Call: HORIZON-HLTH-2021-TOOL-06
Topic: HORIZON-HLTH-2021-TOOL-06-03
Funding Scheme: HORIZON Research and Innovation Actions (RIA)

Grant Agreement no: 101057062

AIDAVA

AI powered Data Curation & Publishing Virtual Assistant

Deliverable No. 4.4

Information Governance Framework and Instruments

Approval by the European Commission Pending

Contractual Submission Date: 31/08/2023

Actual Submission Date: 01/09/2023

Responsible partner: P06: i~HD

Funded by
the European Union

Grant agreement no.	101057062
Project full title	AIDAVA - AI powered Data Curation & Publishing Virtual Assistant

Deliverable number	**D4.4**
Deliverable title	**Information Governance Framework**
Type[1]	R
Dissemination level[2]	PU
Work package number	WP4
Work package leader	P06 i~HD
Author(s)	Nathan Lea (i~HD), Dipak Kalra (i~HD) Isabelle de Zegher (b!loba)
Reviewer	Dominik Steiger (MIDATA) Petros Kalendralis (Maastro)
Keywords	Information Governance, Data Protection, Research Governance, GDPR, Consent, Information Leaflets, Participant Involvement, Regulatory Authority, AI Transparency and Regulation

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA).

Neither the European Union nor the granting authority can be held responsible for them.

Document History

Version	Date	Description
V0.1	21/07/2023	Initial draft for comment by co-authors
V0.5	01/08/2023	Updated Draft shared with T4.1 Collaborators
V1.0	01/09/2023	Final Version Candidate

[1] **Type**: Use one of the following codes (in consistence with the Description of the Action):

R:	Document, report (excluding the periodic and final reports)
DEM:	Demonstrator, pilot, prototype, plan designs
DEC:	Websites, patents filing, press & media actions, videos, etc.

[2] **Dissemination level**: Use one of the following codes (in consistence with the Description of the Action)

PU:	Public, fully open, e.g. web
SEN:	Sensitive, limited under conditions of the Grant Agreement

Table of Contents

List of Definitions

The definitions used in the deliverable are based on the AIDAVA Glossary [ref].

Executive Summary

Deliverable 4.4 describes the Information Governance Framework for AIDAVA. Information Governance relates to regulatory compliance and risk management for information handling. It will also inform technical design and implementation, including security services such as access controls and encryption.

Partner i~HD has therefore under Task 4.1 engaged with the Consortium to conduct the requisite information gathering and risk assessments to ensure high assurance around the handling of health information in line with key Information Governance principles. It has used the Data Protection by Design and Default approach provided by GDPR to engage with the Consortium early on to ensure that it defines the data flows to achieve the goals of AIDAVA, assesses the data protection, security and ethical risks of the project and defines the key instruments that will address them.

The outcome of this is the Data Protection Impact Assessment template for the Consortium, which in turn has assisted with the production of a Data Management Plan published as D4.1. Both deliverables are based on Data Flow Diagrams initiated during a dedicated workshop held in December 2022 between WP1 and WP4, and further refined through joint meetings held throughout the project. The processes have allowed an agreement on the roles of the partners and on the contractual agreements required to govern AIDAVA with progress made on defining these contracts. The contracts themselves include a set of bilateral Data Sharing Agreements developed on a standard template defined within the consortium, including - whenever applicable - existing legal provisions and specific technical provisions; Data Processing Agreements were assessed as not needed.

Task 4.1 has also offered advisory on submissions to Research Ethics Committees - for accessing patient data for annotation purposes and for assessing the prototype - and design choices for the project. The drafting of a Code of Practice is also underway and key challenges are being collated for submission to AIDAVA's independent Ethics Advisory Board which will meet for the first time in early October.

1 Introduction

Deliverable 4.4 describes the Information Governance framework for AIDAVA. Information Governance is a broad term that encapsulates all regulatory and best practice requirements for the safe, secure and legally sound handling of sensitive information that relates to private individuals. Specifically it includes adherence to the General Data Protection Regulation (GDPR) and its implementation across Member States within the European Union and Third States as well as other newer regulations including the Data Governance (DGA) and Artificial Intelligence (AI) Acts.

Information governance also relates to risk management for information handling. It will advise on design choices around information flows and where pseudonymisation or anonymisation is appropriate. It may refer to standards series such as ISO 27000 on information security management and it will also inform technical design and implementation, including security services such as access controls and encryption.

Information Governance cannot operate in isolation from the wider context of the activities it governs. In the case of research projects involving new technologies and interventions, Information Governance frameworks need to incorporate the wider regulatory expectations and requirements that relate to not only the safe design of the studies, but also the approach to recruiting participants. These requirements are enshrined in Research Governance Regulations that are in operation within any sovereign nation’s jurisdiction within and beyond EU borders. The primary instruments that support Research Governance from, at the very least the information handling aspects, are **Informed Participant Consent** to research and independent **Research Ethics / Review Panel approvals.**

The AIDAVA project presents a unique opportunity and challenge to the governance of information and wider research because it is incorporating within the bounds of traditional interventional studies two new paradigms. Firstly, the role of (**Health) Data Intermediaries (HDI)** that have been established under the Data Governance Act and offer the citizen the opportunity to receive, manage and share their own personal information including their health data. Secondly, the use of **Artificial Intelligence (AI)** in novel contexts, i.e. using AI to generate the data sets supporting AI. Indeed the AIDAVA project is developing an intelligent assistant targeted at citizens, and using Natural Language Processing (NLP), Machine Learning (ML) and Deep Learning (DL)and other AI technologies, such as language generation and speech recognition, to maximise curation of unstructured and semi-structured data. The virtual assistant will generate an interoperable and reusable patient longitudinal medical record, in the form of a knowledge graph, called a Personal Health Knowledge Graph (PHKG). From these PHKGs, secondary data sets can be seamlessly derived and processed most effectively by AI programs for research purposes, including to support decision making for patients and practitioners, risk modelling, disease modelling and prediction for clinical researchers.

