within the project by the partners, to be taken forward by those partners; other outputs such as recommendations and standards, underpinning the services but able to form part of the fabric of the EOSC as it relates to persistent identifiers; and high-level outputs, not arising from partners' own developments but nonetheless vital to building the EOSC, such as policy recommendations. These require different approaches to sustainability.

The vehicle for sustainability is the PID Commons. The name implies something about its nature: the word commons suggests a wide range of stakeholders who acknowledge a common interest in the infrastructure and some mutual responsibility for its sustainability. The PID Commons is an ideal, which could be summarised as "open collaborative governance structures with empowered stakeholders". The way to approach the Commons is to take it first as a part of the context in which FREYA's sustainability planning takes place, by virtue of representing ideals of common interest and mutual dependency; but then the structures and mechanisms that implement the Commons will be an outcome of that planning.

FREYA's sustainability work will take place in cycles of engagement, review, analysis and recommendations. There is plenty of background material on which to build: the outputs of THOR, experience of the service providers in FREYA, connections with EOSC and its projects. In this first cycle the aim has been to understand the environment and its implications for sustainability planning. The deliverable considers the "what", "who" and "how" of sustainability and proposes a concrete approach for the next cycle. The aim is to develop the structures and mechanisms that will be integral to the EOSC (and wider stakeholders) so as to establish PIDs as a foundation of the e-infrastructure, not only through existing services and providers but open to others and capable of expansion to encompass new PID types and entities. The EOSC was launched in November 2018, but as Juan Bicarregui, member of the EOSC Executive Board, describes it, it is like a ship after its launch: floating in the water but needing more work before it can make its maiden voyage. FREYA's sustainability effort is part of that work.

2 What to sustain

In the most general terms, sustainability means the capacity to keep some operation or activity continuing into the future. In the context of (e-)infrastructures, the word is often used without definition, and that is perfectly reasonable since the standard meaning of the English word is intended. The EOSC Declaration of 2017[1] uses the word (along with the adjective "sustainable") repeatedly without feeling the need to define it. Nonetheless there is a propensity to restrict its meaning to *financial sustainability*: an assurance that money will be available from some source to allow the ongoing costs of the operation to be met. The influential essay "Principles for Open Scholarly Infrastructures" by Bilder, Lin and Neylon[2] identifies sustainability as one of three areas for establishing trust in an infrastructure, but defines it parenthetically as "funding [the infrastructure]", and allies it with "running the infrastructure (governance)" and "preserving community ownership of it (insurance)".

Undoubtedly financial sustainability is of central importance in enabling the continuance of an operation. Without needing to be specific about what exactly must be done to assure continuance, the matching of income to expenditure allows for staff costs, capital purchases and other types of expenses that might be necessary. However, there might be failures other than of profitability or cash flow that cause cessation of the operation. Usage might decline even if the money is there, due to shifting patterns of behaviour or being superseded by another product. Many social network platforms lost out to Facebook, and not (at least initially) for financial reasons; it would be difficult to call them "sustainable". As Magchiel Bijsterbosch of SURFsara, speaking at the EOSC-hub Week in Prague in April 2019 put it, "Sustainability is more than just the money."

FREYA is committed to "assuring the sustainability of the [PID-related] infrastructure" (Description of the Action, WP6), a goal which is very necessary since "PID infrastructure is established as a foundation service within the European Open Science Cloud", and uncertainty over the future prospects of a foundation service is surely to be avoided. It is therefore the duty of FREYA to consider all aspects of sustainability.

FREYA is concerned specifically with persistent identifiers, and something that is immediately noticeable and worthy of exploration is the relationship of *persistence* to *sustainability*. These words may be used almost synonymously: the THOR Business Plan asks, "What does it really mean for an organisation or service to be persistent?" However, particular requirements are placed on sustainability in the PID arena by the very fact that the identifiers must be persistent.[3] This is also true of long-term archiving repositories or services: a long timescale is fundamental to the job that they do, and is inextricably linked with sustainability of the operation that does the job. It is not however true of all operations that aim to be sustainable. As an example, a service intended for researchers to upload, share and synchronise data has no such long-term ramifications. A researcher making use of such a service would doubtless be annoyed (and very unlucky) to find that it had ceased operation immediately after uploading some data, but it could easily be substituted by another service with similar functionality. With persistent identifiers, a whole complex of access and interconnectedness would be in jeopardy: what if it was no longer possible to assign DOIs, or resolve ORCIDs? A huge investment built up over years would be lost, and could not be retrieved simply by a new provider arriving with a similar offering.

There are many ways of discussing sustainability in particular contexts. As already mentioned the Bilder, Lin and Neylon essay, considering "open scholarly infrastructures", identifies the key attribute as *trust*, and asserts that trust must run strongly across each of the following areas: running the infrastructure (governance), funding it (sustainability), and preserving community ownership of it (insurance). This is certainly one way of approaching sustainability, but it already concentrates on *what must be done*. Before

[1] https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf
[2] https://doi.org/10.6084/m9.figshare.1314859
[3] The report "Persistent Identifiers: Consolidated Assertions" by the Group of European Data Experts discusses different senses of "persistence" in its Appendix 4 (https://zenodo.org/record/1116189#.WvL97pdG1hE).

taking that step, it makes sense to ask what are the facets of sustainability: those qualities that together define what we mean by the word in the context of (e-)infrastructures.

It is proposed that sustainability has three facets[4]:

- **Maintainability.** The capacity of the infrastructure to continue to operate "as is": running its services, maintaining capacity, dealing with problems.
- **Adaptability.** The capacity of the infrastructure to respond to new opportunities, requirements and challenges, which might arise either on the demand side or the technology side.
- **Desirability.** The capacity of the infrastructure to attract and retain users (in a broad sense, including suppliers of third-party services).

