D3.7 Feasibility Assessment of the Implementation of X-Road for Research Data

Author(s)	Miikka Kallberg (CSC), Martti Pelttari (CSC), Joni Virtanen (CSC), Sara Garavelli (CSC), Iiris Liinamaa (CSC), Ilja Livenson (UT/ETAIS)
Status	
Version	Final version
Date	1.0

Document identifier:	
Deliverable lead	CSC
Related work package	WP3
Author(s)	Miikka Kallberg (CSC), Martti Pelttari (CSC), Joni Virtanen (CSC), Iiris Liinamaa (CSC), Ilja Livenson (UT/ETAIS), Sara Garavelli (CSC)
Contributor(s)	
Due date	31/05/2022
Actual submission date	
Reviewed by	Abdulrahman Azab (UiO), Claudio Pica (SDU)
Approved by	
Dissemination level	Public
Website	https://www.eosc-nordic.eu/
Call	H2020-INFRAEOSC-2018-3
Project Number	857652

Start date of Project	01/09/2019
Duration	39 months
License	Creative Commons CC-BY 4.0
Keywords	Service(s), automation, cross-border

Abstract:

This deliverable reports on the second iteration of feasibility assessment of using X-Road for cross-border research data exchange. It [illegible]xperience with integrating a new service with X-Road and provides a set of recommendations for EOSC for cross-border service delivery.

Table of Contents

Table of Figures and Tables

1 Executive Summary

The document describes the continuation of the work performed when analysing feasibility of using X-Road-based solutions, adopted in Estonia and Finland, for exchanging of research data and delivery of services in cross border scenarios. X-Road® is an open-source software and ecosystem solution that provides unified and secure data exchange between organisations. X-Road is developed by the Nordic Institute for Interoperability Solutions (NIIS), a nonprofit association established in 2017 by the governments of Estonia and Finland. Furthermore, the national X-Road environments in Estonia and Finland are federated, which allows for an easy technological way of establishing trustworthy communication between organizations participating in X-Road environments in Estonia and Finland.

In our second report on X-Road, we have evaluated three additional aspects of the cross-border interactions.

- Practical experience and integration effort required for introduction and connection of services not aware of X-Road into the existing environment. For that, a service called Fiona provided by Statistics Finland and operated by CSC has been selected.
- Understanding the process of uptake of an integrated service by users from another country. In particular, what legal and organizational steps would need to be taken. The goal was to understand if X-Road increases or decreases the efforts for establishing collaborations with new clients.
- Finally, a comparison of the service delivery model in X-Road with that of EOSC was performed. During the execution of EOSC Nordic project, a service addressing cross-border delivery of the HPC resources called Puhuri has been developed. While targeted initially at LUMI[1], it has been designed in a generic way and has been added to the analysis as well.

Overall our feasibility study has shown that X-Road can be very useful for enabling services working with sensitive data. Due to current technological limitations working with larger volumes of scientific data can be challenging, however once the fix for a new mode of data transfer has been implemented NIIS, X-Road can be considered a strong choice for the services in the countries, where X-Road has been adopted on the national level.

1.1 Document structure

The document is structured as follows:

- The next chapter, “Motivation for the feasibility assessment”, provides a summary of why we did the work and wrote this summary.
- Chapter 3 concentrates on the steps taken for integrating an existing service with an X-Road based federation as well as analysis of the work.
- Chapter 4 describes the cross-border access process for a service based on X-Road.

[1] https://lumi-supercomputer.eu/resource-allocation/

- Chapter 5 provides a summary of analysis of the delivery process for X-Road, EOSC and Puhuri.
- Finally, Chapter 6 provides recommendations to EOSC based on lessons learnt during the feasibility study.

2 Motivation for the feasibility assessment

X-Road technology has been widely used in the domain of data exchange mostly by public sector organizations in Estonia and Finland. Recently, the countries have started federating their environments, which opened a way for building cross-border services based on a legally binding and sustainable infrastructure.

Furthermore, X-Road based environments are operating under high load, e.g. in April 2022 Estonian environment had over 160 mln requests performed[2] and are used for transferring sensitive data in a safe and compliant manner.

X-Road sustainability is assured by the states using the software as the core data exchange platform. X-Road environments connect both public and private sector users.

Such a basis invites the question if X-Road based environments can also improve the use cases arising in the academic and R&D context. While X-Road is a more query oriented system, it is typical for scientific environments to operate on larger data sets and move the data from storage to processing farms. In our second iteration of the feasibility analysis, we have concentrated on taking a real life use case and analysing what benefits or drawbacks integration with the X-Road environment can bring. Furthermore, we wanted to understand if X-Road can also help with the service access.

Finally, we wanted to analyse to what extent EOSC service delivery differs from the one proposed by the X-Road system. At the time of writing, the EOSC delivery model has not yet been clearly formalised and to complement the analysis, we also included another system aimed at cross-border service delivery - Puhuri. It has been developed in conjunction with the EuroHPC LUMI supercomputer hosted in Finland and used by several partners of EOSC Nordic. This way we can compare 3 systems with different scopes: most specific (X-road), most generic (EOSC) and a middle ground (Puhuri).

3 Integration of a new service

During the first iteration of the feasibility study of X-Road, we took a service that was natively developed for using X-Road. Hence, we could not estimate the effort required for integrating an existing service. In the second iteration, we decided to go with an existing service and partnered with Statistics Finland for that.