The AIDAVA Consortium has worked to understand how these newer paradigms will (positively) disrupt the standard model for conducting research with live patients and their clinical carers. This has inspired the basis upon which Task 4.1 has liaised and engaged with the rest of the Consortium to apply the standard data protection and research governance risk management practices and deploy the best framework and instruments to meet regulatory compliance, protect participants and their data, and offer full transparency to the consortium and the wider public around how these projects are handled. To achieve this, Partner i~HD has under Task 4.1 engaged with the Consortium to conduct the requisite information gathering and risk assessments to ensure high assurance around the handling of health information in line with key Information Governance principles. It has used the Data Protection by Design and Default approach provided by GDPR to engage with the Consortium early on to ensure that they define the data flows to achieve the goals of AIDAVA, assess the data protection, security and ethical risks of the project and define the key instruments that will address them.

The AIDAVA Consortium has also chosen to assess the likely impacts of the forthcoming AI Act given that it will have implications on the use of AI throughout the lifetime of the project. The Consortium is also taking a view on the requirements for Medical Device Certification under the Medical Device Regulation (MDR) from an early stage in its development pending any future need to apply for certification as a Medical Device.

This deliverable describes the approach taken and activities that have contributed to the development of the framework and the results of the activities. It concludes with reflections and further work moving forward with the oversight of AIDAVA and its activities.

2 Description of Activities

The activities required to successfully deliver an effective Information Governance framework for the AIDAVA project focus on compliance with GDPR and other regulations applicable to health data and research. This includes understanding the data flows and identifying key features for limiting personal data use, i.e. through specifying requirements for anonymisation and pseudonymisation.

A cornerstone of achieving these activities is the conduct of the Data Protection Impact Assessment (DPIA) and liaising with the AIDAVA partners on the specifics of the data flows. A preliminary step was the development of the Data Management Plan (DMP) published under D4.1 where an initial specification of data protection requirements was provided after working with Task 1.1 colleagues on examining the necessary data flows from a data protection perspective.

2.1 Approach

The approach taken followed the GDPR specification of Data Protection by Design and Default. This requires that Data Controllers establish the implications for individuals' personal data early on in the development of a new system or data handling activity. In practical terms, this includes engagement with the key stakeholders involved in the new data processing, and culminates in the conduct of a DPIA if the new processing poses significant risk to the rights and freedoms of data subjects and to the data controller themselves. The DPIA and activities involved in completing it, help Data Controllers manage the Data Protection and wider governance risks, identify the key roles and responsibilities especially if other parties are involved.

In the case of large consortium projects such as AIDAVA the challenge is somewhat different in terms of achieving compliance and coherent governance. Consortium partners take on individual liability themselves as Data Controllers or Processors because the Consortium is not a legal entity. Individual partners must therefore consider their risks and conduct their own risk assessments themselves.

From a GDPR perspective, a Data Controller is a legal entity that decides on the mode and manner of processing, the reasons for processing and carries the weight of responsibility for the data protection. Controllers can be distinct or Joint i.e. Joint Controllers will mutually agree on the point, purpose and specifics of the processing. A Processor acts under the instruction of a Controller and is responsible for acting within compliance of GDPR to obey that instruction - whilst also being responsible for advising a Controller or Controllers if their instructions breach GDPR.

This poses a particular challenge because the data processing involved with AIDAVA are complex and conceptually unique involving several partners. AIDAVA is going one step further than traditional data driven interventional studies by involving the participants as actors in the research project and curators of their own data for their benefit. Further the understanding of the newer paradigms of AI and trusted Health Data Intermediaries in the research context needs careful consideration.

From a Data Protection perspective, it is essential to ensure all partners hold a consistent understanding of each others' roles and data processing needs to achieve AIDAVA's goals. The AIDAVA Consortium recognised that this requires engagement with the **developers** involved in the architecting and implementation of the software, with the **clinical practitioners and researchers**, and with

representatives of data subjects - i.e. patients which are represented through patient consultants[3] - who establish the requirements for the overarching solution, and finally with **third party application developers** interested to integrate the resulting medical device prototyped in AIDAVA, in their own solution.

