Project acronym: CS3MESH4EOSC

Deliverable D2.2: Science Mesh service operational procedures

Contractual delivery date	30-06-2021
Actual delivery date	30-07-2021
Grant Agreement no.	863353
Work Package	WP2
Nature of Deliverable	R (Report)
Dissemination Level	PU (Public)
Lead Partner	SURF
Document ID	CS3MESH4EOSC-21-013
Authors	David Antoš, Milan Daneček, Daniel Müller, Renato Furter, Ron Trompert

Disclaimer:

The document reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 863353

Versioning and Contributions History

Version	*Date*	*Authors*	*Notes*
0.1	13.05.2021	Ron Trompert (SURF), Daniel Műller (WWU), David Antoš (CESNET), Milan Daneček (CESNET), Renato Furter (SWITCH)	Initial draft
0.2	21.05.2021	David Antoš (CESNET)	Added comments
0.3	02.06.2021	Ron Trompert (SURF)	Comments incorporated and finished document
0.4	07.06.2021	David Antoš (CESNET)	Clarifications, textual revision, formatting
0.5	08.06.2021	Daniel Müller (WWU)	Textual revisions
0.6	08.06.2021	Ron Trompert (SURF)	Textual revisions and conclusions added
0.7	10.06-2021	Ron Trompert(SURF)	Minor textual improvements
0.8	30.06.2021	Ron Trompert (SURF)	Comments incorporated

Index

Index of Figures

1 Introduction

The Science Mesh is designed to be a highly distributed platform with a very lightweight and almost fully decentralized infrastructure. It consists of several Independent Sites running their own Enterprise File Sync and Share systems. Each of the Sites is expected to be sustainable by itself financially and each of them has their own policies and procedures related to user management, data handling, operations, and security. Finally, each of the Sites has their own SLDs or SLAs with their clients and/or the users for which they run the service. These sites, instead of operating as a disjoint collection of service islands, become nodes in a mesh of interconnected storage, applications and users.

Interoperability between EFSS services is guaranteed by the Open Cloud Mesh standard[1] (OCM). To this "technical" coupling of nodes through the OCM standard, Science Mesh will add agreed-upon policies and procedures with respect to operations and security, together with a governance structure. The goal is to create a coherent Infrastructure that can guarantee a service level having an adequate quality. Operational procedures to achieve this are described in this document and the documents it references.

Science Mesh's small technical footprint will be matched by an equally lean administrative structure. It is our aim to rely as much as possible on policies, procedures, and agreements that are already locally available, and augment them with what is necessary for the Science Mesh to operate. This will also help with the platform's future sustainability and funding of activities, by keeping overhead costs as low as possible.

In this document several acronyms are used. These are explained in the Science Mesh Glossary[2].

[1] https://github.com/cs3org/OCM-API

[2] https://doi.org/10.5281/zenodo.5038663

2 Architecture of the Science Mesh

In this section, a brief recapitulation of the Science Mesh's architecture is given, as described in deliverable D2.1[3]. This provides some context which will be useful to understand the operational procedures that are discussed in this deliverable. For a detailed description of the workflows around monitoring, accounting, establishing trust relationships between Users, data and application sharing and so on, we refer to deliverable D2.1.

The Science Mesh consists of its participating Sites running EFSS services and a so-called Central Component. From a logical point of view, the Central Component is responsible for providing the few global services of the Science Mesh. However, it is by no means necessary that all those services run on a single node of the Science Mesh - they may be distributed across several Science Mesh partners. The interface between a Mesh node and the Mesh's core operational infrastructure is what we call an Executive Module (EM).

Figure 1 below displays the conceptual architecture: The Central Component of the Science Mesh includes the monitoring and accounting infrastructure as well as the Central Database and Helpdesk. The monitoring infrastructure runs probes which return results to the monitoring infrastructure as to if the services running at the Sites are operational. Similarly, the accounting infrastructure collects usage metrics from Sites. The Central Database provides the EMs running with Sites with Science Mesh topology information.