These three facets have not been taken out of thin air. It is clear that the three areas of Bilder–Lin–Neylon do map on to the facets, though not one-to-one: insurance, for example, is an aspect of desirability (by giving trust to users in the infrastructure) but also of maintainability (contingency planning for a particular organisation's cessation of activities). A further correspondence can be made with the breakdown of "features of good PID systems" developed at a high-level workshop held in Singapore in August 2018[5]: the features were divided into G*overnance*, *Technology* and *Operation of service*, which roughly match the three facets, with the understanding that governance is a mechanism of desirability. The THOR Business Plan seems to recognise the point with its "revised assumption" that "sustainability does not necessarily require the existence of a formalised organisation."

The point is that any approach to achieving sustainability should recognise all three of these facets: all are necessary for an assertion of sustainability.

There are other attributes that have been predicated of (e)infrastructures, and indeed some collocations are so often repeated that they come to seem intrinsic properties: trustworthiness and openness being the most prominent. The ODIN project, predecessor of FREYA, defined five characteristics for "trusted identifiers"[6] (implying of course that trustedness itself is on a higher plane altogether, a *sine qua non* of an identifier system):

- unique (on a global scale)
- persistent (resolving as HTTPS URIs persistently with support for content negotiation)
- descriptive (having searchable metadata that describe their most relevant properties)
- interoperable with other identifiers
- governed through an organisation with a sustainable business model, with member organisations and using open technologies

It is generally not clear how these desirable properties should be regarded. Are they sustainability issues at all? If so, are they facets of sustainability in the sense defined above? Are they measures of whether sustainability is being achieved, or assessments that give confidence that sustainability will be achieved?[7] Or are they in fact good-in-themselves attributes with an almost ethical dimension in the world of Open Science?

[4] The FREYA proposal had a different breakdown, with governance replacing desirability. The thinking that has gone into this document suggests that "governance" is the wrong level: it is not a *facet* of sustainability but a mechanism for achieving it.

[5] There is a report on another workshop in this series at https://www.project-freya.eu/en/deliverables/pid-strategy-workshop-nov-2018-summary-websiteversion.pdf

[6] Depicted at https://odinproject2012.files.wordpress.com/2014/05/odin_trusted.pdf

[7] In digital preservation, it is necessary to have some method to "test the basic claim that someone is preserving some digitally encoded information; without such a test this is a meaningless claim." (D. Giaretta, *Advanced Digital Preservation*, Springer, 2011). The same remark could be made of sustainability of infrastructures.

It is also possible to frame the sustainability of (e)infrastructures in terms of desirable outcomes. An international workshop held in Singapore in August 2018[8] identified some key challenges for the international PID community:

- building trust in an environment where PID systems continue to proliferate;
- embedding the use of PIDs in research workflows and harnessing the power of connecting PIDs;
- increasing the adoption of PIDs by researchers and research organisations.

These express what we want to happen: not sustainability itself, not desirable attributes *per se*, but desirable outcomes, presumably underpinned by a sustainable infrastructure.

Rather than expend a great effort on trying to classify these various attributes and aspirations into a all-encompassing framework for thinking about sustainability, we will regard them as considerations to keep in mind in designing and implementing sustainability mechanisms, but not for driving them.

What, though, are the things that must be sustained? Up to now in this document, in the context of (e-)infrastructures, the terms "operation" and "activity" have been used. Is it possible to specify them further, perhaps in terms of "services"? According to the definition of the FitSM[9], a service is a "Way to provide value to customers through bringing about results that they want to achieve". This seems a very general starting point.

Certainly the idea of services is central to the vision of the EOSC. At the highest level it is envisioned as an environment, within which activities take place: according to Commissioner Carlos Moedas, "a trusted environment for sharing and analysing data from all publicly funded research." It is at the next level that services become prominent: "EOSC aims to provide members of Europe's research community with 'a virtual environment with free at the point of use, open and seamless services for storage, management, analysis and reuse of research data, across borders and scientific disciplines'"[10]. The EOSC Glossary[11] pins things down even more: an EOSC Service is

> "*an **EOSC Resource** implemented by the **EOSC System** to provide **EOSC System Users** with ready-to-use facilities. EOSC Services are supplied by an **EOSC Service Provider** in accordance with the **EOSC Rules of Participation** for **EOSC Service Providers**. EOSC Services are approved by the **EOSC Service Portfolio Management Committee** and populate the **EOSC Service Portfolio** and the **EOSC Service Catalogue**."*

where a resource is defined as:

> *any asset made available (by means of the **EOSC system** and according to the **EOSC Rules of Participation**) to **EOSC System Users** to perform a process useful to deliver value in the context of the **EOSC**. EOSC Resources include services, datasets, software, support, training, consultancy or any other asset.*

Though this might seem rather convoluted, it defines some roles and some constraints and highlights the importance of the EOSC Service Catalogue. Ensuring compliance with the EOSC Rules of Participation and presence in the EOSC Service Catalogue are surely aspects of maintainability.

[8] https://www.project-freya.eu/en/deliverables/pid-strategy-workshop-nov-2018-summary-websiteversion.pdf summarises a second workshop in the series.
[9] https://fitsm.itemo.org/wp-content/uploads/sites/3/2018/05/FitSM-0_Overview_and_vocabulary.pdf
[10] https://publications.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1
[11] https://eoscpilot.eu/eosc-glossary

The EOSCpilot project, in its work on the EOSC service architecture, has identified 47 classes of services that will be needed for the EOSC. Within the broad division between *core services* (now called *cross-cutting services*) and *services directly accessible by researchers*, two different groupings of the 47 are presented: one from the user perspective and one from the functional perspective. The cross-cutting services include *PID Handler*, a class of services "to generate and resolve Persistent Identifiers (PIDs)". Although this is undoubtedly the area of interest of FREYA, FREYA's scope goes much beyond simply generating and resolving PIDs. Identifying where else in the service architecture FREYA's outputs should be present will be an important task in establishing sustainability.

Let us now turn things round and ask what FREYA is producing that will enable that vision of the EOSC as "a trusted environment for sharing and analysing data from all publicly funded research"? At the highest level, the vocabulary is imprecise: the role of FREYA is variously described as a "component", "building block", "foundation service", and "framework for collaborative self-governance of PIDs and services". However, the project's milestones correspond in many cases to outputs, and the project partners are relating these to the interaction needed with stakeholders and asking what needs to be sustained. Table 1 lists the milestones per Work Package, inasmuch as they correspond to potentially sustainable outputs, with some additional entries not listed as formal milestones but obviously representing outputs (such as training materials).