Once the key stakeholders were identified, it was critical to get a sense of the flows of data and requisite processing across the proposed system architecture. We followed a classical swimlane approach to draft the data flows, which were then validated during a workshop in Tallinn in December 2022. This helped establish an understanding of what information is flowing where, what processing is involved and what software and hardware components are needed. Most importantly, it also helped define which stakeholders are involved with each step, and what responsibilities they hold in regulatory and contractual terms. Finally, it helped determine whether they are Joint or Distinct Data Controllers or Data Processors, or have only an advisory capacity and are not involved in the processing of data or hold an interest in the results of the processing directly or indirectly. All these aspects were taken into account while drafting the Information Governance Checklist, recommended to be used for establishing a DPIA in each organisation unless they want to use their own tool to deliver such DPIA.

The DPIA process has initiated the Data Flow Diagram development which in turn has helped to inform the IG checklist. It also informed the development of the first DMP and prospectively the strategy for anonymisation and pseudonymisation, and is supporting the implementation of Data Sharing and Processing Agreements, Codes of Conduct .

To support the Data Protection Compliance, a Record of Processing Activity (ROPA) has also been specified that defines the roles and responsibilities for each partner for each processing activity that forms part of AIDAVA and which will be filled out as the processing particulars are finalised in the next six months. The rest of this section describes the engagement, data flow development, roles and responsibilities assessments and the key instruments used to govern the AIDAVA project.

2.2 Consortium Engagement

The AIDAVA Consortium had to ensure that its Partners and their regulatory teams were aware of the requirements to govern the handling of information and that they would be able to fully engage. The purpose was twofold with regards engagement:

1. AIDAVA needed to establish the broad particulars of the data flows for the project and which partner was responsible for which aspect of the data handling;
2. AIDAVA needed to ensure that its Patient Association / Organisation Partners were fully involved and had the opportunity to observe and comment on the establishment of the governance framework and activities.

[3] Patient consultants are patients selected by the 2 patient organisations - based on predefined "inclusion/exclusion criteria'' - who agreed to work throughout the project as consultant/ advisors representing the views from patients. Each organisation selected 4 patient consultants; these patients are not expected to share any personal data

The broad principles and requirements for achieving the goals of T4.1 were therefore outlined at the Kick Off Meeting for AIDAVA in October 2022. It was agreed at that meeting that a dedicated workshop on data flows would be necessary, which occurred in December 2022. At this workshop a clearer understanding of the flows was established and a sense of the roles and responsibilities of each partner were also defined more closely. From this meeting an informal working group was established that included representation from the three medical centres, all HDIs, the research teams involved in development, clinical research and participant engagement, and the Patient Association Partners. The lead of Task 4.1 and the Clinical Coordinator led the group. The group met periodically and helped to define the DMP and contributed to updating the IG checklist . As the project developed, the meetings were held during routine meetings for AIDAVA.

A further engagement approach has been used to support the Consortium Partners. An additional governance requirement is that each relevant AIDAVA partner seeks and receives favourable reviews from its local Research Ethics Committees before patient data can be accessed and before patients can be requested to evaluate the intelligent virtual assistant developed during the project. Concretely, each medical centre and collaborating Health Data Intermediaries needs to submit to their local Ethics Committee a request for review for their participation in the work in AIDAVA, based on the common study assessment protocol developed at consortium level. Task 4.1 supported the consortium in checking data privacy and ethical aspects of this consortium wide study protocol, and will support each partner institution in submission to their local Ethics Committees while keeping a register of approvals processes and bodies along with the results of the reviews.

2.3 Data Flow Development

A key element of understanding the governance requirements was to map out the flow of data from sources (both clinical and collected by participants themselves) and where that data would flow within the Consortium both to individual partners and via the AIDAVA prototype (physical infrastructure and software) that supports processing of the data.

The main workshop was held in December 2022 and helped to gather traction on the details of the data flows themselves. It helped partners understand key processes around how the cataloguing of data would occur and the development of the steps to (semi) automatically generate the Personal Health Knowledge Graph - a semantic based representation of health data of a person, coming from multiple heterogeneous data sources, forming de facto an interoperable longitudinal record. Additionally, partners were able to more fully appreciate the differences around data processing, including the definition of the different outputs across the two use cases defined for the project (see Deliverable 1.1. Use Case description). This discussion also commenced the understanding of the relationships between partners with regards information governance and GDPR.

Establishing the basis of the data flows allowed for ongoing discourse of the following months with the working group. As the design decisions were made and the particulars of the data flows and access requirements were better understood, the DPIA could be better completed and the first DMP developed and published as D4.1.

2.4 Defining Roles and Responsibilities

During the discourse around data flows and the subsequent meetings, the definition of each partner's roles and responsibilities became clearer. This was important for ensuring that the DPIA could be adequately commenced and thereafter that the correct instruments were defined and developed (including Data Sharing Agreements, Codes of Conduct and risk assessments).

The process by which the roles and responsibilities have been defined included a review of each of the partner's contributions in terms of data, software development and data handling. It was reconciled with the nature of their contribution - i.e. whether they were defining the point and purpose of processing, or being instructed to process by another partner.

This reconciliation forms the basis for assigning the role of Data Controller or Data Processor for each partner. This in turn determines the contractual particulars that each partner will sign up to in terms of data transfer including the appropriate templates for Data Sharing, where needed Processing Agreements and how the particulars of processing are defined in the annexes of these agreements.

