TEHDAS

Towards
European
Health
Data
Space

TEHDAS - Towards the European Health Data Space Joint Action
Start date of the project: 1.2.2021
Duration: 30 months

This project has been co-funded by the European Union's 3rd Health Programme (2014-2020) under Grant Agreement no 101035467.

TEHDAS WP7, Milestone 7.3

Report on EHDS architecture and infrastructure implementers’ expectations / experiences

Accepted in Project Steering Group on 31 May 2022

Co-funded by
The Health Programme
of the European Union

Document information

Document authors	
Author	**Partner**
Jaakko Lähteenmäki	**VTT Technical Research Centre of Finland**
Juha Pajula	**VTT Technical Research Centre of Finland**
Juan Gonzalez-Garcia	**IACS Aragon Health Sciences Institute**
Carlos Telleria	**IACS Aragon Health Sciences Institute**
Helena Lodenius	**CSC IT Center for Science**

Disclaimer
The content of this deliverable represents the views of the author(s) only and is his/her/their sole responsibility; it cannot be considered to reflect the views of the European Commission and/or the Consumers, Health, Agriculture and Food Executive Agency or any other body of the European Union. The European Commission and the Agency do not accept any responsibility for use of its contents.

Copyright Notice

Co-funded by
The Health Programme
of the European Union

Objective

- Collect architecture and infrastructure expectations and experiences of stakeholders
  → background for defining options for EHDS architecture and infrastructure to be presented in deliverables D7.1 (March 2022) and D7.2 (January 2023)