Work package	Milestones corresponding to potentially sustainable outputs
WP2 PID Core Services	PID resolution services PID metadata services PID services registry Common DOI search
WP3 New PID Types	New PID evaluation New PID prototypes
WP4 Integrating the PID Graph	Mature service demonstrators New PID types in service demonstrators Integration of PID Graph with EOSC
WP5 Iterative Engagement	PID Forum functional Training materials

Table 1 FREYA milestones and other potentially exploitable outputs

It is striking that few of these are really services in the sense of FitSM and EOSC. Training is a type of EOSC Resource (it is given as an example of a Resource, in fact). Metadata standards, guidelines and recommendations are not set in stone: they will need to be updated and extended, and they must themselves be sustainable (maintainability, adaptability and desirability are all required). Yet they can

hardly be seen as services in themselves. Nor can the mechanisms and governance proposals for introduction of new PID types: it would be very strange to see this appearing in the EOSC Service Catalogue. It does not seem that it could even be considered as an EOSC Resource—it is something more fundamental to the functioning of the EOSC itself, rather than an "asset" of the EOSC.

The conclusion is that the outputs of FREYA (and of other projects that are helping to build the EOSC) need careful analysis to understand what role they play in the infrastructure, and hence what their needs for sustainability are.

Key points emerging:

- All aspects of sustainability should be considered, not only financial viability.
- Services are fundamental to the EOSC, but are not all that FREYA will offer to the EOSC.
- The outputs of FREYA (and other EOSC-building projects) must be positioned within the ontology of EOSC components based on the EOSC Glossary and within the EOSC Service Architecture.
- The fundamental vision of FREYA, the PID Graph, must be related to components of the EOSC, through understanding how it is made up of services, APIs, graphs for particular applications, ….

3 Roles in sustainability

Sustainability involves people: it is people who are responsible for running an operation and ensuring its capacity to continue running in the future. People strive for and judge the success of maintainability, adaptability and desirability. Therefore planning for sustainability must take account of the various roles that people play. Of course it is generally impossible to engage with every individual, present and future, who might play a role, and so the way in which people are grouped becomes important.

In the context of (e)infrastructures, people are generally employed by organisations (universities, funding agencies, service providers, …). However grouping people according to their employing organisation is not necessarily the best way of assessing their roles in sustainability. It is worth examining typical approaches to identifying roles in sustainability to understand their assumptions and constraints.

The word "community" is very frequently used in the context of (e)infrastructures, to the extent that, like "trustworthiness" and "openness", it threatens to lose its significance. A cursory glance at how the word is used shows that there are several quite different concepts being represented. Sometimes it means area of research ("discipline", "field"), especially when used in the plural—this is surely the meaning intended by EUDAT in its statement "Any research community or infrastructure may integrate & deploy EUDAT services". The emphasis is entirely on the *users* of a particular infrastructure. Researchers in a particular field presumably have some common understanding and practices even if there might be divisions and disputes within the field, and it may be that some very specific services are used by everyone who is active in the field: arXiv for pre-prints in high-energy physics, or the Cambridge Crystallographic Data Centre ("The world's repository for small-molecule crystal structures") being examples. These services might include use of particular PIDs, especially if mandated by publishers or funders or just "the way things are done". Indeed some identifiers are created for and used by specific communities (in this sense), such as IGSN (geology) and Accession Numbers (bioinformatics).

It is however a leap of faith to assume—as is often done—that there is or will be a community "consensus" when it comes to advanced information technologies. The further the discipline is from using such technologies as a matter of course, the less likely it is that any consensus exists. It is not necessarily true that awareness of and commitment to Open Science is shared. There will tend to be a few enthusiasts ("early adopters" perhaps) who see themselves as leading the way, but others may hold a range of beliefs: that the new technology has exciting but unclear potential, or is an annoying imposition, or simply something that is being imposed and cannot be questioned.

It might be hoped that the idea of community can be given a more rigorous formulation through "communities of practice", a concept that has been much studied. However, here the community is seen to arise spontaneously within a given organisational environment, rather than something to which one simply belongs[12].

A further and even more severe problem with attempting to encompass sustainability within ideas of research communities is that it is exclusionary. A community almost always refers to a grouping external to the service, typically end-users, and so certainly not including the service providers themselves. Furthermore, the word tends not to be utilised when referring to business actors, though these might be users of infrastructures. It would seem very strange to refer to the "academic publishing community", perhaps because of the element of commercial competition between publishers; "business sector" would be more natural, while STM ("The global voice of scholarly publishing") refers to its ambit as the "scholarly publishing industry".

[12] "Sometimes co-workers who have complementary knowledge will form a group. Often called 'communities of practice,' these self-organized groups are generally initiated by employees who communicate with one another because they share common work practices, interests, or aims." (T.H. Davenport & L. Prusak, *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, 1998, p.38)

Attempts to generalise the idea of community often end in using the word in the singular, which predictably makes the intended meaning so fuzzy as to be almost useless: it becomes a synonym for "anyone who takes an interest". The THOR Business Plan makes a brave attempt: "we define a community as a group of people who are engaged with and have a stake in some activity or enterprise. The community will typically reflect some shared norms and values. In the case of PID service providers, the members of a community may also represent the service provider's membership."[13] However the same document goes on to refer to the "research community at large".

The EOSCpilot glossary already cited tries to characterise the roles of involved parties without using the idea of community. An *EOSC System User* is defined as "the role played by every actor (human or machine) exploiting the EOSC System according to the EOSC Policy. An EOSC System User might be further specialised in roles including EOSC End-users, EOSC Suppliers, and EOSC System Managers." An EOSC End-user is "an EOSC System User consuming EOSC Resource(s) by means of EOSC Service(s) (e.g. the EOSC Portal) to accomplish a task. She can be a Researcher, a Research Administrator, or a Third-party Service Provider." This makes explicit that users may be suppliers as well as consumers of resources.

