DATASPACE 4HEALTH
LUXEMBOURG

INTERCONNECTING DIFFERENT DATA SOURCES

Document ID: DS4H_WP6_D6.3

Author

Dataspace4Health Consortium

Date: 08/07/2026

Internal

Funding

This project has received funding from the Ministry of the Economy of Grand Duchy of Luxembourg under Grant agreement no 20230505RDI170010392869.

Disclaimer

The documents published by the consortium are intended solely for informational purposes and reflect the views and opinions of the consortium members at the time of publication and within the scope of the Dataspace4Health project (DS4H).

The ideas expressed herein do not necessarily reflect the official policy or position of the funding entity.

Efforts have been made to achieve the relevance of the content; however, the consortium does not make any representation regarding its completeness and accuracy.

This document may be subject to further revision.

This document is intended to be published on the Dataspace4Health website.

CONTENTS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
API	Application Programming Interface
CE	Conformité Européenne
DS4H	Dataspace4Health
DT	Digital Twin
EHR	Electronic Health Record
FHIR	Fast Healthcare Interoperability Resources
GSPR	General Safety and Performance Requirements
HCP	Healthcare Professional
ISO	International Organization for Standardization
IVDR	In Vitro Diagnostic Regulation
NGS	Next Generation Sequencing
MDR	Medical Device Regulation
MTB	Molecular Tumor Board
NB	Notified Body
PII	Personal Identifiable Information
PoC	Proof of Concept
QMS	Quality Management System
SPE	Secure Processing Environment
TTP	Trusted Third Party
WP	Work Package

EXECUTIVE SUMMARY

1. **Purpose of the document:**
   This deliverable outlines data workflows for interconnecting health data sources within the Dataspace4Health (DS4H) initiative in Luxembourg.
2. **Key objectives:**
   - Define interoperable and secure data workflows across institutions.
   - Illustrate implementation through both a diabetes Digital Twin-enabled and Clinical AI Decision Support (DT) and Molecular Tumor Board (MTB) use cases.
   - Address technical, governance and where applicable regulatory requirements.
   - Explore deployment scenarios and pathways to clinical adoption.
3. **Strategic importance:**
   The document supports the development of a federated national health data ecosystem, enabling data-driven care, cross-institutional collaboration and alignment with data space initiatives.

1. INTRODUCTION

Healthcare data is often fragmented, limiting its use for clinical decision-making and innovation. DS4H aims to address this by enabling secure, federated and interoperable data sharing in Luxembourg.

This deliverable presents data workflows supporting two use cases: a DT for diabetes and an MTB for oncology care. It also considers deployment models and where applicable regulatory requirements, providing a foundation for moving from pilot solutions to implementation.

The concepts and implementation approaches presented in this deliverable are based on the DS4H reference architecture as described in Deliverable D3.1, “Dataspace4Health Reference Architecture”, which provides the architectural foundations for secure and interoperable healthcare data exchange across the DS4H ecosystem. Hence, please refer to Work Package (WP) 3 deliverable D3.1 for more details on the technical architecture.

2. DIABETES USE CASE

Diabetes is a chronic disease with a high prevalence and a significant risk of developing long-term complications such as nephropathy, retinopathy and cardiovascular. Predictive models and personalized interventions are essential to – in future – minimize risk of complications and thus ultimately aim to improve patient outcomes.

The DT concept enables a virtual representation of a patient, continuously updated with real-world data, allowing authorized Healthcare Professionals (HCPs) to simulate treatment scenarios and predict complication risks.

The Dataspace4Health (DS4H) initiative provides – amongst others – a federated, secure and interoperable data-sharing infrastructure based on GAIA-X principles, ensuring data sovereignty**,** GDPR compliance and trustworthy exchange between healthcare stakeholders.

This chapter will address how the DT decision support tool *could* be utilized in the DS4H context, where the model *could* be deployed and finally describe a high-level regulatory roadmap how the model *could* transition from a Proof of Concept (PoC) research project to a solution that can be used within the healthcare environment.

2.1 WORKFLOW OF THE DT CLINICAL AI DECISION SUPPORT SYSTEM

This subchapter will describe *a potential* workflow of how the DT *could* technically be utilized. It will address different prerequisites and describe how the DT can be initiated and used during a consultation.

2.1.1 PREREQUISITES

Different prerequisites that need to be met prior to being able to utilize the DT will be addressed here.