Co-funded by
the Health Programme
of the European Union

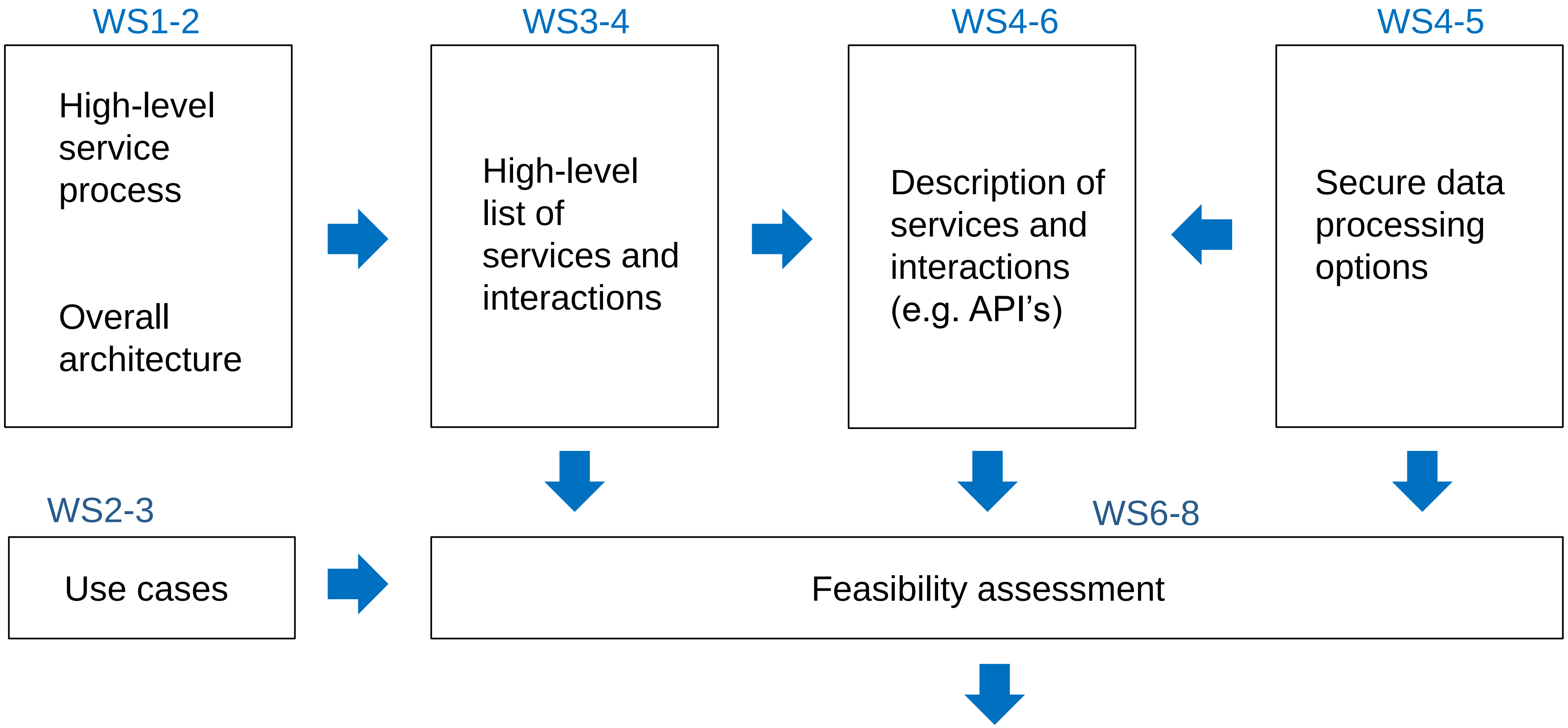
Method

- A group of experts was formed in the beginning of the TEHDAS project in spring 2021: **Work package 7 advisory group (WPAG7)**

- WPAG7 has been gathered for a series of workshops to collect views and opinions on EHDS architecture and services (see workshop topics next slide)

Co-funded by
the Health Programme
of the European Union

Workshop plan



WS1-2

High-level service process

Overall architecture

WS3-4

High-level list of services and interactions

WS4-6

Description of services and interactions (e.g. API's)

WS4-5

Secure data processing options

WS2-3

Use cases

WS6-8

Feasibility assessment

<u>Results:</u>

- EHDS service catalog
- EHDS architecture options

Co-funded by
the Health Programme
of the European Union

Workshops

Reported earlier in milestone 7.2 report:

- Workshop 1 (online), 18.05.2021, 40 participants
- Workshop 2 (online), 22.06.2021, 46 participants
- Workshop 3 (online), 14.09.2021,   39 participants

Reported in this milestone 7.3 report:

- Workshop 4 (online), 7.12.2021, 38 participants
- Workshop 5 (online), 15.02.2022, 36 participants
- Workshop 6 (online), 10.05.2022,   39 participants

Workshops to be organised in autumn 2022:

- Workshop 7, planned: 4.10.2022
- Workshop 8, planned: 29.11.2022

Co-funded by
the Health Programme
of the European Union

Workshop practices

- In all workshops, options and approaches for EHDS architecture and services have been presented to the meeting participants
- The materials have been provided to the participants two weeks before the workshop to ensure time to prepare for the workshop
- The participants have been invited to comment the presented architectural approaches:
  - orally or in chat during the workshop (Teams meeting)
  - on whiteboard (Jamboard) during or before the workshop

- Inputs have been recorded in minutes shared after the workshop to all participants