A second approach to characterising the roles in sustainability is through *stakeholders*, those who have some stake or interest in the (e)infrastructure or the services it provides. Again, stakeholders are groupings for which it is not self-evidently true that they will have a common view or practice, though the idea does have the strong advantage of inclusivity: service providers and third-party suppliers comfortably fit within the definition of stakeholders, just as much as end-users. However there might well be persons or groups who have an influence on an infrastructure and its sustainability but do not have a stake in it (or are not aware that they do)—in the sense of not caring directly about the fortunes of the infrastructure. High-level policy makers, perhaps, might not take into account the infrastructure for persistent identifiers when formulating their declarations on Open Science, but those declarations will certainly have an impact, whether positive or negative. As long as it is understood that stakeholders might include parties who influence the environment without having a real stake of their own, the term can be used safely.

It is not uncommon to find detailed stakeholder analyses in the reports of projects that aspire to sustainability of their outputs, and FREYA itself has produced one its deliverable D5.2 "Communications and Stakeholder Plan". Five main stakeholder categories are listed:

- service providers;
- research stakeholders;
- users;
- structural stakeholders (e.g. funding agencies or government bodies);
- commercial stakeholders.

These are further classified according to their information needs and main communication approach that will be applied by the FREYA project team, followed by the desired outcomes: what the group is expected to do. For example, the stakeholder groups "service providers" and "research stakeholders", responding to the information need "Does FREYA support my domain?", are expected to produce the outcome "Contribution to PID Graph, possibly feedback on demonstrators". The assumption is that meeting the information needs will be enough to get the group to behave in such a way as to produce the outcomes.

The five-way breakdown refers to major projects building parts of the EOSC (EOSC-hub and OpenAIRE Advance are mentioned by name), but the governance of the EOSC as a whole is not mentioned, since at the time of preparation of the deliverable it had not been confirmed. There is a need to include EOSC governance more explicitly.

[13] There is an intention to distinguish this view of community from the broader target audience of a service—an interesting parallel with the distinction between the Designated Community (for whom data is being preserved) and the wider users made by the OAIS standard in digital preservation.

As validation of this breakdown, the stakeholder types adopted by EOSCpilot (deliverable D2.1) can be compared:

- European e-Infrastructures (FREYA: Service providers)
- Data/Research Initiatives (FREYA: Research stakeholders)
- Cloud providers (FREYA: Service providers/Commercial stakeholders)
- Research funders (FREYA: Structural stakeholders)
- Cloud community (FREYA: Research stakeholders/Users)
- Research Communities and Institutions (FREYA: Research stakeholders/Users)
- Research Infrastructures[14] (FREYA: Research stakeholders)
- Policy makers (FREYA: Structural stakeholders)

D5.2 is a communication plan so focusses on channels and content. FREYA's Work Package 6 needs to build on that and go further: not only to communicate but to engage; what does FREYA need them to do, and why would they want to do that? In FREYA, this is precisely the work of building the PID Commons. But it is possible to make some general statements about different types of stakeholder—what kind of stake they hold and what are the implications of that. Stakeholders are confronted with the three facets of sustainability—maintainability, adaptability and desirability—and will influence or respond to those in different ways:

- *reactive*: they make judgements based on what they perceive of the infrastructure, and modify their behaviour accordingly (typically: end-users);
- *proactive*: they contribute to building and (wittingly or not) to its sustainability (typically: service providers);
- *context-setting*: they take decisions or actions that are outside the scope of the infrastructure but have a bearing on its evolution (typically: policy makers).

Table 2 presents a first cut at how the five broad stakeholder groups of FREYA D5.2 map on to these types.

Stakeholder group	Nature of stake
Service provider	Proactive Context-setting (for service providers with scope wider than PIDs)
Research	Reactive Proactive (in the case of research infrastructures for particular domains)
User	Reactive
Structural	Context-setting
Commercial	Reactive Proactive (when building services of their own)

Table 2 Stakeholders and their stakes

[14] Research infrastructures provide thematic/domain-focussed infrastructures, in contrast with the "horizontal" e-infrastructures. An examples is CESSDA in the social sciences.

Key points emerging

- Great care should be taken in framing sustainability in terms of “communities”.
- Thinking in terms of stakeholders is valuable, bearing in mind that some might not have explicit stakes.
- Stakeholders should be mapped not only to their information needs but to desired behaviours and their own motivations.
- In the FREYA context, there is a need to revisit the stakeholder analysis to take account of ongoing developments in the EOSC and its governance.

4 How to sustain

In principle the path to sustainability is clear: identify what is to be sustained, then select and put in place structures and mechanisms that will ensure the three facets of sustainability for those things, right now and into the future. The mechanisms will have to be implemented somehow, and this is the role of the stakeholders; so the corollary is that the mechanisms must be congruent with the needs, wishes and capabilities of the stakeholders. Of course, the identification of stakeholders makes assumptions about how things will be in future and might tend to perpetuate the status quo; nonetheless, the stakeholder groups do exist, already play roles in the infrastructure, and any mechanisms that do not align with them will be unworkable or at best fragile. The question of how to sustain therefore depends on the choice of mechanisms that are both sufficient to ensure the three facets, and at the same time implementable by the stakeholders, with powerful motivation for them to do so. In the rest of this section we will examine some major issues that arise from this approach.

There are some pitfalls to avoid. One has already been mentioned: a tendency to perpetuate current structures, to assume that current groupings of stakeholders are immutable and that sustainability must be implemented only in terms that they already presume. The working techniques introduced in the next chapter will help to avoid this mindset. Another pitfall might be excessively detailed levels of granularity. When thinking of maintainability of some particular PID services, it would not be right to specify a need for "ticketing system for recording and tracking problems with PID assignment"—this is simply too fine-grained, and a reasonable service provider organisation would be trusted to handle it.