System integration

The DT model *shall* be integrated into the hospital’s Electronic Health Record (EHR) with a dedicated interface, e.g., via Fast Healthcare Interoperability Resources (FHIR)-based Application Programming Interfaces (APIs), which would ensure seamless access for the relevant and authorized HCPs. The DT module is thus envisioned to be accessible through the EHR of a patient.

The access to the DT shall be role-based, so that only authorized personnel can access it.

Patient consent

Each individual patient needs to provide explicit consent for usage of the DT during their consultation. Hence, this consent needs to be given prior to the actual consultation(s) with HCPs that intent to utilize the DT.

Please refer to WP 6 deliverable D6.2 for further information on the proposed consent management.

2.1.2 INITIATION AND USE OF DT DURING CONSULTATION

When the different prerequisites are met, to initiate the DT during the consultation, authorized HCP are thus envisioned to be able to open the patient's record in the EHR and select the DT module.

By doing so, the DT module shall perform different controls:

- Consent verification: Confirms patient consent for utilization of requested data.
- Access control: Validates HCP authorization to the DT platform.

Once all checks are validated, the DT model will update the risk predictions based on the available EHR data for the specific patient.

With the updated DT model, the authorized HCP *would* thus view the updated health state and risk predictions for the patient. Additionally, the DT *would* allow for simulations using "what-if" scenarios, e.g., "What if the patient improves diet by 20%?" or "What if medication is switched to SGLT2 inhibitor?".

The DT will run such simulations in real time using predictive algorithms and will display their outcomes accordingly.

The authorized HCP *would* then be able to use these insights to e.g.,:

- Adjust the treatment plan.
- Educate the patient on lifestyle changes.

All actions *would* be logged for auditability and compliance.

Please refer to WP6 deliverable D6.2 for further information on how, when and by who the DT can be utilized in the envisioned future process.

2.2 DEPLOYMENT OF THE MODEL

DT consists of two sub-modules:

1. The predictive simulation model
2. The conversational decision support AI agent

Either of the sub-modules of the DT tool may be deployed following different scenarios:

Scenario 1: Local level, hospital environment

Location: On-premise hospital servers or private cloud.

Advantages:

- Direct integration with EHR.
- Low latency for real-time consultations.

Challenges:

- Limited (no) scalability across institutions.
- Requires strong local IT governance.

Scenario 2: National level

Location: Secure, nation-wide-compliant cloud infrastructure.

Advantages:

- Federated access to multiple data sources.
- Centralized compliance and audit mechanisms.
- Supports multi-institutional use cases.

Challenges:

- Requires robust identity and consent management.
- Network dependency for real-time updates.
- May require data harmonisation between different data holders (due to the existing data structure within different hospital EHRs), with regard to required variables of interest for DT model.

While the predictive simulation and the AI agent sub-modules may be deployed independently, deployment across different scenarios would necessitate an additional local-to-cloud system integration.

Common key principles

Regardless of the deployment scenario, common key principles exist, namely:

- Federated data access: The DT does not store data permanently; it fetches data just-in-time via connectors.
- Data sovereignty: Patients retain control over their data through consent management.

2.3 HIGH-LEVEL: REGULATORY ROADMAP

This subchapter will – on a high level – describe the regulatory roadmap for the DT, how the current PoC research pilot *could* transition to a solution that can be used within a clinical setting.

There is a distinguishment between the regulatory roadmap under the Medical Device Regulation (MDR) or under the In Vitro Diagnostic Regulation (IVDR), which will be detailed below.

Furthermore, a clinically deployable DT will (likely) qualify as a high-risk Artificial Intelligence (AI) system under the AI Act, because it is intended for healthcare and diagnostic decision support. This imposes additional obligations, such as risk management, human oversight and robustness, which will be also further addressed in this chapter.

2.3.1 MDR

As the DT is intended to support diagnosis, prevention, monitoring, prediction, prognosis and/or treatment of a disease (namely: diabetes), it would need to be qualified as a Medical Device under Article 2(1) of MDR (EU 2017/745) [1].

In order to qualify as a Medical Device, the following shall be met/accomplished:

- Qualification & classification
  Software intended for medical purposes is regulated under the MDR. Due to the nature of the DT, the likely classification will be either Class IIa or IIb, per Rule 11 of Annex VIII [1].