[2] https://www.x-tee.ee/factsheets/EE/#eng

3.1 Background about Fiona and Statistics Finland

Statistics Finland started developing a Secure Remote Desktop-environment for researchers back in 2008 and it was migrated to the CSC environment and re-branded to FIONA in 2015. Current version of FIONA was developed as part of Sitra's ISAACUS -project and it replaced the previous version in 2021.[3]

The FIONA remote access system is a data secure environment for processing research data for unit-level data needed in research, such as Statistics Finland's micro data. Micro data or unit-level data are available from Statistics Finland for scientific studies and statistical surveys.

The Research Services offer ready-made data and tailoring of data according to research need. Data are available from both enterprise and individual databases in a variety of ways. The data, their tailoring and remote access are subject to charge.

Statistical legislation and data protection and confidentiality practices specified in legislation are applied in compiling and releasing the data. The releasing of data is subject to a user licence.

[4] Working in FIONA takes place via the desktop that opens in the browser window. FIONA is a closed system, which means that no information is transferred between the user and the environment except for the display image and keyboard and mouse input.

Currently data delivery to FIONA is a manual process and it requires human resources to deliver to FIONA for the researchers. During Sitra's ISAACUS project CSC did a proof of concept to test data transfer via X-Road, but at that point there weren't technical capabilities to integrate FIONA to X-Road. [5]

Research results and other material may be transferred outside the system only through a screening process. The screening ensures that no individuals or businesses can be identified from the published data. Statistics Finland screens all material that users request to transfer outside of the system. In addition to Statistics Finland's data, data from other authorities can also be combined in FIONA in the same environment. The calculation and storage capacity of FIONA is available as needed. FIONA also contains several analysis programmes, such as STATA, R, Python, SAS and SPSS.

[3] https://www.sitra.fi/en/articles/one-stop-shop-well-data-isaacus-laid-foundations-future/
[4] https://www.stat.fi/tup/mikroaineistot/index_en.html
[5] https://www.sitra.fi/en/articles/one-stop-shop-well-data-isaacus-laid-foundations-future/

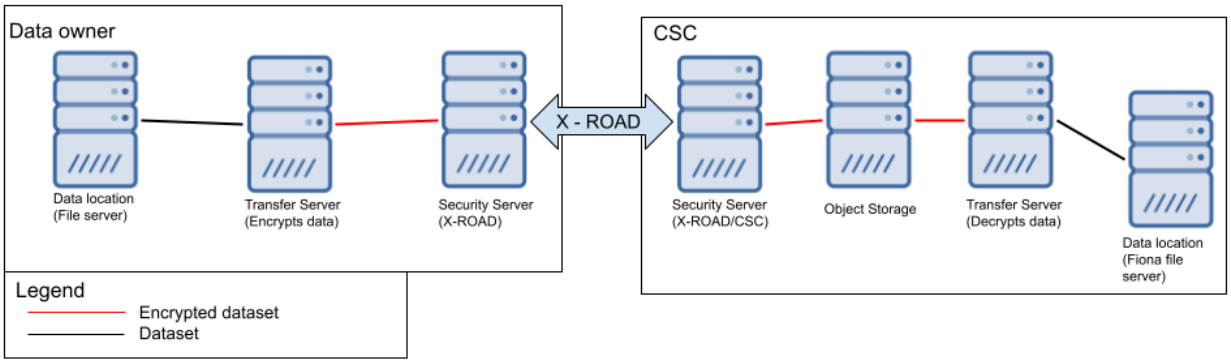


Figure 1 X-Road datasets legends

3.2 Data transfer overview

Basic idea behind this process is that the data that is transferred via X-Road is labeled protected data and it needs to have additional encryption on top of X-Road's security layers. Additional encryption is added to ensure that this solution can be used to transfer classified or other non-public data. Encryption keys for this additional encryption are handled in a separate process that is not described in this study. Because of that encryption requirement, the dataset needs to be prepared and encrypted before the transfer and decrypted after the transfer.