Sustainability of the structures and mechanisms themselves is another danger area. The OAIS standard (see box below) explicitly allows for recursion in the provision of what it calls Representation Information allowing the understanding of an object by its Designated Community: what supplementary information is needed to understand the object, and what in turn does that require if it is to be understood? The spectre of infinite regress is less appealing in sustaining e-infrastructures, though it can certainly be imagined. If for example a particular oversight committee is called for to monitor the introduction of new PID types, how will that body itself be sustained? Who pays for it, how is its membership rejuvenated? We leave this question unanswered for the time being, noting only that it will require thought at some stage. It is likely that general EOSC governance, assumed to be supported at the highest levels, will permit termination of recursion.

Structures and mechanisms are created and live within a context. Planning for sustainability is not a purely theoretical exercise, designing a world of ideal interactions from scratch. In the case of FREYA, by far the most important aspects of the context are the existing PID infrastructure and the European Open Science Cloud. The former constrains the structures through the existence of a highly developed and successful system of service providers, with more or less robust business models and established interactions with their stakeholders and with each other. The latter introduces disruptive elements such as its "rules of participation", whose applicability to well established infrastructure components will require negotiation, but also offers new opportunities.

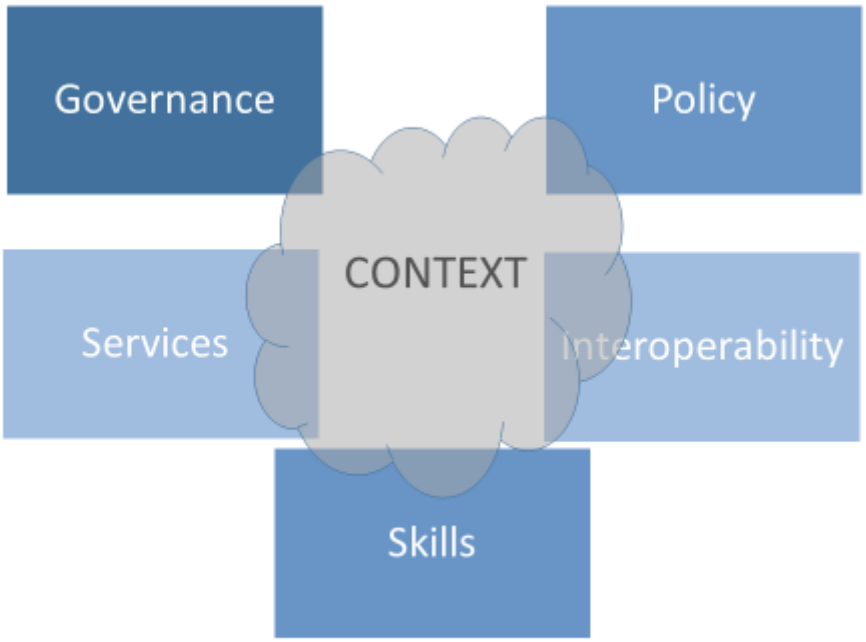

Figure 1 The themes of EOSCpilot

PIDs play an undeniably important role in the EOSC—indeed they are fundamental to the FAIR principles that underlie the EOSC. From the sustainability point of view, PID-related services must conform to the requirements and expectations of the EOSC. The formulation of services in EOSC recognises a distinction between:

- EOSC Services offered by EOSC Service Providers to EOSC System End Users;
- and EOSC Services offered by EOSC Service Providers to enable the EOSC System.

These have different implications for sustainability, the latter surely more stringent inasmuch as the EOSC as a whole depends on them.

In any case, it would be wrong to think that "services" are all that matter. Mechanisms for developing and admitting new PID types to EOSC are a part of its sustainability, as is the onboarding of new service providers—the environment must be open to these. The structures and mechanisms must exist, but (as already noted) these will certainly not appear as "services" in the EOSC Portal. Are they aspects of "governance"? Yes, for they deal with the orderly functioning and evolution of a particular part of the EOSC. How to set up or adapt governance structures for these purposes is precisely the issue that FREYA must grapple with in the next cycles of work on sustainability.

Context means not only the environment within which the PID infrastructure will develop, but also what can be learnt from other initiatives that have tackled similar issues, perhaps from a different perspective. This is the background that influences the perceptions of what is possible or what has been tried in the past, or gives novel views on the problems. A number of reports have been studied as part of FREYA's work—see Annex A and Annex B—and included outputs of EOSCpilot as well as more general papers.

The development of structures and mechanisms for sustainability is the work of the next cycles. However it will not be a great surprise if structures are tightly constrained by what already exists or what is on the horizon: existing PID service providers, existing e-infrastructures with their forms of governance, the

evolving structures of the EOSC. Adapting these groupings of stakeholders to the purposes of sustainability of the PID infrastructure will be essential. However, as concerns mechanisms—what the stakeholders do through the structures in which they participate—it is possible to delineate what their scope will be. Without considering how stakeholders relate to these, several general mechanisms can be identified immediately:

- consultation;
- measurement/monitoring;
- tracking of external environment (technology, policies, …);
- development;
- governance.

Where does this leave the PID Commons? The PID Commons is how FREYA will achieve sustainability of the PID infrastructure, though simply labelling it such says nothing substantive. There are of course hints in the name, and some aspirations, even if not always clearly articulated, in the project's Description of the Action. There the PID Commons is more an assemblage of things needed for sustainability—defining the Commons by what it does rather than what it is:

> "defining the roles, responsibilities and structures for good self-governance based on consensual decision-making."

and also:

> "The main outputs of work package 6 will be: an established PID Commons, with a committed group of relevant stakeholders, a governance structure and a future activity roadmap; an approach to sustainability for FREYA outputs and services; an evaluation of the results of FREYA, with recommendations for future activities."

It would be possible to take an emergent, bottom-up view of the PID Commons, as just what emerges from the detailed planning that is outlined in this deliverable: the thinking about what needs to be sustained, stakeholders and their roles, structures and mechanisms that bring them together. But that would surely miss something of the intent of the Commons, captured in its name, representing a body of shared resources in which stakeholders have a common interest and common commitment. The way to handle this intent is probably to treat it as another form of context, influencing the planning for sustainability as the EOSC environment does.