- General Safety and Performance Requirements (GSPR)
  The DT will need to compliance with Annex 1: demonstrate clinical benefit, risk management as well as cybersecurity compliance [1].
- Conformity assessment
  Class IIa/IIb devices require engagement with a Notified Body (NB) to assess its conformity to the MDR requirements. Technical Documentation must be established accordingly and meet Annex II and III requirements [1].
- Clinical evaluation
  Required under Annex XIV, the safety and performance of the DT needs to be demonstrated using clinical data [1].
- Quality Management System (QMS)
  A QMS needs to be established which complies with International Organization for Standardization (ISO) 13485:2016 [2]. This includes implementing post-market surveillance and vigilance systems [1].
- Conformité Européenne (CE) marking & registration
  Finally, the DT must be CE-marked and registered in EUDAMED [3].

This can be further broken-down to the following steps:

1. Establish/confirm that the DT indeed is a Medical Device
2. Classify the DT
3. Implement a QMS
4. Prepare Technical Documentation
5. Appoint an EU Authorized Representative
6. Appoint a NB
7. Obtain the EU Certificate (after successful NB audit and subsequent Technical Documentation assessment)
8. Register in EUDAMED
9. Prepare a Declaration of Conformity
10. Affix the CE-mark

2.3.2 IVDR IN-HOUSE DEVELOPMENT

Another possible regulatory roadmap is under the IVDR (EU 2017/746). Because the DT will interpret and analyse in vitro diagnostic data (e.g., HbA1c, glucose levels) to provide diagnostic insights, the argument can be made to fall under the In-House Exemption (Article 5(5)) of the IVDR. Conditions for this are the following [4]:

- The tool needs to be developed and used within a single health institution
- The tool cannot be transferred to/used in another legal entity
- The tool shall be used under direct professional responsibility
- The tool shall be compliant with equivalent safety and performance standards

The following requirements are (still) applicable:

- Establish and maintain a QMS (not necessarily ISO 13485, but nonetheless robust), including implementing post-market surveillance and vigilance systems [4].
- Provide a public declaration of compliance.
- Ensure traceability and documentation of development and use.

Hence, the clear disadvantage of this regulatory roadmap is that the DT can only be utilized within a single institution.

Finally, there is also a possible regulatory pathway under the MDR for in-house development, based on Article 5(5) of Regulation (EU) 2017/745. This exemption also allows health institutions to manufacture and use medical devices internally without CE marking, provided strict conditions are met. However, this pathway is less common compared to the in-house development under the IVDR.

Although the MDCG guidance document *MDCG 2023-1: Guidance on Article 5(5) of MDR and IVDR* [5] applies to both regulations, practical adoption differs. In-house development under MDR lacks historical precedent for software-based medical devices such as a DT, whereas IVDR in-house exemptions are widely used for e.g., Laboratory Developed Tests in clinical laboratories, which have a long-standing tradition of internal test development and thus making compliance more practical and common.

The lower adoption of MDR in-house development for software is due to several factors:

- Complexity of compliance: MDR requires adherence to Annex I GSPR, which includes cybersecurity, usability and software lifecycle requirements and thus documentation similar to CE-marked devices, which are challenging for healthcare institutions to implement internally.
- No strong historical practice: Unlike under IVDR for e.g., Laboratory Developed Tests, healthcare institutions rarely develop software medical devices internally.

2.3.3 AI ACT

As a (likely) high-risk AI system, the DT must additionally comply with the requirements as outlined in the Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (the AI Act) [6].

These obligations include [6]:

- Risk management system
  Continuous identification, analysis and mitigation of risks throughout the lifecycle of the AI system (Article 9).
- Data governance and quality
  Ensure training, validation and testing datasets are relevant, representative, free of errors and complete to minimize bias (Article 10).
- Transparency and information to users
  Provide clear instructions for use, including the system's capabilities, limitations and intended purpose (Article 13).
- Human oversight
  Implement measures to ensure that the AI system can be effectively overseen by natural persons during operation (Article 14).
- Accuracy, robustness and cybersecurity
  Design and develop the AI system to achieve appropriate levels of accuracy, robustness and resilience against manipulation (Article 15).
- Logging and traceability
  Maintain automatic logs of system operations to enable traceability and auditing (Article 12).

3. ONCOLOGY USE CASE

Precision oncology depends on integrating clinical and genomic data to guide personalized treatment decisions. In Luxembourg, the MTB plays a central role in this process, but current workflows are manual and fragmented, limiting efficiency and scalability.

This chapter describes how to technically establish a digital MTB workflow within the DS4H context, enabling secure and federated data sharing. It outlines data extraction, data offering, data analysis and return of data insights.