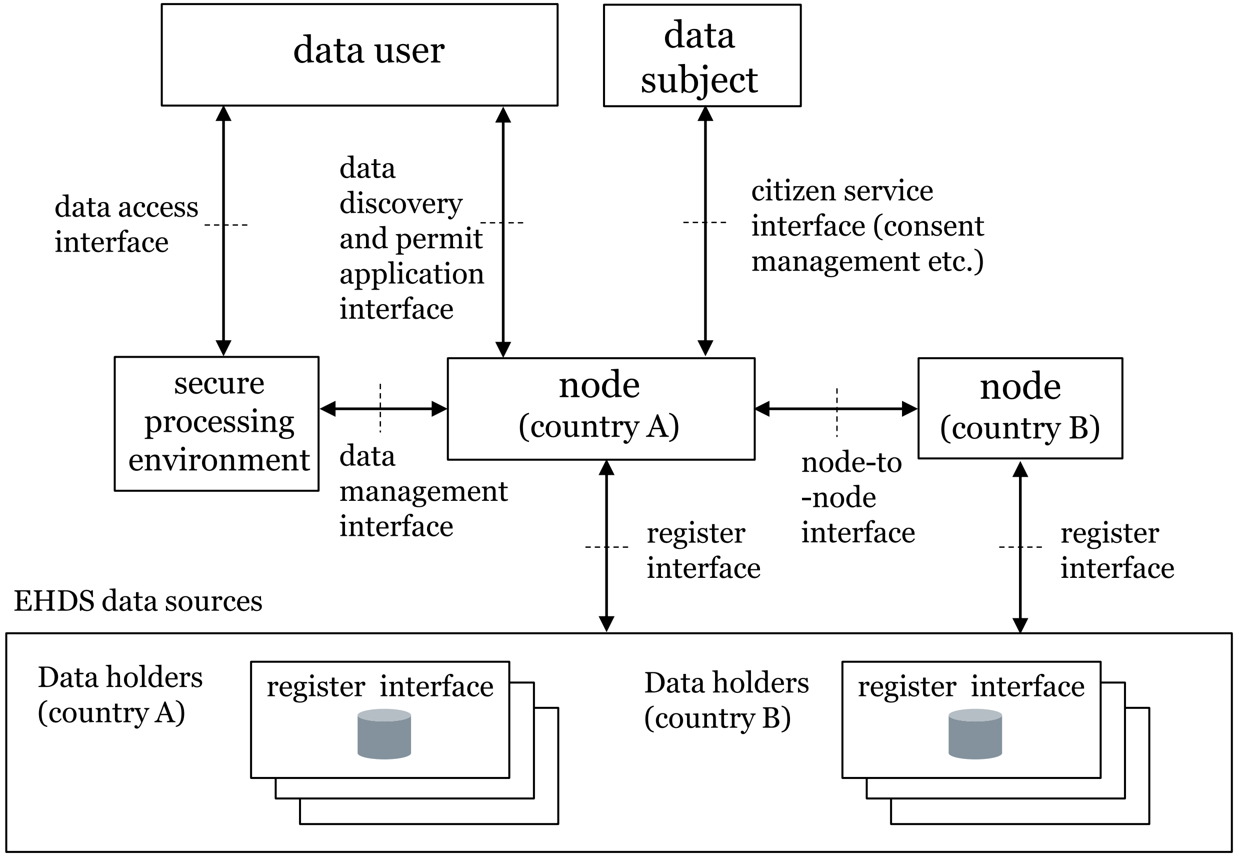
Co-funded by
the Health Programme
of the European Union

Workshop 4

- Invited presentation of TEHDAS WP6 (Excellence in data quality) given by Enrique Bernal-Delgado (IACS)
- Architecture approaches for comments:
  - interfaces for the overall architecture discussed in earlier workshops
  - secure processing environment (SPE) approach
- Additionally discussed topics:
  - legal and governance model related issues
  - terminology
  - nature and amount of nodes
  - data user's own data (usage combined with EHDS resources)
  - myData
- Jamboard with comments (link)

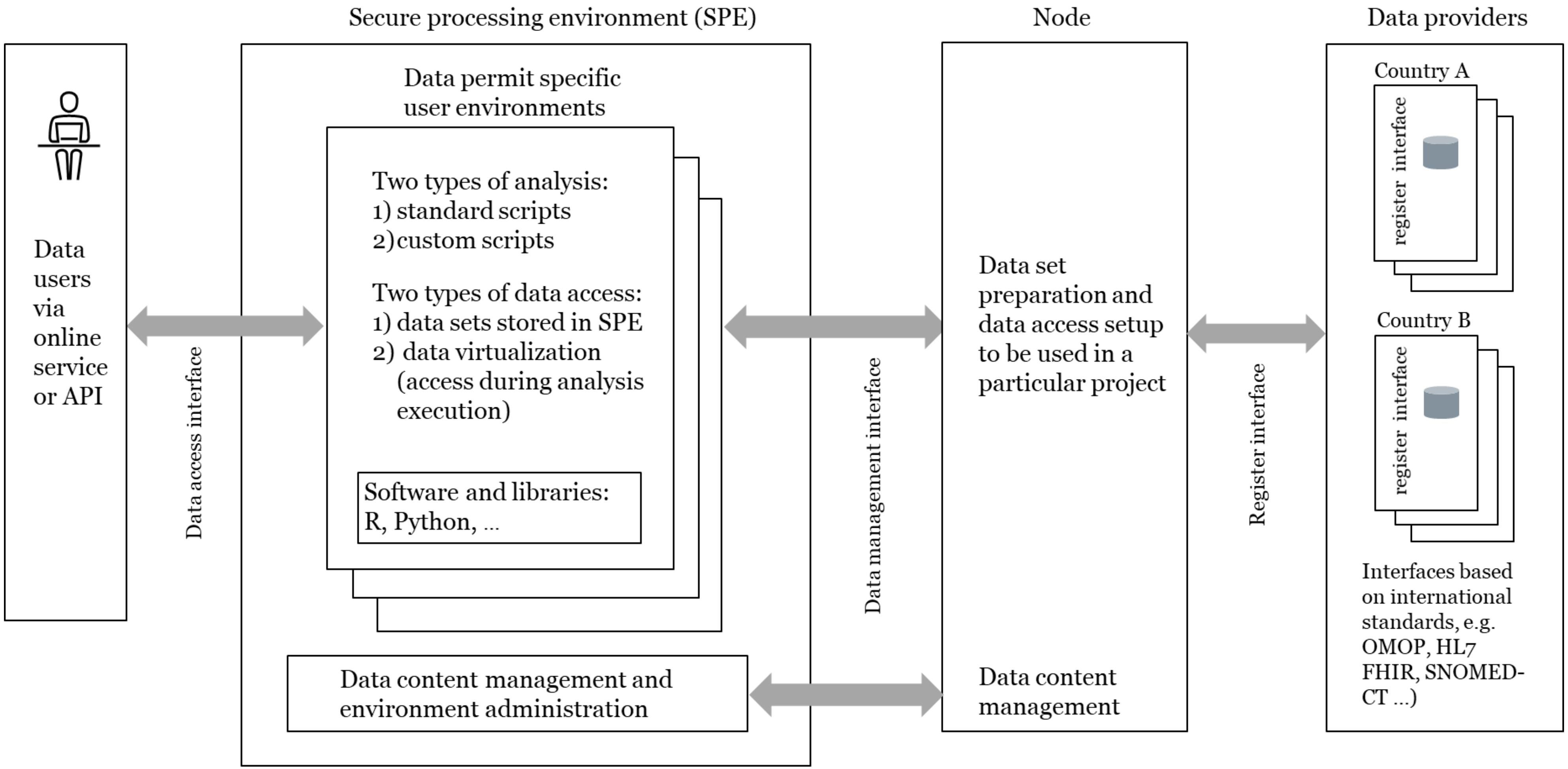
Co-funded by
the Health Programme
of the European Union