Figure 2 attempts to depict the general model that has emerged from all the foregoing discussion. It is a simplified view, but does make clear the major factors that the planning in the next cycle must take into account: making sure that the "what to sustain" is necessary and sufficient to support the e-infrastructure; correctly analysing the context and understanding its implications for structures and mechanisms; the same for stakeholders.

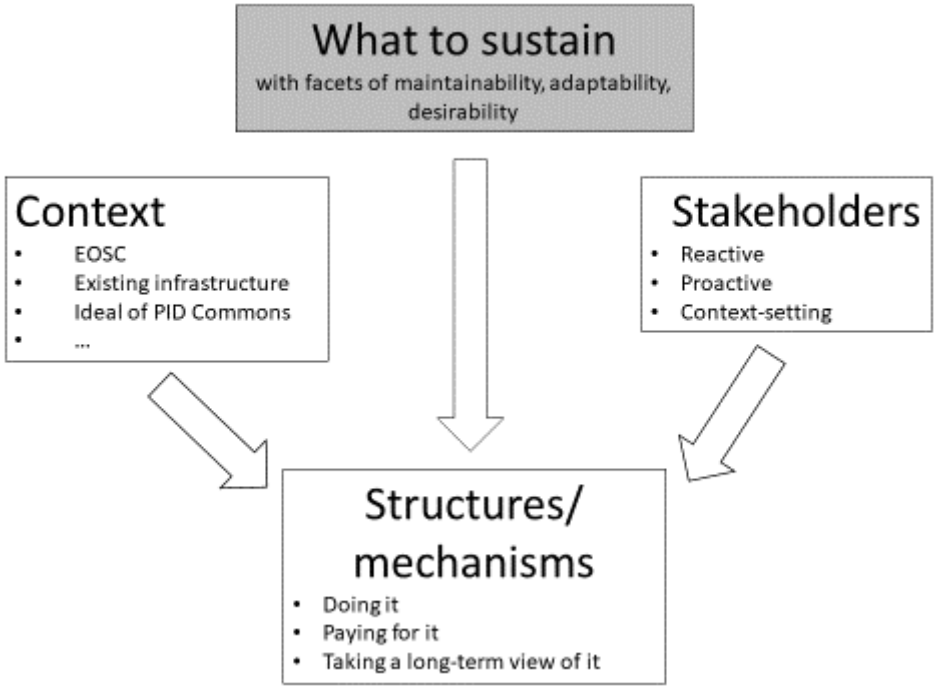

Figure 2 General model of sustainability planning

A final point to consider is the relationship between FREYA's efforts towards sustainability and the exploitation of its results by the FREYA partners. This exploitation is in the classic sense of European-funded projects since time immemorial, giving an assurance to the funder that their investment is being put to good use, that it will not simply support some staff working on a project with no lasting impact but will lead to worthwhile results that are taken forward. It has to be admitted that there is a certain tension, in projects such as FREYA that are building the wider e-infrastructure, between this view of exploitation and the wider striving to sustainability. It is not that they are incompatible, but that the priorities are different: exploitation, expressed through business plans, focusses on what the partners themselves will do with the results; e-infrastructure sustainability entails a more selfless attitude, allowing that other actors will have roles to play that might even compete with the FREYA partners, but that creating the conditions for that is for the greater good. Not incompatible, for sure: the fact that FREYA partners can exploit the project outputs gives evidence that those outputs are sound and can be deployed in reality. The FREYA DoA puts it well when it distinguishes four types of exploitation by project partners:

- core PID providers (Crossref, DataCite, ORCID) are dependent on access to PID developments to enhance their current services;
- e-infrastructure operators (STFC, CERN, EBI/EMBL) embed the work of the PID providers into the fabric of European Open Science;
- domain specialists whose knowledge enriches the Open Science infrastructure (BL, CERN, DANS, EBI/EMBL, PANGAEA);
- researchers who require ease of access to that knowledge, and the publishers, libraries and other institutions where it is kept (BL, ANDS, PLOS, Hindawi).

Key points emerging:

- The design of structures and mechanisms for sustainability must avoid pitfalls on several sides.
- The most important aspects of the context are the existing PID infrastructure and the European Open Science Cloud.

- The PID Commons may be seen as providing a context with its ideals of common interest and mutual support of stakeholders.
- The relationship between sustainability and exploitation of project results is delicate.

5 Moving forward

The preceding chapter has explored the "how" of sustainability: the structures, mechanisms and constraints are available or must be taken into account. This is not "how" in the practical sense of what the members of the FREYA project will now do develop the sustainability planning. A parallel may be drawn with building a house: the "what" tells us what a house is for and what attributes it must have; the "how" describes the use of bricks, mortar, timber and tiles, the importance of foundations and the function of roofs, doors and windows; but there is still a need, when confronted with the desire for a particular house, to have methods for planning how to use those materials and those substructures to achieve the intent of *that particular house*.

It could be argued that the collective expertise and experience of the FREYA partners is sufficient; that putting our heads together will produce a good solution, since it will represent and balance a wide range of domains and of stakeholders in the PID landscape. Of course this is one of the strengths of the FREYA consortium, but even the best and most comprehensive panel of wise heads may suffer from groupthink, or assume that the future will be and must be an extrapolation of the present. Any group undertaking a planning exercise is grateful for tools that will help to structure and direct its thinking.

Two possibilities areas likely to be especially valuable for the next stages of FREYA's planning for sustainability of the PID infrastructure: *backcasting* and *scenario planning*.

INTRODUCTION TO BACKCASTING

Backcasting is a planning technique that has been widely applied to complex systems such as transport, energy and environmental sustainability. It is explicitly normative in that it starts by defining a desirable future state and works backwards to connect the present to that future. Backcasting is regarded as "particularly helpful when problems at hand are complex and when present trends are part of the problems."[15] It helps to overcome adherence to incremental or marginal change, but does assume that the time horizon is long enough to allow scope for deliberate choice[16].