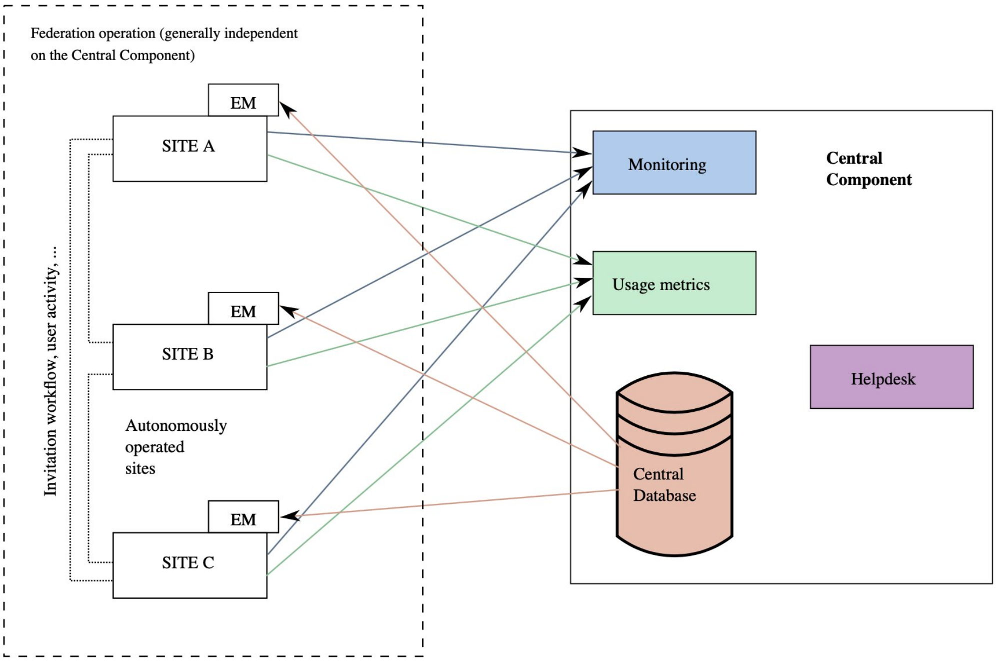

Figure 1 Schematic overview of the Science Mesh architecture showing the information flows between the Central Component and the Sites

[3] https://edu.nl/vfew9

Topology information about the Science Mesh is contained in the Central Database. This includes Site names, endpoints, services running at the Sites and further meta-information. Executive Modules consume this information and use it to perform service discovery for the users (similar to a WAYF service in identity federations) in order to establish trust relationship between Users. This serves mainly to reveal the User’s home EFSS and/or access services running at a particular Site. Executive Modules also enforce sharing policies implemented at the Sites.

Various functionalities of the Sites are periodically monitored. Accounting data is collected as well, containing gross aggregated usage statistics such as the number of shares and users, as well as the amount of storage used. None of the collected data can be traced back to an individual.

The Science Mesh is therefore a “share-nothing” infrastructure where all Sites can function independently of each other and do not depend on the Central Component to provide basic functionality. If the monitoring or accounting services go offline, the main service is not affected and only monitoring/accounting operations will be stopped. If the Central Database goes offline, the EMs cannot be updated. However, this only means that the information within the EMs may be stale, but the EMs remain functional. Since the Central Database will be based on a relational database management system, it will possible to opt for a geo-redundant solution with high availability, if it’s deemed that downtime is not tolerable.

3 Governance of the Science Mesh

3.1 Introduction

The governance model that has been selected for the Science Mesh draws heavily from the governance model adopted by GÉANT's eduGAIN federation, which has been successfully in use for several years. This model is described in detail in the Science Mesh Policy Framework Constitution[4]. In this document, we provide a summary of important principles.

3.2 Science Mesh Governance Bodies

There are three central governance bodies in the Science Mesh, which deal with oversight and technical issues, respectively. First, there is the Science Mesh Executive Board (SMEB), which consists of representatives of the Sponsors - those are the Sites which financially enable the Science Mesh to operate, e.g. through being a recipient of a grant and/or project.

The SMEB is responsible for ratifying the following decisions made by the Science Mesh Steering Group:

1. Changes to this Policy Framework Constitution.

2. Changes to and new proposals of any documents that contain mandatory elements for the Sites.

The Science Mesh Steering Group (SMSG) is a body consisting of Sites' representatives, and has an oversight and decision-making role in the Science Mesh Service. Participation of Sites in the SMSG is strongly encouraged but it is not mandatory. Each Site is invited to appoint a delegate and deputy to the SMSG.

The SMSG is mainly responsible for:

1. Approving changes to the documents in the Science Mesh Policy Framework, such as introducing, updating, or removing documents.

2. Approving the disqualification or temporary suspension of member Sites.

3. Appointing the Chair and non-voting observers to the SMSG.

The SMSG decides by voting. Decisions are made by a majority vote.