Architecture and interfaces under discussion in workshop 4



Co-funded by
the Health Programme
of the European Union

Secure processing environment (SPE) under discussion workshop 4



Secure processing environment (SPE)

Node

Data providers

Data users via online service or API

Data access interface

Data permit specific user environments

Two types of analysis:
1) standard scripts
2) custom scripts

Two types of data access:
1) data sets stored in SPE
2) data virtualization (access during analysis execution)

Software and libraries: R, Python, …

Data content management and environment administration

Data management interface

Data set preparation and data access setup to be used in a particular project

Data content management

Register interface

Country A

Country B

Interfaces based on international standards, e.g. OMOP, HL7 FHIR, SNOMED-CT …)

Co-funded by the Health Programme of the European Union

WS4 Conclusions (1)

Legal and governance related

- The EHDS should include a common approach for the legal basis for data processing and consent (when it is need and what kind of consent is needed)
- As this is not yet known (for the time being) the architecture shall take into account different alternatives and should be adaptable to different approaches of legal data processing basis and consent.

Secure processing environments

- The presented approach was considered feasible in general. The common opinion is that individual-level data should be processed in the SPE and not exported from there. Common requirements and standards for an SPE should be agreed and any organization compliant with them, should be able to provide the SPE.
- It was stressed that a good trade-off between high security and smooth access for data user should be achieved.

Co-funded by
The Health Programme
of the European Union

WS4 Conclusions (2)

Architecture

- The presented architecture interfaces were considered relevant. However, there are different opinions on the possibility to transfer individual-level data across borders or even within countries.

Terminology

- The terminology used in the architecture shall follow the Data Governance Act were applicable.

Nature and amount of nodes

- The architecture should be essentially decentralized with the idea that at least one node would be in every country and that each node should be empowered to maintain the “standard” user journey and provide the related services.
- The need for some centralized services was however recognized.

Co-funded by
The Health Programme
of the European Union

WS4 Conclusions (3)

Data consumer's own data (usage combined with EHDS resources)

- This kind of linking should be enabled by EHDS, but the risk for re-identification of data subjects is considerable and needs to be taken into account

My Data and clinical trial data

- It was concluded that data donated by individuals (e.g. for research) can be an important component. There were different opinions on the implementation: (1) whether it should be a function of the EHDS directly to the donor or (2) whether the actual donation is taken care outside EHDS and brought in to EHDS scope by the corresponding data controller. Similarly, different opinions on handling clinical trial data were expressed.