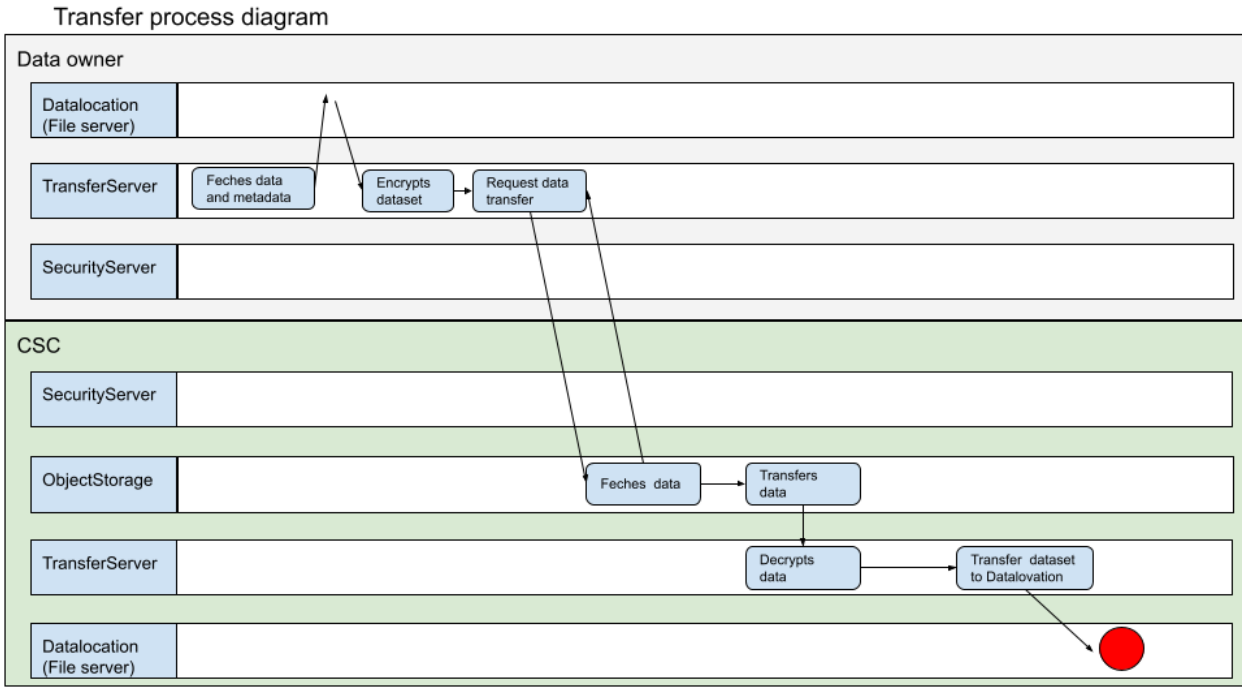


Figure 2 Data transfer process diagram

3.3 Technical solution

X-Road supports SOAP and REST requests. The standard use case for X-Road has been to query for limited data sets or individual records based on parameters. With file transfer it was important to find a solution that has high reusability, good amount of documentation, open source applications and would be able to handle large amounts of data being sent and received. Another beneficial feature would be support for REST or SOAP requests. At the first stage CSC focused on two alternatives: object storage and SOAP requests with attachments. Object storage seemed the better suited alternative for the use case and the team at CSC had already used it with other use cases. Objects are also supported in the ePouta OpenStack environment even though the Fiona environment at the moment relies on a different storage solution.

Object storage provides data storage architecture based on objects. Almost all object storage solutions also provide support for REST API and different software development toolkits (SDK). X-Road also supports REST API, so operations done to object storage via the REST API would be transferred over X-Road. First idea was to publish data to be fetched with X-Road. It was then seen that the idea of having a point for receiving data from different organisations had more reusability than requiring the organizations to publish an object based storage system to X-Road. FIONA environment could receive data from different organizations so it would be difficult to require those organizations to host their own object storage solutions.

The development environment consisted of a Minio cluster[6] for the object storage part and a containerized version of the X-Road Security Server[7] running on a dedicated host machine. This docker container could also run on any container cloud, but CSC does not have a container cloud option in our development environment. This way the CSC team could set up relatively quickly an S3 compatible object storage to have full control over.

3.4 Analysis of integration between X-Road and the object storage

Integration to the X-Road Finnish development environment was relatively simple. Most time-consuming part was to wait for the firewall requests to be fulfilled by the Finnish X-Road support team. The documentation for running the X-Road docker container was sufficient and X-Road itself was easy to use. Setting up the required clients and services was more complicated and CSC had to make a few shortcuts to make them work.

As Minio is S3[8] compatible it uses Amazon Web Services signatures (version 4) for authentication. The authentication model is different and has downfalls when combined with a service like the X-Road. AWS signature is calculated from the following information:

- Endpoint Specification
- Action

[6] https://min.io/
[7] https://github.com/nordic-institute/X-Road-Security-Server-sidecar
[8] https://aws.amazon.com/s3/

- Required and Optional Parameters
- Date
- Authentication Parameters[9]

With X-Road the endpoint is not the same as the actual hostname of the S3 object storage. Signature calculated for the client is different from what S3 expects it to be. You can code your own client to send the signature using the end-point of the S3 instead of X-Road Security Server, use HTTP/2 requests or calculate the signature at the backend and use a token-based authentication for the X-Road requests. For CSC's use case public buckets were used that were published in the X-Road so that they required no authentication. The X-Road client calling the service was authenticated with a TLS client certificate and a private key.

Another pitfall with using object storage REST requests with X-Road was the way that X-Road handles Content-Length headers.

"Content-Length
- The Security Server MAY change the transfer encoding, thus removing or adding a content-length header as necessary."[10]

Performance of the larger data set transfers can be an issue in X-Road[11]. Transfer between security servers is done by X-Road's own transfer protocol and while CSC did research the REST protocol before our development phase, did not understand the impact of this on the method. Sending large and or binary data can be done in two different ways: Multipart/Form-Data or S3 multipart upload. Both of these methods require that the content-length header is present and valid. With X-Road the content-length header was missing on the requests. As a workaround a proxy for object storage can be set up that calculates the content-length and this proxy could also calculate the signatures and set up a token-based authentication. One solution like this is the Nginx S3 Gateway, but that only works for getting objects from storage and not for sending them.[12]

Getting objects from object storage via X-Road was easy and the transfer speeds where similar for when getting data directly from object storage or via X-Road.

We made a support request for the X-Road development team and they responded that they have made a feature request from the use case with the estimate that it would be released in Q4/2022 with the X-Road version 7.2.0. The issue can be seen in the X-Road backlog[13] (requires registration). As X-Road has been used for transferring smaller amounts of data it has not had this issue before.