Backcasting is not only of academic interest. Similar techniques are used in business planning, one going under the name "Remember the Future"[17]. In this case the focus is on product development, asking "What will the product have done for me at such-and-such a point in the future?" This type of question enables more fruitful thinking than simply "What should eht product do?" A related technique has actually been employed in a European-funded (Framework 7) project for sustainability planning: the photograph below shows the result of a collective working session with the timeline running left to right (future at the right) and different aspects of the system represented in different colours. The writing on the Post-Its, generated by workshop participants, describes what must be in place at that particular time to lead to the future state envisaged.

[15] J. Holmberg & K.-H. Robert, "Backcasting — a framework for strategic planning", *International Journal of Sustainable Development & World Ecology*, Vol 7 Issue 4 (2000).
[16] Karl H. Dreborg, "Essence of backcasting", *Futures*, Vol 28 No 9 !1996).
[17] https://www.innovationgames.com/remember-the-future/

Backcasting is considered appropriate for FREYA because the PID infrastructure being developed by FREYA is embedded in the European Open Science Cloud, and that is a collective endeavour to create something that does not yet exist and is considered desirable—so the striving towards the future state fits perfectly with the essence of backcasting.

INTRODUCTION TO SCENARIO PLANNING

Scenario planning, like backcasting, is not concerned with predicting what is going to happen (forecasting). Scenario planning is perhaps more objective and more rational than either forecasting or backcasting: it entails studying the possibilities of tomorrow, recognising that there is uncertainty and unpredictability, and developing plans that are robust in those alternate futures. It does not assume stability and predictability, but recognises that "today's world is better characterised by turbulence, uncertainty, novelty and ambiguity - conditions that contribute disruptive changes and trigger the search for new ways of coping."[18]

One of the major challenges of scenario planning is the development of the scenarios themselves. Several approaches are possible, including consulting a range of experts and synthesizing their opinions; "morphological approaches" in which different future states for all the key driving forces are combined, then reduced to manageable numbers; and "cross-impact" approaches[19].

A variant of scenario planning is the so-called Delphi technique, based on asking experts in their various fields to estimate individually the probability that certain events will occur in the future. The goal is to converge on future views by comparing their answers with those of the other experts.

Scenario planning may be helpful for FREYA because of its requirement for looking at a range of futures, not assuming that things will necessarily develop as they have in the past and that the same factors will influence the developments. Its focus on "stories" about the future helps to break down mental blocks and encourage new ways of thinking.[20]

[18] R. Ramírez &A. Wilkinson, *Strategic Reframing: The Oxford Scenario Planning Approach*, Oxford UP, 2016.
[19] Gill Ringland, *Scenario Planning*, Wiley, 1998.
[20] A possible starting point for scenarios in the context of FREYA, and indeed the EOSC in general, might be the future of "openness" or of "FAIR". Do these trends gain an unstoppable momentum that sweeps all stakeholders along, or will there be blockers in some areas? What then would be the implications for PID services and policies? A range of scenarios immediately springs to life.

Whatever techniques are adopted, there will be two special emphases of the work in the second cycle:

- the PID Commons;
- the relationship with EOSC at large.

According to the plan for WP6:

> "In the second cycle, the lineaments of the PID Commons will emerge as feedback is obtained from stakeholders. Roles and responsibilities will be outlined and related together; necessary and sufficient conditions for the functioning of the Commons will be drawn out."

It is all very well to make plans within the confines of the FREYA project consortium, but these plans will entail engagement of other stakeholders, and during the lifetime of the project is when that engagement must start. The "iterative engagement" work of the project will shift gear, not only creating awareness of the progress and achievements of FREYA, but drawing the stakeholders into the framework of sustainability through the Commons.

Second, the relationship of FREYA's PID infrastructure to the EOSC at large will be explored and clarified. Work has already started through establishing contacts and preliminary discussions with the major EOSC-building projects EOSC-hub and OpenAIRE Advance. Alignment towards the common goals of the EOSC is important, and a number of areas have been identified, including:

- identification, development and sustainability of core services for the EOSC;
- the relationship between the FREYA PID Graph and other PID-related services and infrastructure in the EOSC, both currently existing and potentially emerging in the future;
- policies and rules of participation impinging on the PID infrastructure.

Finally, the whole process presented in this deliverable can be summed up in Figure 3, setting out what FREYA partners will do in the next cycle:

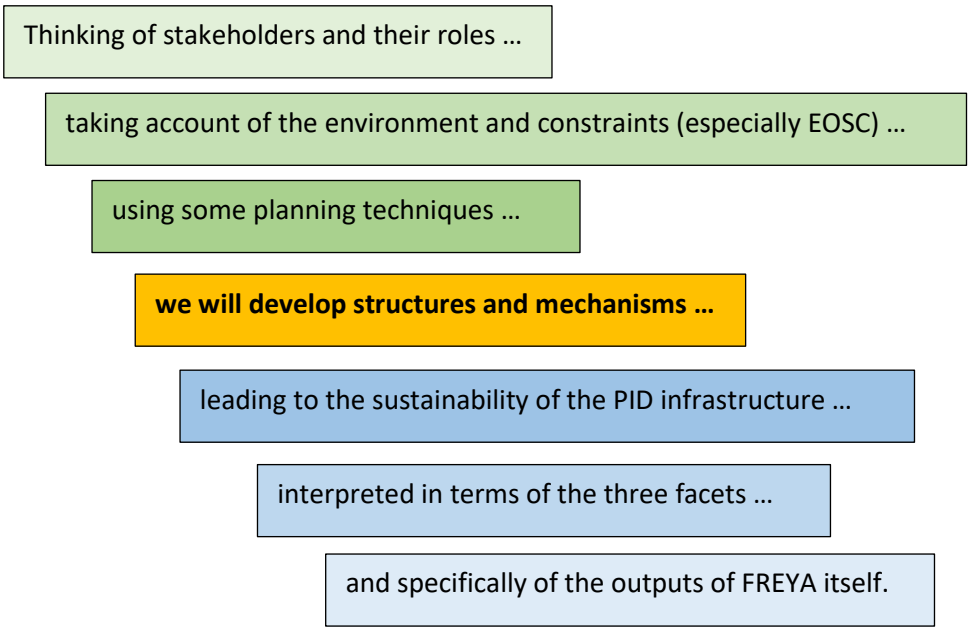

Figure 3 The next cycle of sustainability work

Annex A: Examples of analysis of background material

"Realising the European Science Cloud"
http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