Co-funded by
The Health Programme
of the European Union

Workshop 5

- Invited presentation of eDelivery solution for cross-border data sharing given by Bogdan Dumitriu (DG DIGIT, European Commission)

- Architecture approaches for comments:
  - centralized services added to the overall architecture
  - renaming of some architecture elements to be aligned with Data Governance Act
  - approaches for reducing privacy risks in cross-border setting (three policy scenarios concerning data transfers and two options concerning data access)
  - privacy-preserving techniques and their applicability to policy scenarios

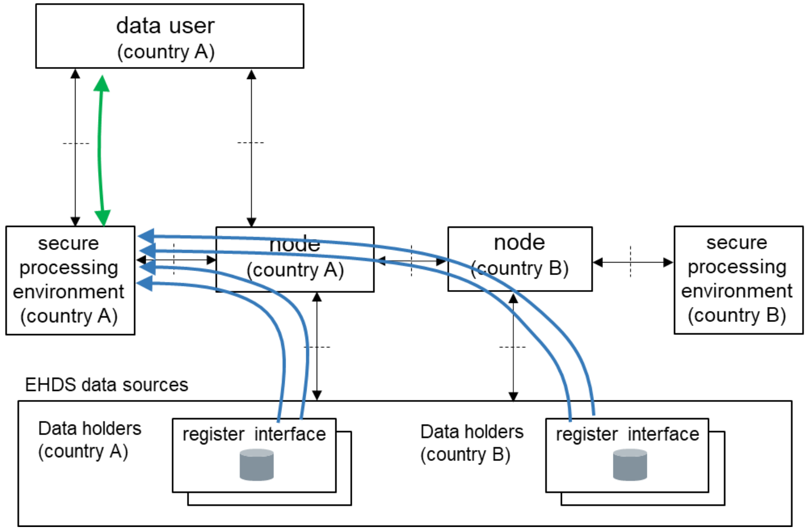
- Additionally discussed topics:
  - anonymization/pseudonymization
  - SPE security

- Jamboard with comments (link)

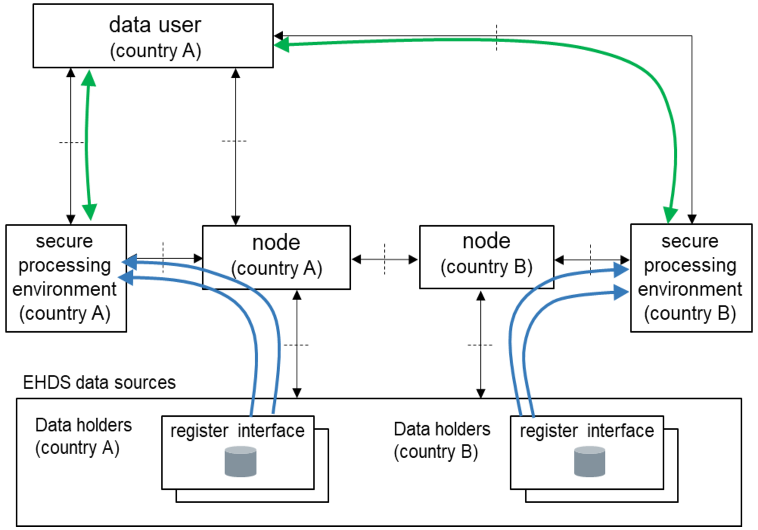
Co-funded by
the Health Programme
of the European Union

Policy scenarios under discussion in workshop 5

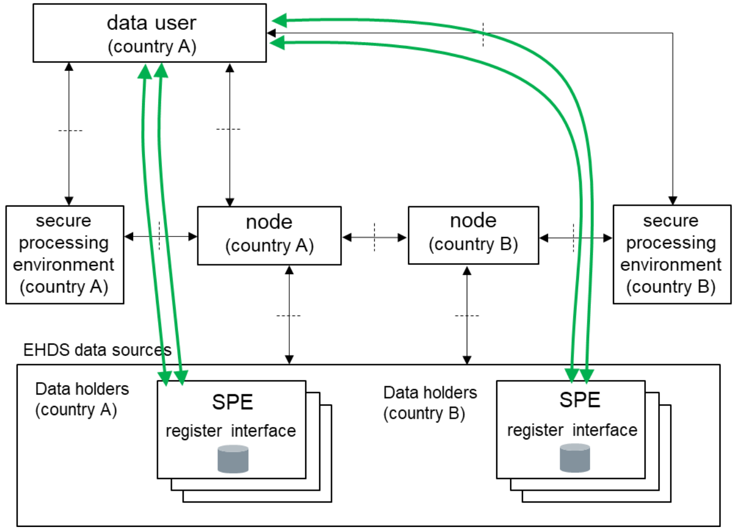
Data transfer across borders is accepted