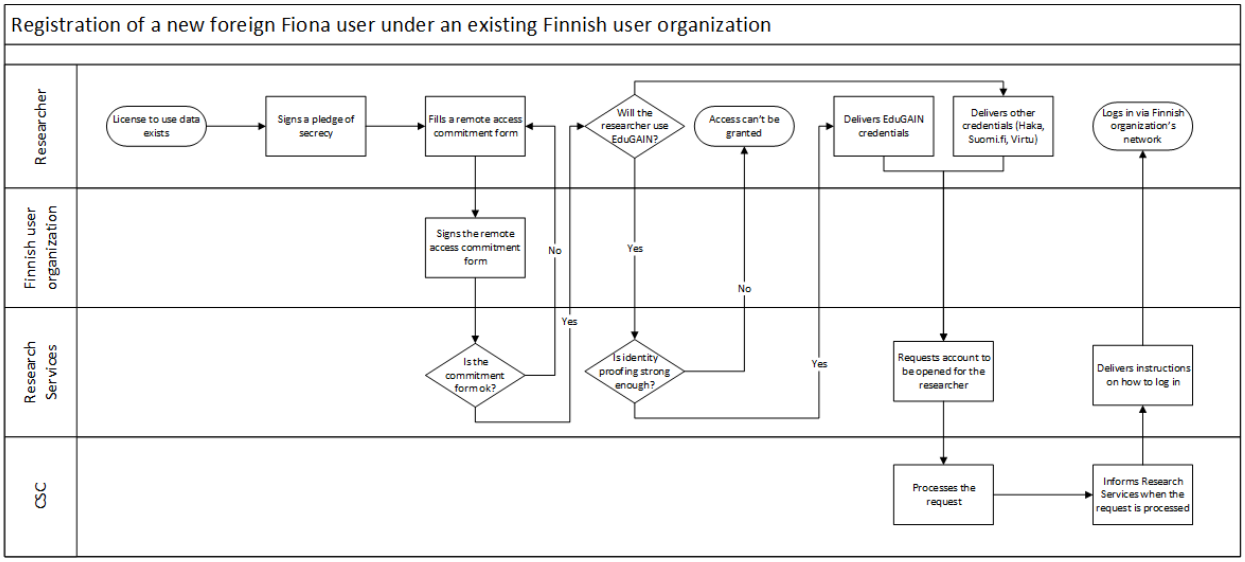
---

[9] https://docs.aws.amazon.com/general/latest/gr/sigv4_elements.html
[10] https://www.x-tee.ee/docs/live/xroad/pr-rest_x-road_message_protocol_for_rest.html
[11] https://zenodo.org/record/6274890#.YhjO_YqZM2w, 3.1.1 Performance
[12] https://www.nginx.com/blog/using-nginx-as-object-storage-gateway/
[13] https:/jira.niis.org/browse/XRDDEV-2022

4 Accessing X-Road-based service from the other country

Adoption of the technology for cross-border data exchange can have an impact on the service usage by end-users, as the natural follow up is increase of interest for a service from abroad. To understand better if “xroadified” service could be accessed from abroad, we performed the analysis of access process. While the process below is specific to Fiona, we believe that it is representative of similar data analysis services.

4.1 How can one become a remote user of research data?



Registration process of the new foreign user Fiona diagram is in Appendix 1 in Figure 4 as well.

There are prerequisites

1. User has a valid license to use data.
2. User is affiliated to a Finnish organization that has an agreement of remote access with Statistics Finland.
3. User operates out of an EU country or from a country that is listed in a list of countries that have an adequate level of data protection.[14]

14

https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en

4.2 Organisation agreement on remote access use of research data

For a researcher to be able to work in the remote access system, the researcher's organisation must conclude a remote access agreement with Statistics Finland. In connection with the processing of a user license application, the Research Services supply to the customer for filling in the organisation agreement on remote access use and a request for data protection and data security practices connected to the organisation's remote access connection.[15]

The customer is responsible for ensuring that the premises used by researchers are suitable for research and remote use and that data can be processed there so that the data protection of the data is not endangered. Statistics Finland must have the possibility to view the customer's work premises.

The organisation agreement on remote access use specifies the contact persons responsible for administrative and technical matters relating to the remote access use of the organisation. The contact persons must take part in training related to remote access use.

4.3 User and workspace-specific agreement

Each remote user must submit the user and workspace-specific remote access commitment to Statistics Finland before starting the remote access use. All remote access locations and addresses must be given by a remote access commitment.

Researchers can apply for a license to use the remote access system from home or abroad (EU and non-EU countries whose level of data protection has been recognised as sufficient by the European Commission). For use from abroad, the research project must be connected with a Finnish organization. In addition, it must be mentioned in the user license for the research project that remote access use takes place abroad.

The customer organization is responsible for their employees' remote work and remote work connection.

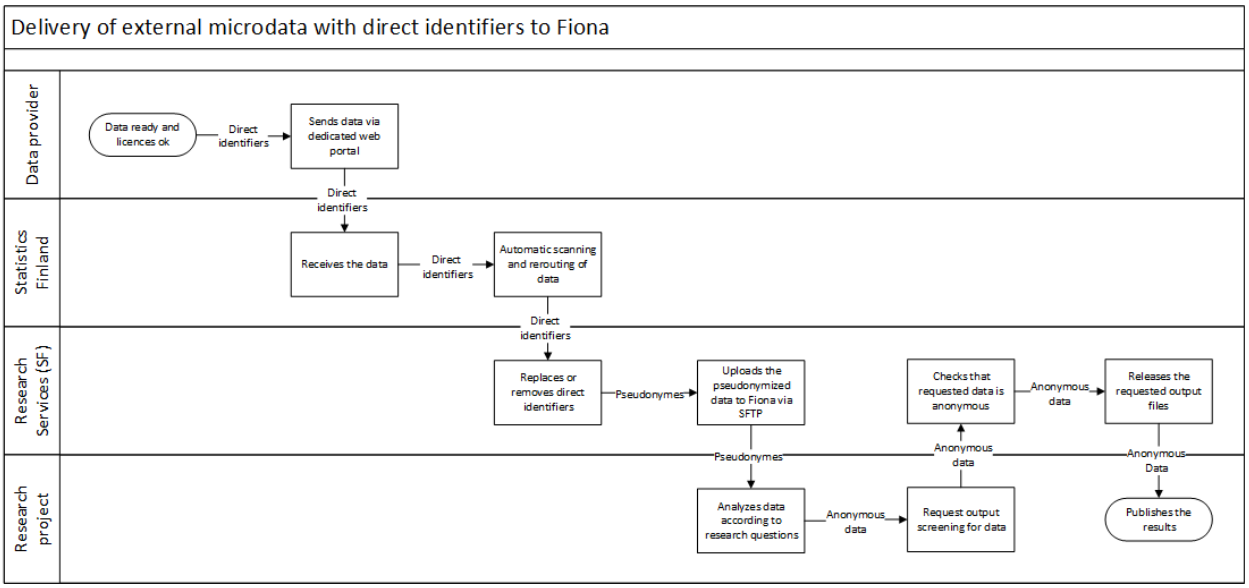
Logging in to the remote access system requires the use of federated identification services, of which the following are in use:

- Suomi.fi
- Finnish Government identification service
- Haka
- EduGain

Currently data delivery to FIONA is a manual process and it requires human resources to deliver to FIONA for the researchers. During Sitra's ISAACUS project CSC did a proof of concept to

[15] https://www.stat.fi/tup/mikroaineistot/etakaytto_en.html

test data transfer via X-Road, but at that point there weren't technical capabilities to integrate FIONA to X-Road. [16]



The process can be found on Appendix 2 in Figure 5: Delivery of external microdata with direct identifiers to Fiona.

Using X-Road for data transfer would help to automate the process and speed it up. X-Road integration would also make it possible to fetch data from data sources that are exposed to X-Road in any country. X-Road's security architecture makes it easier to build trust between organizations.

5 X-Road vs EOSC vs Puhuri service delivery models

X-Road-based environments come with a set of rules governing operations of the environments. The same is true about the ecosystem that EOSC is working towards. Hence, it is valuable to compare and analyse the commonalities and differences that X-Road systems and EOSC have. For the reference X-Road environment we have chosen an Estonian production environment[17] - the other existing ones are very similar with only minor differences in specifics that do not matter for this analysis.

Regarding EOSC, we base analysis on the D2.4 The EOSC Delivery Chain[18] as well as participation in meetings and discussions around the EOSC in various forums.

[16] https://www.sitra.fi/en/articles/one-stop-shop-well-data-isaacus-laid-foundations-future/
[17] https://www.ria.ee/en/state-information-system/x-tee/introduction-x-tee.html
[18] https://zenodo.org/record/5541424#.YqZbD-xBwqJ

In addition to X-Road, we have also included into analysis Puhuri[19] service - a service co-funded by NeIC and several national Research Infrastructure, for providing access to LUMI supercomputer[20] from LUMI partner countries: Belgium, Czech Republic, Denmark, Estonia, Finland, Iceland, Norway, Poland, Sweden and Switzerland. Each of the countries have nominated one or more organizations for providing LUMI services to their target communities. In total at the moment of writing, there were 26 organizations providing access to LUMI services. Puhuri is an important case as it was designed to cater for automated delivery of services across the border and was chosen to be also the basis for the EOSC Nordic service gateway.

5.1 Overview of Puhuri service

Puhuri provides resource allocation and authentication infrastructure for service providers. As the system is new we provide additional details of its architecture below. Puhuri's service delivery is based on the agreements that Service provider and Resource Allocators have established prior to the usage of the system, e.g. deciding on the quotas of the resources that can be allocated in a digital way.

Puhuri AAI is part of MyAccessID authentication and authorisation infrastructure (AAI) services based on GEANT eduTEAMS.

Figure 3 below summarises the architecture of the Puhuri consisting of two layers: identity layer and Infrastructure Service Domain (ISD) layer.

The **identity layer** is responsible for delivering consistent identity information across clients and services of Puhuri. The registration creates a unique identifier (Community Unique Identifier, CUID) for the user in MyAccessID registry, which is used for referencing and linking user identity across the different components. The user registration process can be started from the existing portal of Resource Allocators (e.g. national allocation portal) or from a Puhuri Portal, which is provided as a reference solution. Identity layer also assures that supported identity providers release the attributes about user identity in a common way so that services could be built with end-to-end user authentication.

Users can optionally register SSH public keys with their MyAccessID profile, in which case they become available to the service providers that user gets access to.

On the **Infrastructure Service Domain layer**, Puhuri Core, operated by University of Tartu, is the resource allocation service exposing API for Resource Allocators and Service Providers for managing Projects, Members (using CUID of Puhuri users as references) as well as passing information on Resource Allocations, their usage and lifecycle.

[19] https://puhuri.io
[20] https://www.lumi-supercomputer.eu/

The Puhuri services deal with personal data and have been analysed from the perspective of GDPR compliance. Required contracts have to be signed prior to access to the production environment of Puhuri Core.