Reviewed by	René van Horik
Date	July 2018
Purpose of the document	First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud (HLEG). Advice on governance and funding of an Open Science Cloud (= an open and trusted environment where research data can be safely stored and made openly available). Trends of open science and their relevance for the EOSC are described. Currently 50% of all research data and experiments is considered not reproducible, and the vast majority of data never makes it to a trusted and sustainable repository. Recommendations are given to realise the EOSC (policy recommendations, governance recommendations, implementation recommendations). The EOSC is technically conceived as an Internet of FAIR data and services. It is the ambition of the HLEG that this report can be the basis for a formal paper to be approved by the Commission and subsequently endorsed by Member States.
Direct relevance to PIDs (if any)	PIDs are mentioned directly and indirectly in the report. Several observations have relations with the application of PIDs. • "decaying hyperlinks" are a problem of the current science system. • To implement the EOSC the HLEG advises an approach which is based on minimal rigorous standards. • The EOSC should distinguish domain specific standards and protocols (e.g. Preferred Persistent Identifiers in a discipline) and protocols for concepts and data formats that are of general utility. • Generic protocols should be the responsibility of international scientific organisations both formal and informal (e.g. ORCID for researcher PIDs) … • Implementation recommendation: actively stimulate and support multiple ESFRI-type communities in the same broad domain to collaborate on these issues and collectively set a minimal set of norms for a Preferred Persistent Identifier (PPID) scheme in their domain as well as mappings to other PID. • Another recommendation: Stimulate cross domain collaboration at ESFRI level for more generic semantic types such as people, organisations and geographical locations.
Position on e-infrastructures	(e-infrastructure: refers to all ICT-related infrastructures supporting ESFRIs or research consortia or individual research groups, regardless of whether they are funded under the CONNECT scheme, nationally or locally) At policy and governance level the EOSC should take an approach similar to the of the successful ESFRI roadmap: preparatory phase followed by an implementation phase. (but we cannot afford a preparatory phase of many years).
Position on governance	Governance recommendations: 1. Aim at the lightest possible, internationally effective governance 2. Guidance only where guidance is due (relates to technical issues, best practices and social change) 3. Define Rules of Engagement for service provision in the EOSC 4. Federate the gems and amplify good practice Implementation recommendation: Develop a governance plan for the EOSC, as lightweight and inclusive as possible.
Position on sustainability	"An important aspect of the EOSC is professional data management and long term data stewardship. The latter aspect is presently lacking". (This is an incentive to pay attention to sustainability - in which PIDs play an important role of course. Recommendation: establish minimal technical standards for the EOSC and plan for their long-term maintenance and compliance..
Other notable points	

"Business plan for sustaining the THOR federated PID infrastructure and services" - final version of THOR D5.4

Reviewed by	Simon Lambert
Date	26 June 2018
Purpose of the document	The deliverable is entitled "Business Plan for Sustaining the THOR Federated PID Infrastructure and Services" and obviously focuses exclusively on THOR. It is titled as a business plan but is really an examination of the idea of sustainability and issues around it, though it does examine the DataCite, ORCID and Crossref set-up and business models using the "business canvas" structure.
Direct relevance to PIDs (if any)	Of course highly relevant! It deals specifically with PID services.
Position on e-infrastructures	The word "infrastructure" is in the title, but the word seems to signify simply the collection of services related to PIDs. DataCite and ORCID are called infrastructure providers. There is no recognition that these services might be part of a larger e-infrastructure, nor any mention of the EOSC.
Position on governance	The word "governance" is always used as "open governance" or "community governance" - these are regarded as desirable, indeed essential for sustainability, but no models are given.
Position on sustainability	Sustainability means the ability to continue into the future. The report acknowledges a shift of understanding from sustainability of particular organisations to sustainability of services. Community engagement/ownership is important, as is a commitment to openness. A set of "open questions" is presented around the sustainability of PID services under the headings **Openness**, **Dependencies**, **Centralisation**, **Persistence** (not of the PIDs but of the organisations/services), **Sustainability of service adopters**; these are examined and some assumptions revised - e.g. sustainability does not necessarily require the existence of a formalised organisation. The report concludes with "four key elements that factor into sustainability of PID services: **Community, Trustworthiness, Persistence, and Dependencies**". Dependencies are particularly interesting: will the whole infrastructure collapse if one service provider fails?
Other notable points	The report reveals a very different scope from FREYA. There is an explicit assumption that PID services developed in THOR will be absorbed into the operations of existing service providers (DataCite and ORCID) at marginal cost.

Annex B: List of background material analysed

Bilder, G., Lin, J., & Neylon, C. (2015). Principles for Open Scholarly Infrastructures https://doi.org/10.6084/m9.figshare.1314859

“Business plan for sustaining the THOR federated PID infrastructure and services” - final version of THOR deliverable D5.4

“Implementation Roadmap for the European Open Science Cloud” http://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf#view=fit&pagemode=none

“Realising the European Science Cloud” http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

“The European Open Science Cloud: Who pays for what?” - report from Science|Business https://sciencebusiness.net/report/european-open-science-cloud-who-pays-what

EOSCpilot D2.5 “Recommendations for a minimal set of Rules of Participation”

EOSCpilot D5.1 “Initial EOSC Service Architecture”

EOSCpilot D5.2 “EOSC Service Portfolio”

EOSCpilot D5.3 “EOSC Federated Service Management Framework”

EOSCpilot D6.2 “EOSC architecture design and validation procedure”

RDA GEDE report “Persistent identifiers: consolidated assertions” https://zenodo.org/record/1116189#.WvL97pdG1hE