Please refer to WP5 deliverable *Data Flow Diagrams for Oncology Case* for more details on the dynamics of the anticipated data flow. Additionally, please consult WP6 deliverable D6.2 for more details on the envisioned future process for the MTB.

3.1 DATA PREPARATION

Data extraction is the foundational step in enabling a digital MTB workflow. It involves gathering relevant patient information from multiple sources to create a comprehensive dataset for precision oncology. This process ensures that both clinical- and genomic data are standardized, pseudonymized (for secondary use of data) and prepared for secure integration within the DS4H ecosystem. By harmonizing data formats and applying privacy-preserving techniques, data extraction lays the groundwork for subsequent analysis and decision support in MTB sessions.

Data extraction

Data extraction involves collecting structured- and unstructured data from multiple sources, such as:

- Clinical data (hospital): Extract patient demographics, diagnosis, treatment history, etc. from the hospital's EHR. Standardization would be achieved by using the OncoBox data model [7], which ensures interoperability and quality assurance for oncological data.
- Genomic data: Next Generation Sequencing (NGS) generates raw (unstructured) sequencing data files. Bioinformatic workflows generate alignment files and variant call files. These files may be stored in Scality RING object storage [8].

Pseudonymisation

For secondary use, Personal Identifiable Information (PII) shall be replaced with a pseudonym. The data holders may pseudonymize the data internally or optionally use the national pseudonymization service provided by the Trusted Third Party (TTP).

3.2 DATA OFFERING

Once the data is prepared, it can be published by the data holder as a data offering in the DS4H catalogue:

- Its metadata will describe the data type (clinical or genomic), access conditions as well as technical standards.
- Access control: Role-based permissions for data users within the DS4H ecosystem.
- Federated access: For secondary use, data is not transferred to the data user, but remains in a Secure Processing Environment (SPE) [9] for a limited duration as defined in a digital contract established between the data holder and the data user respectively. The data user queries data through DS4H connectors without centralizing sensitive data.

Note: Approval of the data request is out of scope for this deliverable and will not be further described.

3.3 DATA ANALYSIS

Once the data can be accessed, the data analysis is conducted in three steps:

1. Annotation of genetic and other molecular data:
    a. Evaluating and curating the identified genetic variants using various tools, such as annovar or Variant Effect Predictor [10].
    b. Interpreting the results by comparing to a population of patients.
2. Knowledge databases such as Clinical Interpretation of Variants in Cancer [11] or OncoKB can provide treatment recommendations on cancer-related genetic variants.
3. The results are curated and interpreted by dedicated and trained personnel.

3.4 RETURN OF DATA INSIGHTS

After data analysis, for primary care, a report including identified, actionable genetic variants identified and drug recommendations are returned to the MTB.

Ultimately, outcomes of the MTB discussion are returned to the treating physician in the hospital and documented in the EHR.

4. CONCLUSION

This document demonstrates how DS4H enables secure and interoperable data workflows to support advanced healthcare use cases.

The proposed approaches demonstrate the importance of data standardization, governance and regulatory compliance for real-world adoption.

5. LITERATURE

1. European Parliament and Council. Medical Device Regulation (MDR), Regulation (EU) 2017/745. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745
2. International Organization for Standardization. ISO 13485:2016 – Medical devices – Quality management systems. https://www.iso.org/standard/59752.html
3. European Commission. EUDAMED – European Database on Medical Devices. https://health.ec.europa.eu/system/files/2023-01/md_eudamed-overview_en_0.pdf
4. European Parliament and Council. In Vitro Diagnostic Regulation (IVDR), Regulation (EU) 2017/746. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0746
5. Medical Device Coordination Group (MDCG). Guidance on Article 5(5) of Regulation (EU) 2017/745 (MDR) and Regulation (EU) 2017/746 (IVDR), MDCG 2023-1. https://health.ec.europa.eu/system/files/2023-01/mdcg_2023-1_en.pdf
6. European Commission. Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) COM/2021/206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206
7. XML-OncoBox Data Model. https://www.xml-oncobox.de/de/Home/FuerZentren
8. Scality Object Storage at LNS. https://itnation.lu/news/le-stockage-objet-de-scality-au-laboratoire-national-de-sante/
9. Luxembourg National Data Service (LNDS). Secure Processing Environment. https://www.lnds.lu/secure-processing-environment/
10. McLaren W, et al. The Ensembl Variant Effect Predictor. Genome Biology. 2016;17:1-4. https://www.ensembl.org/info/docs/tools/vep/index.html

11. Griffith M, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nature Genetics. 2017;49(2):170-174. https://civicdb.org