2.5 Key Instrument Development

In understanding the data flows and activities of each of the partners, the appropriate instruments can be prepared. The key instruments include the DMP, Information Governance Checklist for developing the DPIA, Data Sharing Agreements, Data Processing Agreements, any requisite licensing agreements and a Code of Conduct. All of these agreements will be read alongside the Consortium Agreement where they will refer to the particulars of the Consortium Agreement to be fully executed.

The **Data Management Plan (DMP)** broadly defines what information is needed, how it will be handled, what broad security requirements exist and how the data can be made Findable, Accessible, Interoperable and Reusable in line with the FAIR Principles. This DMP is a requirement for all EU funded projects and is usually completed by Month 6 and reviewed and updated in the fourth year of a project. The first draft of the DMP was submitted as D4.1.

The **Information Governance Checklist** is the core of this deliverable. It is important to note that this checklist is proposed as a recommendation to be used for establishing a DPIA in each organisation unless they want to use their own tool to deliver their DPIA.

A Record of Processing Activity (ROPA) defines the roles and responsibilities for each partner for each processing activity that forms part of the AIDAVA project and serves as the foundation to demonstrate how the Consortium achieves regulatory compliance.

Data Sharing and Processing Agreements are needed to govern the transfer of data between parties. **A Data Processing Agreement (DPA)** is required when a party is receiving instructions from a Data Controller or Controllers and acting as a processor. This agreement defines the particulars of processing and the expected behaviour of a Processor. Whilst none of the Consortium Partners are likely to take on the role of a Processor, they should all retain the option of appointing a Processor outside of the Consortium where this may be optimal or necessary to achieve AIDAVA's project goals.

A Data Sharing Agreement (DSA) is required, in the case of Controller to Controller transfers. It outlines the expected handling and security requirements between parties, the purposes data may be used for and establishing who is responsible for upholding data subject rights and liabilities and the technical specifications of the data to be transferred.

In the case of Joint Controllers, a Joint Controller Agreement (JCA) which is very similar to a Data Sharing Agreement is required that defines each of the roles and data handling for all of the parties involved in defining the purpose and particulars of the processing. The main difference between a JCA and a DSA is around the number of parties involved in the document signatures: a DSA tends to be bipartite, covering Controller to Controller processing, while a JCA is a multiparty agreement, covering processing across a number of parties who have mutually agreed the mode and manner of the processing and its purpose.

Code of Conduct and Policies defined the particulars for how every consortium partner will conduct themselves. These are operational materials that will be drafted by M24 of the project after the key agreements have been reviewed and a further development iteration is achieved. These will be developed and integrated with existing policies within partner practice in line with the processes defined in ISO 27001 and ISO 27002, particularly where partners have achieved or are in the process of achieving certification to ISO 27001.

3 Results & Discussion

3.1 Data Management Plan (DMP) and Data Protection Impact Assessment (DPIA)

Please refer to D4.1 for the DMP and Annex 1 for the content of Information Governance Checklist, to be used by each site for their DPIA. Note that both these instruments should be considered living documents where in the case of the DMP there will be an updated version provided by in the final year of AIDAVA. In the case of the DPIA Information Governance Checklist, it will be updated periodically at least every six months, or in the event of a significant change in processing or issue. Latest versions will be made available to all Consortium Partners.

The DPIA executed via the Information Governance Checklist represents the results of the activities outlined in Section 2; it includes checks related to the data flows, data processing, roles and responsibilities and compliance checks. Points to note in particular relate to the assignment of roles from the GDPR perspective, the development of the data management plan.

In short, the Information Governance Checklist summarises the main changes to data processing to support the AIDAVA Consortium, identifies the broad partners and articulates the framework for Controllers to identify the lawful basis for processing (Consent, Public Task or Legitimate Interest depending on the partner). It also confirmed the use of Special Category Personal Data (health, ethnicity amongst others) where the justification is consent or scientific research.

The compliance checks outline how the regulatory compliance requirements will be achieved and the broad quality and specification of the transparency notices - including information leaflets for participants. The IG Checklist also confirmed that whilst the processing as proposed in AIDAVA requires a full assessment check that has been completed, it does not pose so significant a risk to rights and freedoms of participants or to the parties involved with the processing that Supervisory Authorities need to be informed of these risks.

This process also provided the basis to define the key organisational involvement within the AIDAVA Project, including the patients who agree to curate their data, Healthcare Providers who hold data for care purposes, the Clinical Researchers and expert curators who develop and maintain clinical registries to perform their research, the HDIs who hold data on behalf of the Data Subjects participating in AIDAVA, the Research Institutions and Software Developers who are developing, deploying and testing the AIDAVA software prototype. This helped to dispel and address myths early in the project, including that AIDAVA was not a legal entity and that it was a consortium of legal entities whose GDPR roles and liabilities needed to be determined.

A specific discussion arose on the role of the Health Data Intermediaries. It made clear that the patients using AIDAVA - and participating in the evaluation of the prototype - were not themselves Data Controllers, rather Data Subjects who were participating in research and directing the HDIs to manage their data on their behalf. This imposed Controller or Processor roles on the partners who were offering the services to Data Subjects where the data would need to be governed in line with the consents and requirements of GDPR and Research Governance. An additional nuance highlighted during the DPIA discussion was ensuring that the data use would be expressly consented for when it goes beyond the scope of consent to participate in AIDAVA and wider services offered by the HDIs.