Finally, there is the Operational Team, which is responsible for the operation of centralised components in the Science Mesh, as well as technical and communication procedures regarding Site

[4] https://doi.org/10.5281/zenodo.5040152

membership. The Operational Team is responsible for:

1. Operations within the central Science Mesh infrastructure (e.g. Configuration Database, web site, etc.).
2. Collaboration with the operators of member Sites.
3. Reviewing and approving membership of new Sites.
4. Receiving enquiries about the Science Mesh and forwarding them to the appropriate body.
5. Processing applications to the Science Mesh and reviewing necessary technical aspects of them.
6. Providing support to the Sites regarding operations and configuration of the Science Mesh and its services.
7. Performing availability tests of the Sites (i.e., running tools to perform such tests), evaluating their Quality of Service, and handling security incidents regarding Science Mesh operations.

3.3 Declaration

From a formal point of view, sites join the Science Mesh by signing a Science Mesh Declaration[5] which is intentionally kept as minimal as possible. The Declaration requires the Site to publish their technical metadata, be willing to cooperate with the Science Mesh Operational Team, and to publish its own Terms of Service and Privacy Policy. It is explicitly stated that signing the Declaration creates neither legal nor financial obligations between Sites. The Site is required to comply with the Science Mesh Constitution (which itself refers to other technical documents).

[5] https://doi.org/10.5281/zenodo.5040087

4 Operational Procedures

4.1 Introduction

The collection of documents describing the Science Mesh Operational Procedures is called the *Science Mesh Operational Practice Statement*. The documents are described in the sections below. These sections also provide references to the documents.

4.2 Science Mesh Operational Practice Statement

4.2.1 Site Admission Procedure

The Site Admission Procedure[6] describes the technical requirements needed to be fulfilled by Sites to join the Science Mesh. In addition, it describes the procedural steps that need to be taken. The administrative requirements are covered by the Science Mesh Policy Framework Constitution[7].

4.2.2 Site Exit Procedure

The Site Exit Procedure[8] describes the consequences of leaving the Science Mesh that need to be dealt with properly in order to achieve a smooth transition. Steps that need to be taken by the leaving Site as well as remaining ones are described there, too.

4.2.3 Site Suspension Procedure

The description of the Site Suspension Procedure[9] discusses the possible reasons for Site suspension, the consequences for Sites which have been suspended, as well as the steps that are taken to suspend a Site and to lift an existing suspension. It should be noted that there are no requirements as to the Availability and Reliability of a Site. The reason for this also lies with the fact that we chose to keep the overhead needed to run the Science Mesh at a minimum. In addition, as mentioned above, the Science Mesh is a very loosely coupled e-Infrastructure. This means that each of the individual Sites does not depend at all on any other Site or services running within the Science Mesh in order to operate normally. If a Site encounters problems and goes offline, only its Users together with Users of other Sites with whom it has shared data or applications will be affected. While this certainly is a nuisance for the impacted Users, it does not prevent other Users of the Science Mesh to do their work.

[6] https://doi.org/10.5281/zenodo.5039989
[7] https://doi.org/10.5281/zenodo.5040152
[8] https://doi.org/10.5281/zenodo.5040130
[9] https://doi.org/10.5281/zenodo.5040639

In other words, the impact of a Site going offline is limited.

4.2.4 Support

The document called Science Mesh Support Procedure[10] describes how support for both Users as well as Sites is realised within the Science Mesh. It describes the formal support process as well as the informal community-based channels.

It is expected that Sites handle incoming issues and requests from their own Users. A central Science Mesh helpdesk will be there for Site admins - this is the formal part of the support. We expect that community-based support will play a valuable role in alleviating the burden of the central helpdesk system, by allowing admins to ask questions and discuss possible solutions.

4.2.5 Security

For a distributed infrastructure like the Science Mesh, where trust relationships between people have to be established using the various Enterprise File Sync- and Share systems and data as well as applications are shared, security is a primary concern. Well-established procedures around security must therefore be applied in order to provide guidance to site administrators as well as users and the Science Mesh governance bodies.

As already mentioned above, the philosophy underlying the Science Mesh is to make it as lightweight as feasible and to reduce the centralised overhead as much as possible. This holds for the central services that need to be run, for the support workflows, as well as the procedural overhead. We therefore do not impose a Science Mesh security policy upon Sites, but instead require that every Site has sufficient security policies in place that are applicable for all services they run, their EFSS included. The security policies related to the Science Mesh will therefore only focus on security aspects impacting the Science Mesh as a whole and not an individual Site. At the time of writing, these policies and procedures are in development and will be described in deliverable D2.4 “Full implementation of security policies and procedures” which is due in month 24 of the Project, which is December 2021.