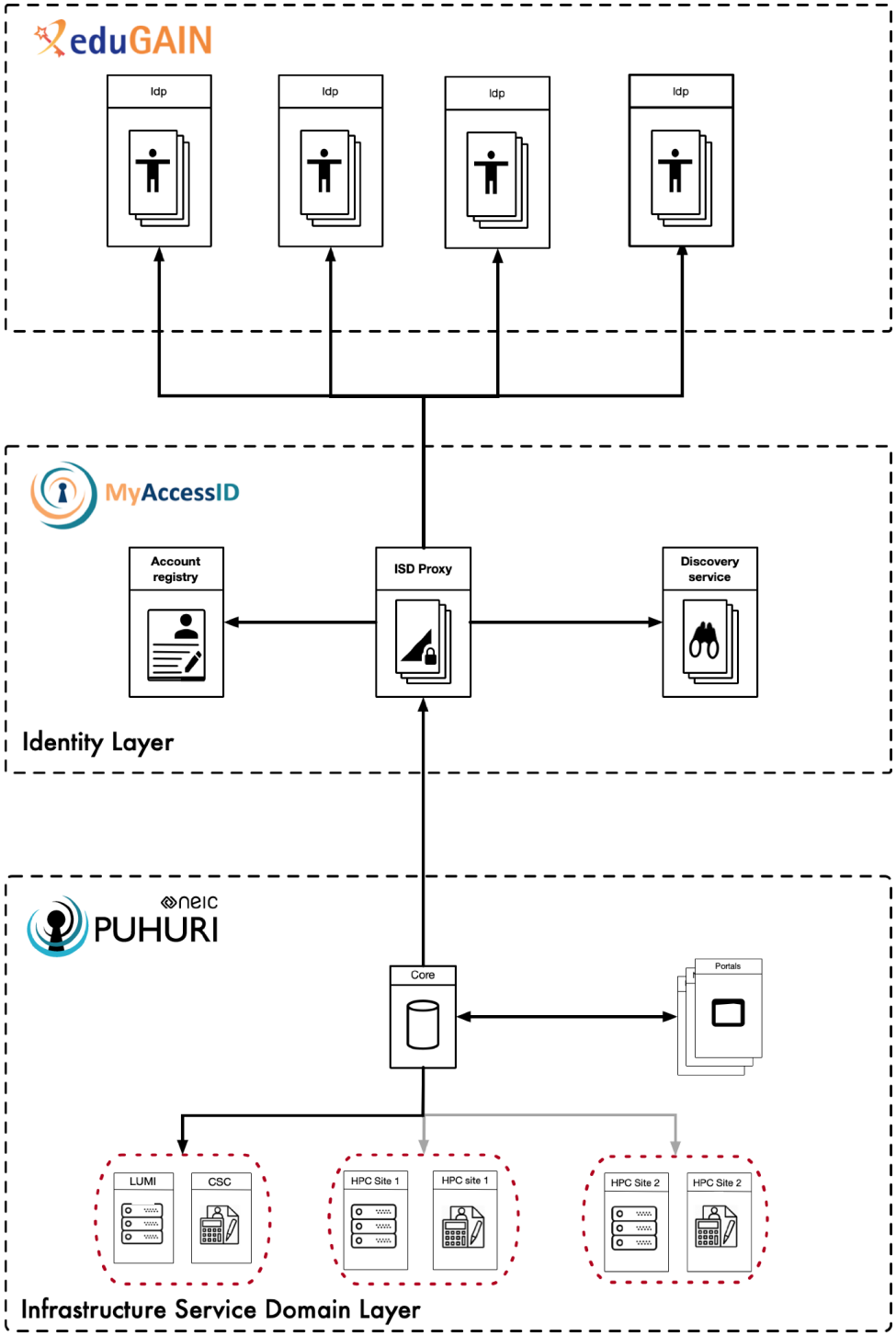


Figure 3 Overview of the Puhuri architecture

5.2 Defined service delivery aspects

To perform the analysis, we have selected several aspects that are common in a service delivery. The aspects and the motivation for selection are as follows:

1. **Rules of participation**: explanation of the governance of participations
2. **Stakeholder verification process**: information exchange typically means transfer of data from one trusted party to another. Verification of stakeholders can be used for increasing the trust in the system.
3. **Service providers**: typical service providers
4. **Service consumers**: typical consumers of services
5. **Services**: typical services
6. **Central services**: common services provided by the environment
7. **AAI**: identity management and authorization within the environment
8. **Accounting**: service usage accounting
9. **Monitoring**: how health of the environment is assured
10. **Service publication**: how publishing of new services is organised
11. **Order management**: how requesting of a service access is organised
12. **Transactions (cost)**: approximate cost of a single transaction in the environment
13. **Transaction (volume)**: approximate volume of transactions in the environment
14. **Sustainability**: how sustainability of the environment is assured
15. **Motivation for service providers**: what motivation is provided to service providers for participating in the environment
16. **End-user access to service**: how end users access environment services
17. **GDPR for transactions**: to what extent the environment is GDPR aware

5.3 Comparison of the service delivery aspects in X-Road vs EOSC and Puhuri

Aspect	X-Road	EOSC	Puhuri
Rules of participation	Defined by the legislation of the country.	Defined by EOSC Rules of Participation, evolution is not very clear but probably governed by EOSC Association.	Defined by a steering group represented by delegates of the consortium.
Stakeholder verification process	Very strong. Validation of user identities and	Weak, based on user-provided info.	Strong, based on signed contracts with central service

	representation rights for the legal persons.		operators.
Service providers	Mostly public agencies, some private companies	Research performing organisations, research infrastructures	Providers of services aimed at cross-border usage,
Service consumers	Public agencies and private companies	Researchers, Research Infrastructures	Allocators of resources
Services	Mostly data services and registry data	Wide range of different services	Primarily infrastructure services
Central services	Registry, CAs, Timestamping	EOSC Core: Supply/Demand portal, Accounting, Monitoring, Helpdesk	Puhuri AAI and Puhuri Core
AAI	eIDAS-compliant, relies on x509 certificates	EOSC AAI as a planned standard, EOSC Portal AAI	MyAccessID AAI-based service
Service publication	Self-service for validated service providers	Self-service via Supply portal for validated service providers	Self-service via Puhuri core for validated service providers. Manual activation by Puhuri operators for entering production state.
Order management	Access to services is granted based on access policies agreed outside the environment.	Process is supported by EOSC Demand portal, but no common policy for ordering	Formalised order management process that is based on digital service delivery
Transactions (cost)	Low cost data lookups	Varies depending on service	Small to large, depending on the specific allocation type
Transaction (volume)	Very high number of transactions	Low volume	Average volume
Sustainability	Guaranteed by states operating X-Road	Unclear	Revenue generation via subscriptions for service providers (e.g. LUMI), partially

			sponsored development
Motivation for service providers	Legal mandate, easier integration of services by customers	Better visibility of services	Easier reach of users across the border, lower cost of service delivery via automation
End-user access to service	Service specific	Varies, partially open access	Service specific
GDPR for transactions	Outside the scope of X-Road	Not currently handled	Analysed with contract signed as part of on-boarding

Table 1 Comparison of service delivery aspects

5.4 Summary of comparison

All three systems - EOSC, X-Road and Puhuri - are aimed at improving cross-organizational usage with support for cross-border interactions. The level of support and prioritization are however different.