3.2 Data Flows

The key data flows of the AIDAVA prototype, deployed within the hospital environment, are described in Figure 1 below: yellow bullets and lines indicate external data flow, subject to Data Sharing or Data Processing Agreement, green bullets and lines indicate internal data flow not further discussed in this deliverable.

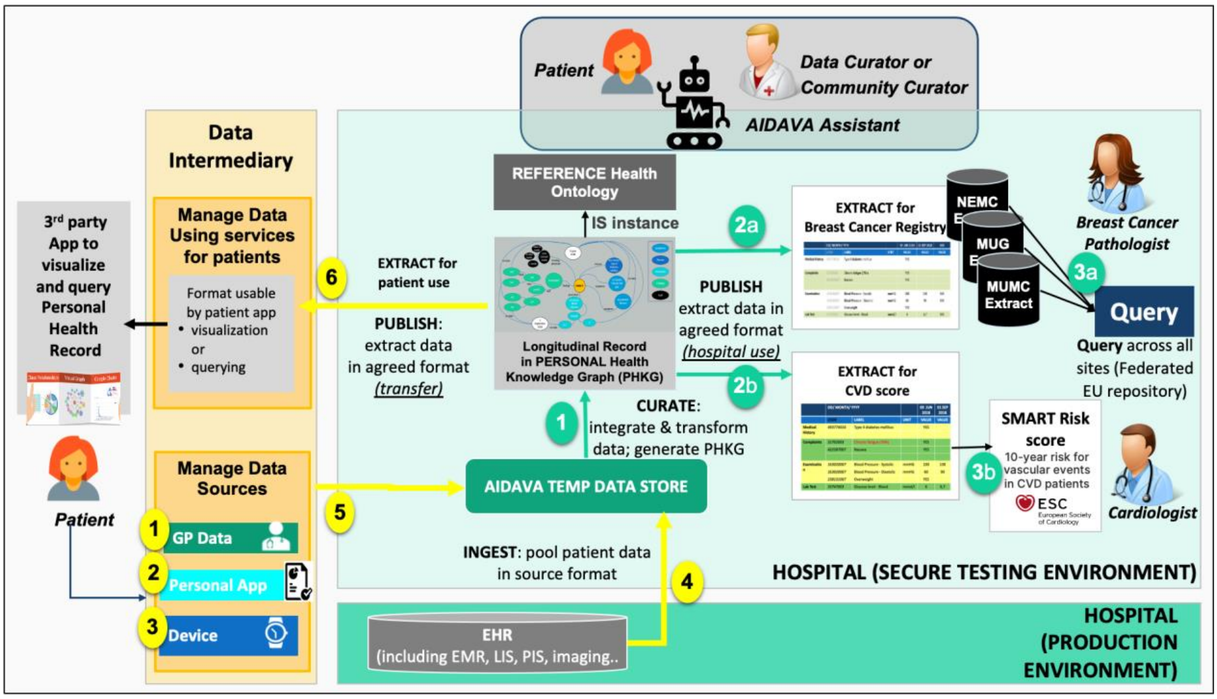

Patient
Data Curator or
Community Curator
AIDAVA Assistant
Data Intermediary
Manage Data Using services for patients
Format usable by patient app
• visualization
or
• querying
3rd party App to visualize and query Personal Health Record
Manage Data Sources
1 GP Data
2 Personal App
3 Device
Patient
6
EXTRACT for patient use
PUBLISH: extract data in agreed format (transfer)
REFERENCE Health Ontology
IS instance
Longitudinal Record in PERSONAL Health Knowledge Graph (PHKG)
2a
2b
PUBLISH extract data in agreed format (hospital use)
1
CURATE: integrate & transform data; generate PHKG
5
AIDAVA TEMP DATA STORE
INGEST: pool patient data in source format
4
EXTRACT for Breast Cancer Registry
NEMC
MUG
MUMC Extract
3a
Query
Breast Cancer Pathologist
Query across all sites (Federated EU repository)
EXTRACT for CVD score
SMART Risk score
10-year risk for vascular events in CVD patients
ESC
3b
Cardiologist
HOSPITAL (SECURE TESTING ENVIRONMENT)
EHR (including EMR, LIS, PIS, imaging..
HOSPITAL (PRODUCTION ENVIRONMENT)

Figure 1: High Level Data Flows for AIDAVA prototype

#	Provider	Recipient	Data Shared	Comments
1	Holder of GP data based on national practice	HDI	GP Data	● The Netherland: MedMij ● Estonia: extract from national system
2	Holder of data 3rd Party app (including Eq5D)	HDI	QLY data and other personal input	Need to confirm 3rd party app ASAP
3	Holder of medical device data (Withings BPM)	HDI	Blood Pressure	Only for CVD patients
4	Hospital (production environment)	AIDAVA	All patient data in scope of use case	May not require legal provisions, only legal provisions
5	HDI	AIDAVA	All data from HDI	
6	AIDAVA	HDI	Patient IPS in HL7	

The key elements for AIDAVA involve the sharing of data from clinical sources both in non-hospital environments (GP, Participants' own data that they collect through their apps and wearables), through

Health Data Intermediaries (estimated to be around 60% of health data) and Hospital environments. The collection from both Health Data Intermediaries and Hospitals will be in an AIDAVA temporary data store held by the Hospital partners. The data store is then used to curate the data and generate PHKGs, with the help of the Participants in the curation of their own data whenever relevant. The data from multiple PHKGs can be pseudonymised and where possible anonymised and provided to the Breast Cancer Registry for wider research use; the data from a single patient PHKG can also be used for the Cardiovascular Use Case risk prediction scoring. Finally the full PHKG can also be passed back to the Participant via the HDI for their own use.