4.2.6 Personal Data Handling

The Science Mesh requires every Site to have a privacy policy in place. For that reason, the Science Mesh privacy policy will describe procedures of handling personal data within the Science Mesh. The Sites are advised to refer to the Science Mesh Privacy Policy in their particular Privacy Policies.

In terms of compliance with existing data privacy laws such as the GDPR, we have focused on collecting as little personal data as possible for the operation of the Science Mesh. Naturally, Science Mesh users might need to work with personal data and share it with others, but as far as regulations go, those users will be considered data controllers and have the responsibility the applicable legislation imposes on them. The Science Mesh Sites are data processors acting on behalf of their users.

[10] https://doi.org/10.5281/zenodo.5040028

The Science Mesh deals with three general types of personal data:

1. (operators) personal data of site operators, administrative contacts, members of Mesh governance bodies, and other relevant personnel,
2. (users) personal data of sync-and-share system end users, e.g. their email addresses, identities of the users in their particular sync-and-share systems,
3. (user’s data) user's data may contain a broad spectrum of personal Information as well.

Personal data stored centrally by the Science Mesh are strictly limited to contact information of Site administrators and other relevant personnel. The Science Mesh collects no end user personal data into centralised systems. The accounting data that is collected centrally will only contain gross usage statistics like the total number of shares, the total amount of data stored, and similar highly aggregated information. It will not contain any personal data.

A cornerstone of the Science Mesh operation is to enable users to discover their identities in their sync-and-share systems. To recap the procedure, users can invite one another to share a resource through any textual communication medium, like e-mail for instance. The invitee (i.e. the target user) initiates a discovery procedure identifying their sync-and-share system and the user ID in the target system during this process. It is vital to reveal this information for the resource share establishment procedure to work. In other words, it is a legitimate interest of the infrastructure to deal with this information; the user is free to abort this procedure at any point, which naturally leads to the sharing not to be established. The information is used for establishing the resource sharing only and then forgotten.

Repeating the full procedure for each and every case of resource sharing would be quite inconvenient for the users. Thus, to make the end user’s experience more pleasing, caching of information is available, providing an option to keep track of the target user's identity and their home EFSS at the originating site. As caching is not strictly necessary as a basic functionality, it is purely optional and based on the target user's consent.

Finally, a user’s data itself may contain personal information. This is beyond the scope of the Science Mesh and must be handled by the individual end Sites. In such cases, the end users are solely responsible for handling their own collections of personal data in a lawful manner.

Before the Science Mesh enters the production phase, a Science Mesh Privacy Policy will be put in place.

5 Concluding Remarks

This document gives an overview of the work which has been done in order to set up Science Mesh's operational infrastructure policy-wise. The following documents have been produced as the Science Mesh Policy Framework:

1. Science Mesh Site Admission Procedure[11]

2. Science Mesh Site Exit Procedure[12]

3. Science Mesh Site Suspension Procedure[13]

4. Science Mesh Support Procedure[14]

5. Science Mesh Policy Framework Constitution[15]

6. Science Mesh Declaration[16]

7. Science Mesh Glossary[17]

The Science Mesh Privacy Policy is still work in progress. We expect this to become available soon. One of the guiding principles while creating this Science Mesh Policy Framework[18] was to make the Science Mesh e-Infrastructure a federated infrastructure with the least possible central overhead, i.e. making things as lightweight as possible. Another important guiding principle has been, with the GDPR in mind, to collect and use as little personal data as possible.

We consider this set of documents to be just a starting point which we nonetheless consider to be sufficient to move the Science Mesh infrastructure into a production state. Naturally, as the Project advances, lessons will be learned and the necessary adjustments will be made to the Science Mesh Policy Framework.

[11] https://doi.org/10.5281/zenodo.5039989
[12] https://doi.org/10.5281/zenodo.5040130
[13] https://doi.org/10.5281/zenodo.5040639
[14] https://doi.org/10.5281/zenodo.5040028
[15] https://doi.org/10.5281/zenodo.5040152
[16] https://doi.org/10.5281/zenodo.5040087
[17] https://doi.org/10.5281/zenodo.5038663
[18] https://edu.nl/q436n