Data transfer is only accepted within country



No data transfer out from original data holder is allowed



🟩 ⟷ access to individual-level data

🟦 ⟷ transfers of data managed by the node

Privacy-preserving techniques under discussion in workshop 5

Technique	Description
View-only remote desktop	Data user can see the data on screen, but cannot download the data.
Anonymization	Data user can freely process the data (incl. download). The data is pseudo-anonymized by masking, generalisation, swapping, perturbation or some other means.
Synthetic data	Data user can freely process the data (incl. download). The data is artificial data generated based on the original data set.
Custom scripts	Data user uploads own analysis scripts (programs, algorithms, …) and executes them without access to data.
Standard “certified” scripts	Data user selects among ”certified” scripts presinstalled in the SPE and executes them without access to data.
Binary closed source SW	Data user uploads and runs binary closed analysis software on data without direct access to data.
Data virtualization	Data is dynamically retrieved to SPE for the duration of algorithm execution. Data is not stored in SPE and the data user does not have access to data.
Federated learning	Machine learning model is trained across multiple SPE’s or registers. The data user gets the model parameters from the distributed compotation sites and can combine the results without direct access to data.
Secure multi-party computation	Cryptographic technique enabling parties (nodes or data holders) to jointly compute a function using data from all parties while keeping the data private.
Homomorphic encryption	Cryptographic technique enabling computations on encrypted data. Data holders may disclose data for the data user in encrypted form.

WS5 Conclusions (1)

Cross-border scenarios

- Discussion concerning if and how individual-level data can be transferred out from original location to be integrated with data of other data owners. There are different opinions on how data can be transferred.
- It was noted that under consent of data subject all scenarios are possible. However, consent is not always possible to be asked or may cause bias in many research cases.
- Several ways to minimize exposure of personal data have been identified (see previous slide). All approaches have merits, but there are still remaining questions to be solved.
- Federated learning, secure multiparty computation and homomorphic encryption are options enabling data to stay in its original location. The maturity of these technologies in specific use cases still needs to be demonstrated. Potential increase of complexity to the architecture may impact the applicability (e.g. an additional processing layers and needs by data owners to implement computing resources).

WS5 Conclusions (2)

- The approach of “standard scripts” may be a good approach to enable data processing without exposing data to the data user. However, it is unclear how the correct performance of the scripts and quality of data can be verified without direct access to data.
- Several anonymization techniques have been presented. More information is needed about their applicability in different use cases. When used, it is important to remember that 100% anonymization will be difficult to ensure. Differential privacy is an important technology for improving the "conventional" anonymization approaches especially to protect distributed data processing (e.g., federated learning).
- For all approaches hiding the individual level data, it will be challenge to report findings back to initiate appropriate healthcare interventions.

Co-funded by
The Health Programme
of the European Union

Workshop 6

- Invited presentation on privacy preserving technologies, in particular focusing in using differential privacy for protecting distributed computing (federated learning) – presentation given by Antti Honkela (University of Helsinki)
- Invited presentation of the PHIRI project focusing on a federated computing approach enabling the use of distributed data resources for population health, in particular COVID 19 - presentation given by Francisco Estupiñán-Romero (IACS)
- An introduction to the published WP7 deliverable 7.1. (Options for the minimum set of services for secondary use of health data in the EHDS) – presentation given by Juan Gonzalez (IACS)
- Discussion on the impact of the EU's legislative proposal on the EHDS architecture and services - groupwork led by Jaakko Lähteenmäki (VTT)
  - Jamboard with comments (link)

Co-funded by
the Health Programme
of the European Union

Discussion topics on the legislative proposal

- Purposes (Art 34(1))
- Health Data Access Bodies (Art 36(1))
- Obligations towards natural persons (Art 38)
- Deadlines for permit application and data delivery (Art 41 & 46)
- Data minimization (Art 44)
- Data access application (Art 45)
- Data permit (Art 46)
- Requirement to publish results (Art 46(11))
- Reporting clinical findings (Art 46(12))
- Access to data from a single data holder (Art 49)
- Secure Processing Environments (Art 50)
- Cross-border architecture (Art 52)
- Dataset description (Art 55-57).