Figure 3 below describes the Components of the data management within the AIDAVA solution that is provided to the Participant and their Expert Curators. Data can then be provided back to the HDIs for Participants to access and use their data as needed, where at each step in the data flows the data quality can be assessed as part of T4.2.

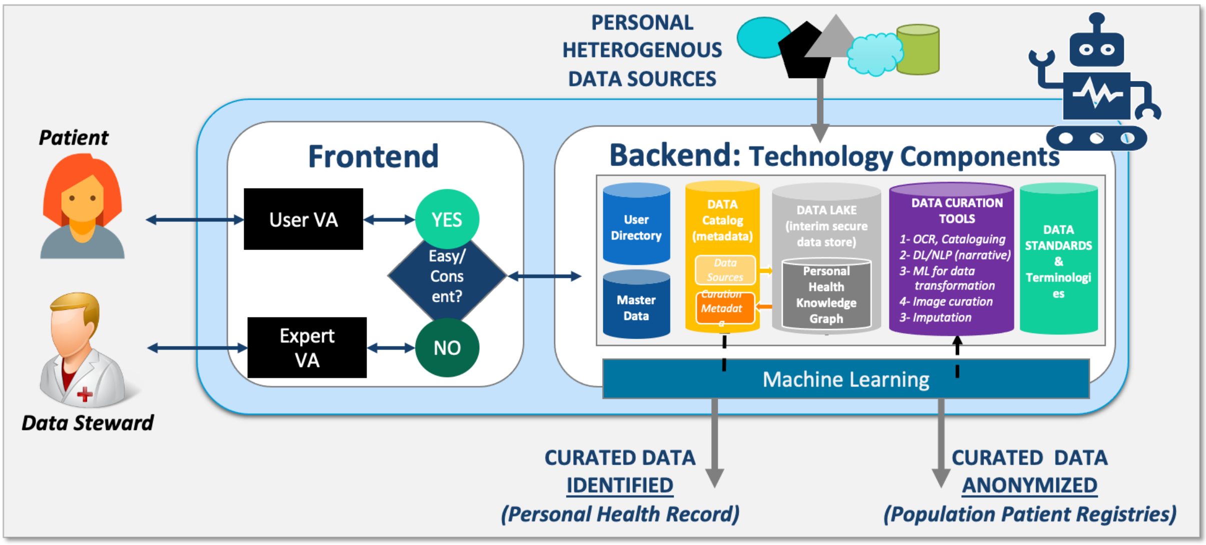


Figure 2: Overview of Backend Components that Power the AIDAVA Virtual Assistant, PHKG and Registry and Risk Scoring.

The data flows were described in detail through swimlanes and discussed and validated at the Tallinn workshop in December 20202. While exploring flows for transferring data within AIDAVA (from hospital and from HDI), as well as transferring data from AIDVA to HDI, we need to first onboarding the data i.e. agree on legal aspects, though Data Sharing/Processing Agreement, ensure agreement from the Ethical Committee through describing the intended data processing in an EC Approval form and finally describe the technical specification through a Data Transfer Technical Specification (attached to the Data Sharing Agreement) that allow to deliver a catalogue of the data sources to be further curated. This is described in Figures 4 and 6, respectively for Flows 4 and 5. The actual data transfer of patient data - with subsequent curation is detailed in Figures 5 and 7.

The symbols used on these different data flow is provided in Figure 3 below:

Process	Process to be supported by the virtual application
ELSI Process	Process to be managed (manually ?) by local Ethical Committee and Technical Committee
Curation Process	Process to be automated as much as possible with curation tool - **with audit trail**
Human in the loop	Interaction with user - Human in the Loop - through explainability module
Data Store	Data store with IDENTIFIABLE data . Most data managed by the AIDAVA VA are identifiable
Data Store	Data store with PSEUDONYMISED data
Data Store	Data store with ANONYMIZED data
Check	Decision point