X-Road as well as Puhuri are following Business-to-Business model (B2B), where the information flows between strongly authorized parties, which then build solution on top for the end users. In such a model, there are fewer stakeholders that need to be strongly authenticated, however the number of transactions is generally higher as stakeholders can represent larger user communities, for example, a family doctor office serving multiple patients in case of X-Road or a national allocation portal serving researchers of the country in case of Puhuri.

EOSC has a much wider scope, aiming at capturing interactions also with end-users. While the exact vision of EOSC is still being formed at the moment of writing of the deliverable, the wide scope of EOSC unfortunately leads to very generic Rules of Participation and unclear Sustainability model from the perspective of Service Providers. On the other side, X-Road and Puhuri, due to high requirements for the authentication of users as well as service provider verification and expectation on the fully digital delivery of services are more intrusive and have higher threshold of integration. EOSC requirements are more relaxed and lead to easier on-boarding, however the lack of sustainability offered to service providers means that the environment will struggle to maintain its appeal to target users as service providers might not be willing to integrate with the proposed order management system and hence the EOSC services in EOSC registry will be rather advertisements than actually available services.

6 Recommendations to EOSC

Experience gained during the piloting of X-Road for the scientific cases allows us to formulate several recommendations for EOSC to succeed as an environment from the service provider perspective.

1. X-Road was chosen for the pilot project as both in Estonia and Finland the law is in place that communication among government systems should use X-Road channel. Among other things, this provides an environment with clear and consistent trust guarantees for service providers and reduces effort required on re-validation of parties by service providers. EOSC could consider introducing more clear and stricter policies for service providers willing to engage in digital transactions.
2. Re-use of the solution that was natively designed for a specific environment comes at a very low cost, as was demonstrated by experience from D3.6. However, integration of a service with a new environment is not trivial and requires extensive resources from the service provider. As such, a motivation needs to be present for the service provider to uptake such an effort. In case of X-Road, this has been twofold - mandatory requirement coming from the legislation as well as reduced effort on service provider organization due to a high trust within the system.
3. X-Road as a solution is a very interesting candidate for cross-border data exchange. However as long as streaming of larger datasets is not easily doable due to a technical limitation, it most probably can be used only as a control channel with the need of having a separate data channel. The issue has been raised and accepted by the X-Road developers, so hopefully will be fixed eventually. It is not entirely clear how a potentially useful data transfer system could be integrated with EOSC - or if it should be at all. One option is to consider X-Road as an access protocol and promote it as a technical capability of a service. It can be an important step towards composability of EOSC services, however not much work has been found by the authors of this deliverable on that in EOSC.

APPENDIX 1: Registration of new user to Fiona

Prerequisites:

1. User has a valid license to use data
2. User is affiliated to a Finnish organization that has an agreement of remote access with Statistics Finland
3. User operates out of an approved country: https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en