Co-funded by
the Health Programme
of the European Union

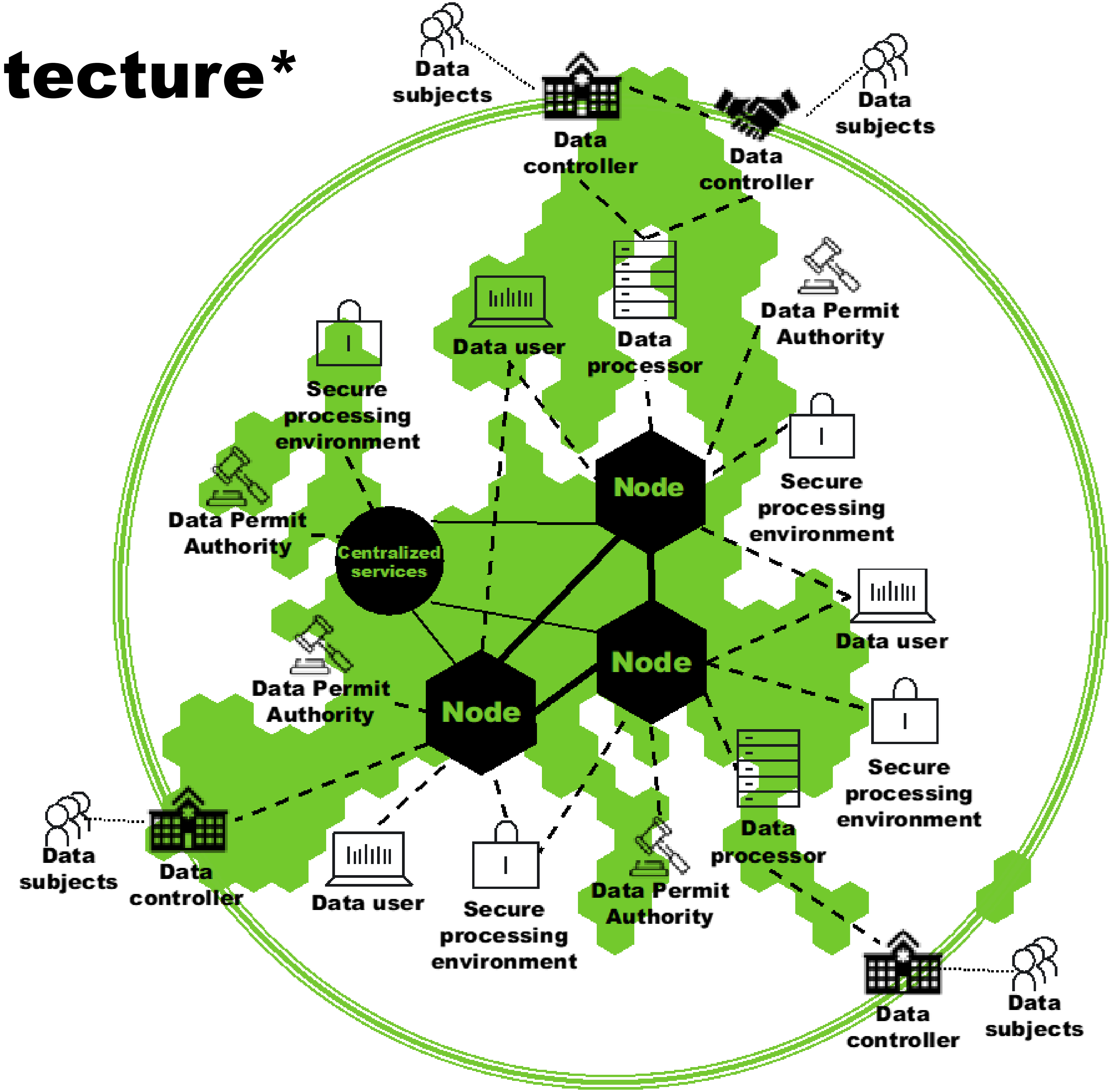
Summary of discussion

- In general, the overall architecture and services outlined in Deliverable 7.1. (see next pages) are in alignment with the EHDS legislative proposal.
- Terminology needs to be updated, e.g., ”node” → ”health data access body” (HDAB).
- HDAB has a large scope and high importance: represents both ”node” and ”permit authority”.
- The role of secure processing environments (SPE’s) is important. Further specifications to harmonize the approaches in different countries are needed.
- Obligations towards data subjects need clarification.
- The categorization of data into two groups: (1) anonymized, (2) pseudonymized remains unclear and need elaboration.
- Archival of data sets and results with the objective of results verification and data set enrichment are not adequately defined in the proposal.
- Mechanisms and requirements for reporting clinical findings need further clarification.
- The proposal supports data transfers to a local SPE in a different country or EU’s centralized SPE. On the other hand the proposal encourages the principle: “bring questions to data instead of moving data”. Further clarification and criteria for selecting the appropriate approach in a specific case is needed.

Co-funded by
the Health Programme
of the European Union