Figure 3: Symbols used in the data flows described below

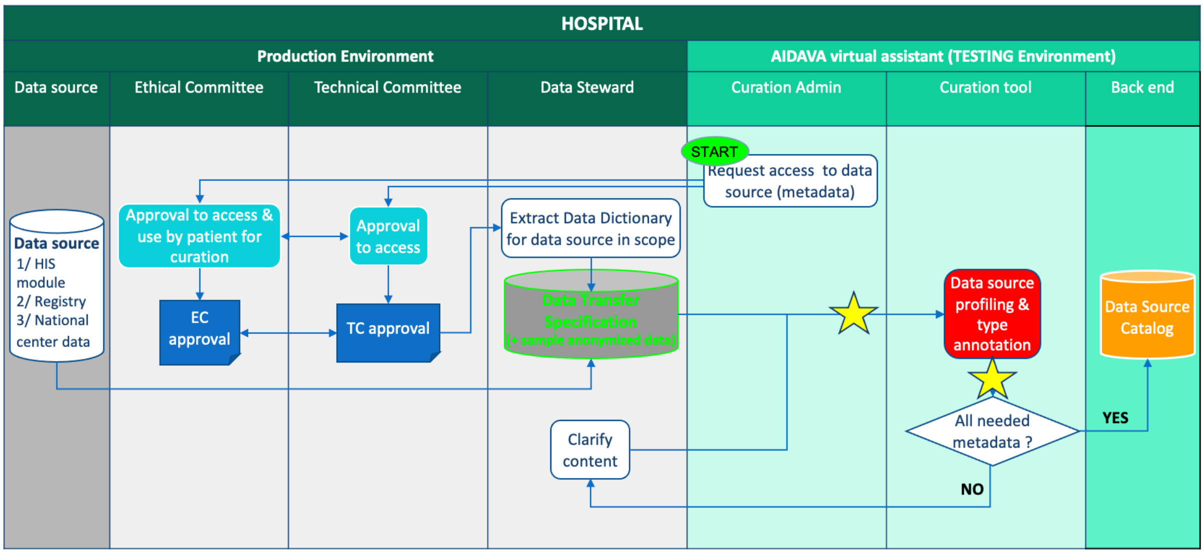
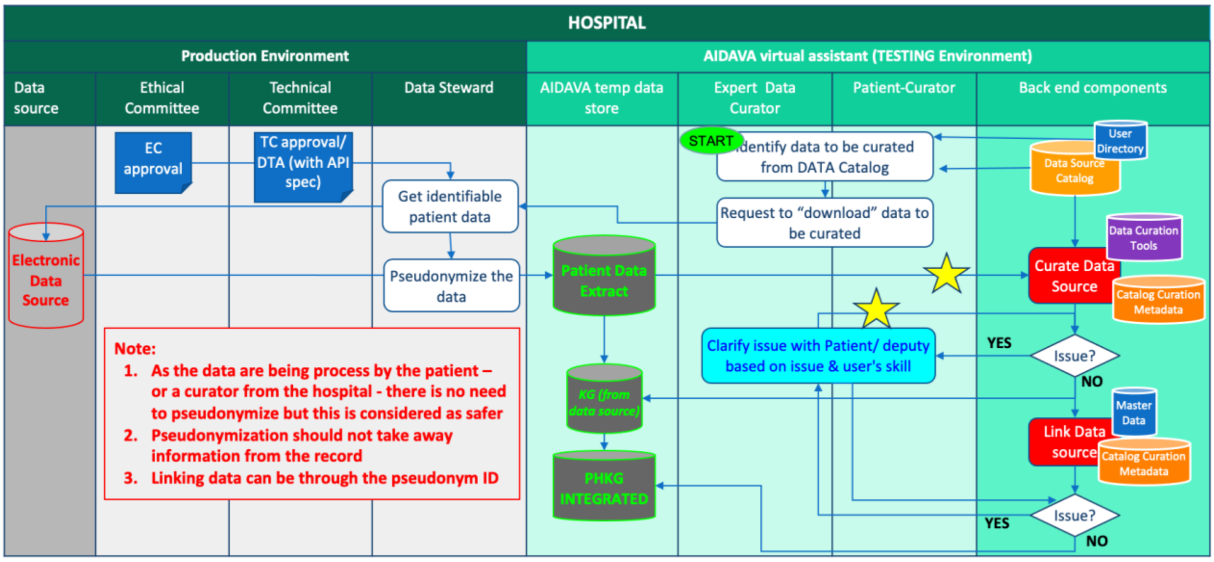


Figure 4: Detailed description of data onboarding for Flow 4 (Hospital to AIDAVA)



Note:
1. As the data are being process by the patient – or a curator from the hospital - there is no need to pseudonymize but this is considered as safer
2. Pseudonymization should not take away information from the record
3. Linking data can be through the pseudonym ID

Figure 5: Detailed description of data transfer - and curation - for Flow 4 (Hospital to AIDAVA