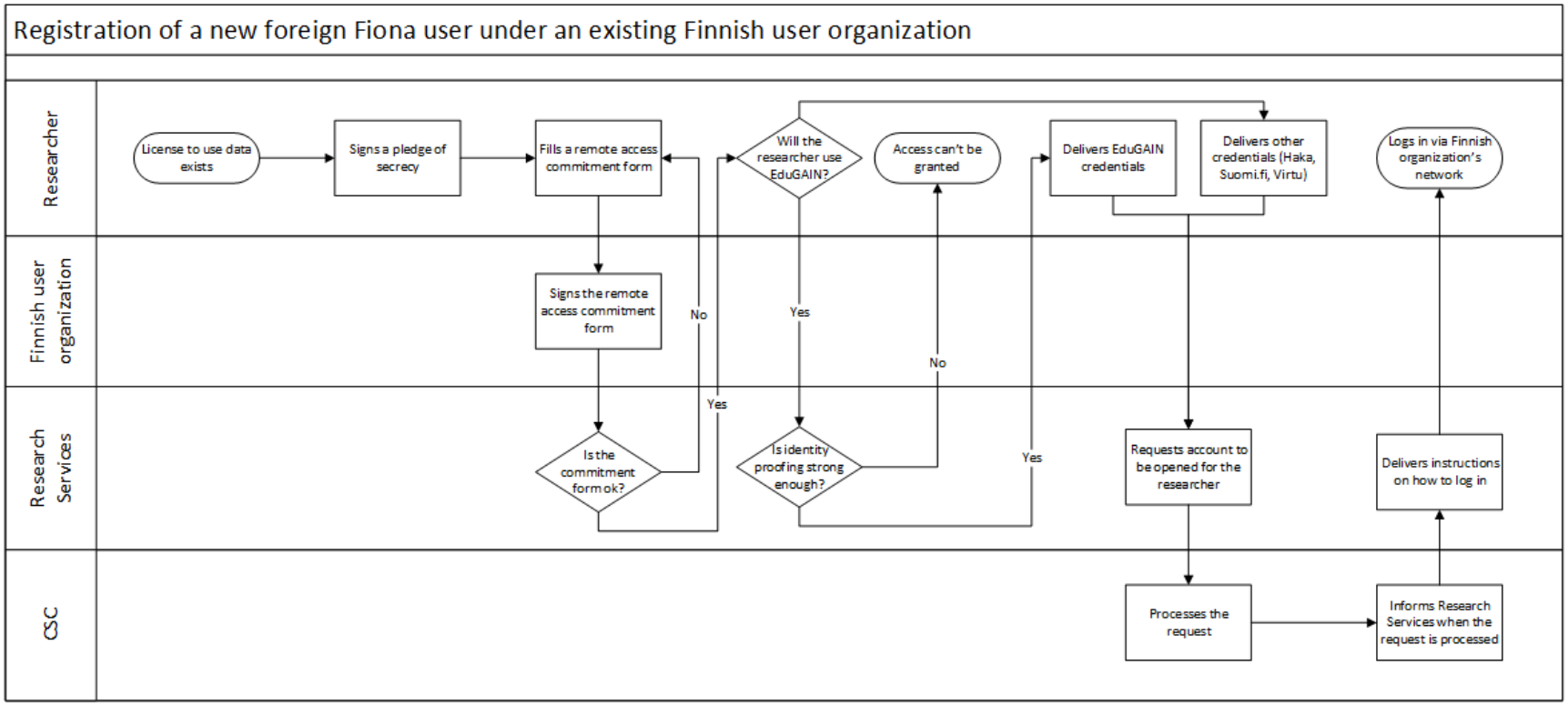


Figure 4 Registration of a new foreign Fiona user under an existing Finnish user organisation

APPENDIX 2: Delivery of external microdata with direct identifiers to Fiona

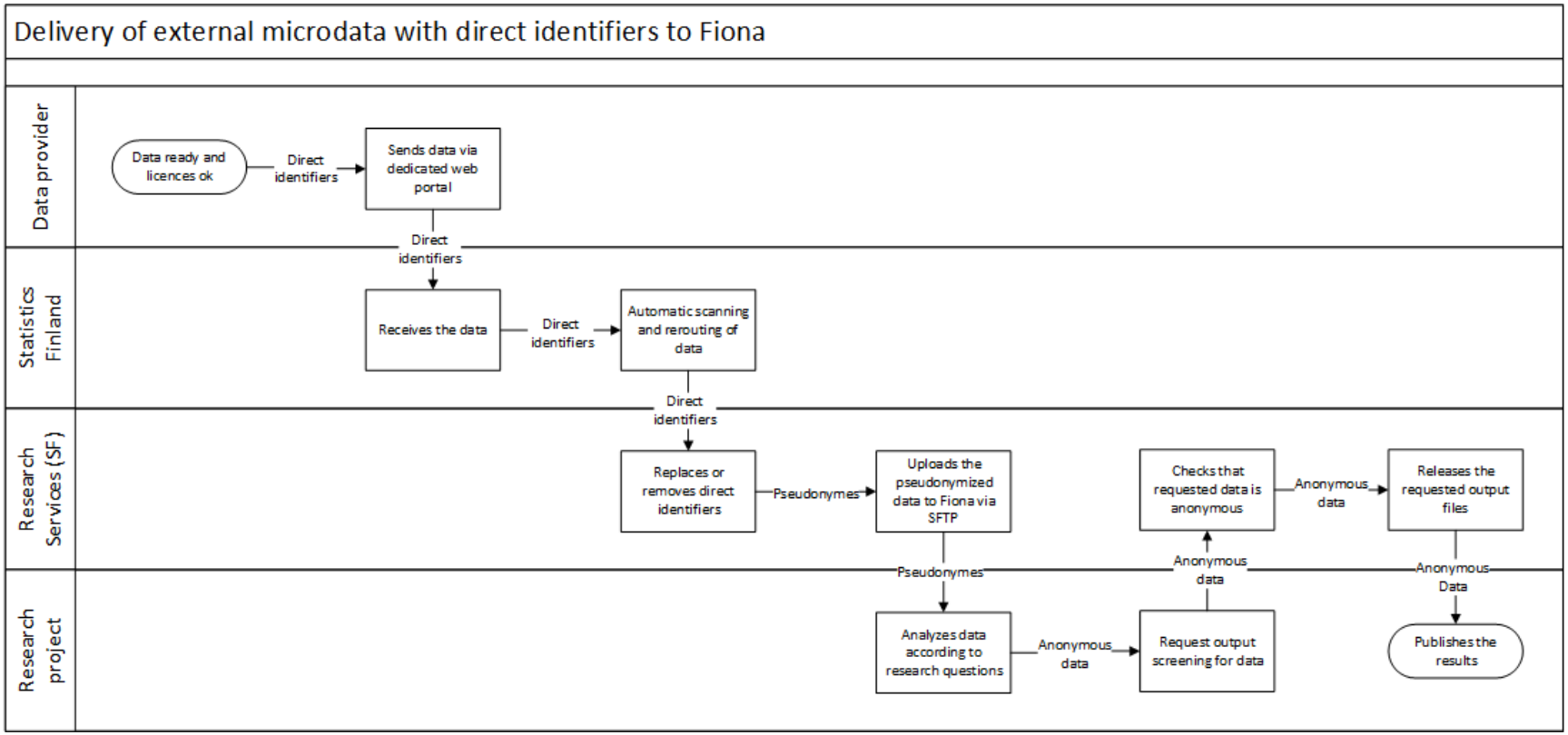


Figure 5 Delivery of external microdata with direct identifiers to Fiona