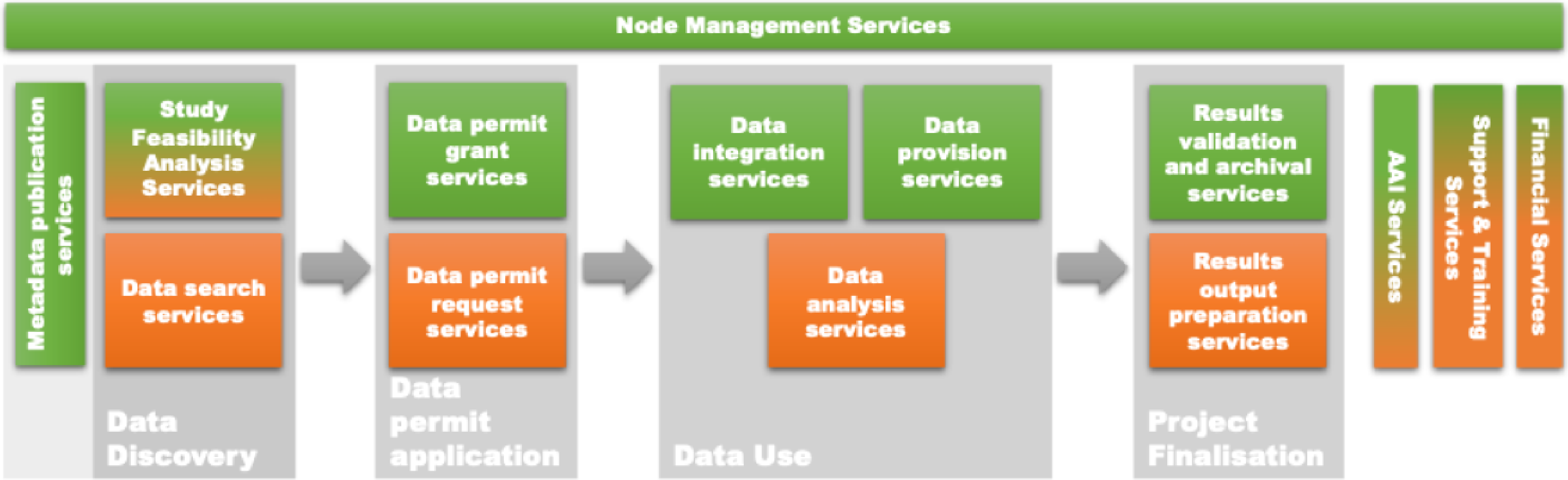
Overall architecture*

(Deliverable 7.1)



Co-funded by
The Health Programme
of the European Union

User journey (Deliverable 7.1)


Node Management Services
Metadata publication services
Study Feasibility Analysis Services
Data search services
Data Discovery
Data permit grant services
Data permit request services
Data permit application
Data integration services
Data provision services
Data analysis services
Data Use
Results validation and archival services
Results output preparation services
Project Finalisation
AAI Services
Support & Training Services
Financial Services

Co-funded by
the Health Programme
of the European Union

Summary

- Three workshops have been held to collect views on EHDS architecture and services
- The views on overall architecture and services are well reflected in the EHDS legislative proposal published in May 3
- Several points in the proposal still need clarification and will be worked on the in the remaining two workshops will be held in autumn 2022
- Final set of architecture and service options to be collected in deliverable D7.2

Co-funded by
the Health Programme
of the European Union

Thanks

- The WP7 task leader team wishes to thank all external experts and TEHDAS WP7 representatives for active participation and contributions in the workshops

Co-funded by
the Health Programme
of the European Union