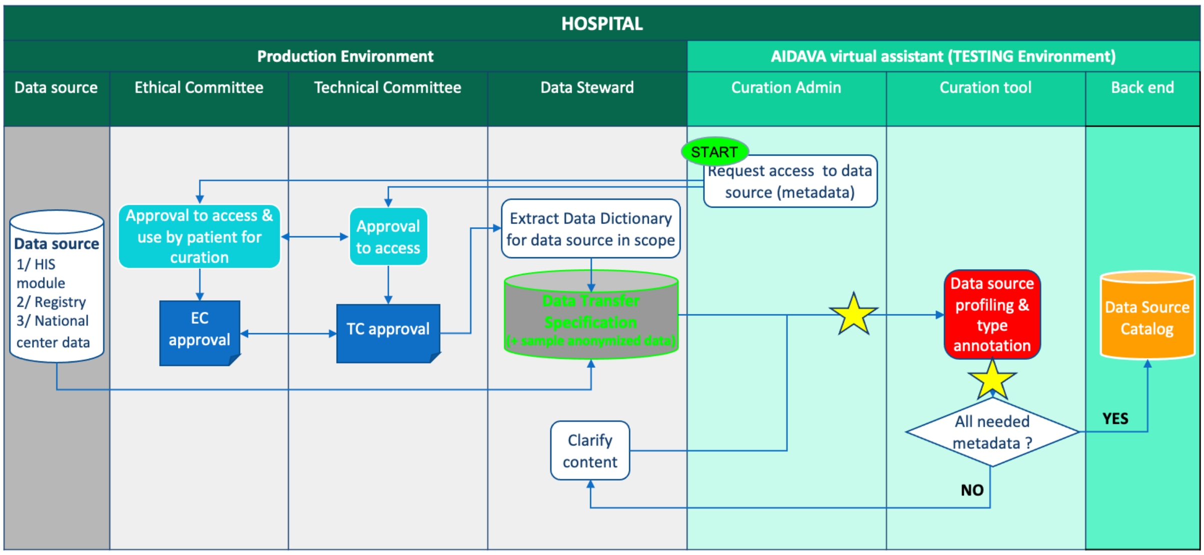
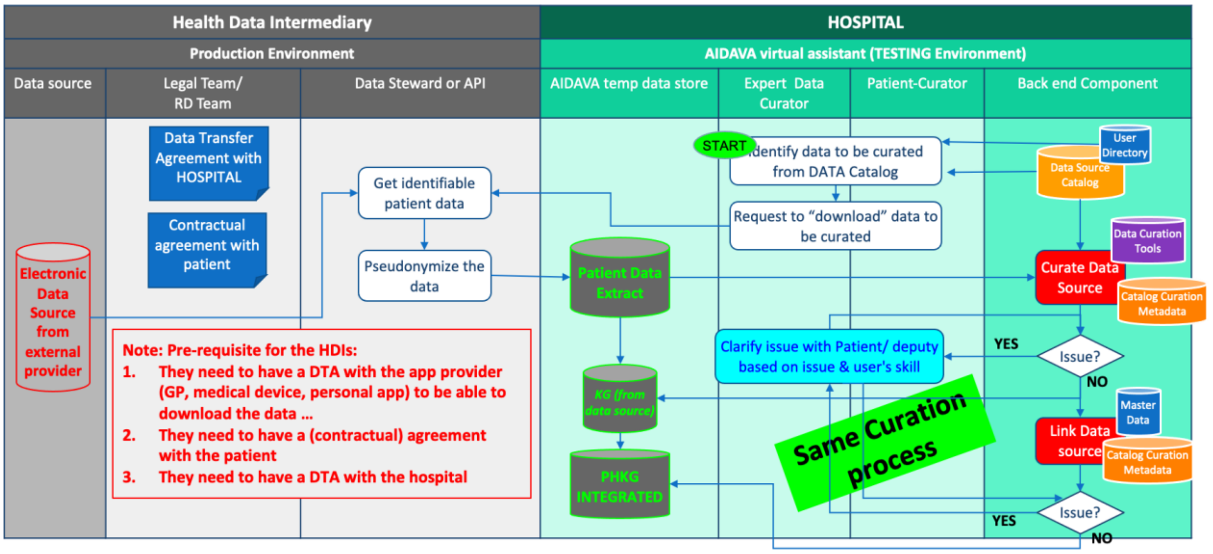


Figure 6: Detailed description of data onboarding for Flow 5 (HDI to AIDAVA)



Note: Pre-requisite for the HDIs:
1. They need to have a DTA with the app provider (GP, medical device, personal app) to be able to download the data …
2. They need to have a (contractual) agreement with the patient
3. They need to have a DTA with the hospital

Figure 7: Detailed description of data transfer - and curation - for Flow 5 (Hospital to AIDAVA

Key learning from the assessments of the data flows are that AIDAVA is a complex project with an ambitious delivery. The governance processes need to ensure that they clarify and support the data processing and not hinder it. This is why significant time has been spent on ensuring that the AIDAVA partners are in agreement and have a common understanding.

3.3 Record of Processing Activity (ROPA)

A ROPA is a framework that has been established by several Supervisory Authorities across Europe as a means for Data Controllers and Processors to log the personal data they are currently handling, identifying why they are doing so and specifying who in the organisation is responsible for them and what agreements apply to them.

These are useful records as organisations can rapidly review their liabilities and responsibilities for data that they are handling. This is an important record for AIDAVA as it will allow the Consortium to list the data that is involved with the data flows that have been identified and will allow the Consortium Partners to tag the Information Governance Frameworks that are deployed to each of the items that are being shared.

This will also be important to relate back to the AIDAVA Curation Metadata and Curation and Publishing Tools (as outlined in Deliverable 2.3 Figure 2) and serve as a basis to understand better the development of the data and its protection as the dictionary is agreed. Additionally this will be shared with Consortium partners so that they can update their own organisations' ROPAs as they see fit.

The ROPA will be completed within the next six months and will be finalised as the testing of the AIDAVA solution is designed and specified. The source template is based upon the Commission Nationale de l'Informatique et des Libertés (CNIL) framework where the template can be viewed in the supporting documentation.

3.4 Data Sharing Agreements (DSA)

The DPIA Information Governance checklist and discussions around the roles of the various partners provided a basis for determining how the handling of data should be controlled between partners in terms of contractual agreement. Large consortia need to assess whether partners are Data Controllers or Processors. Where they are Controllers, they are often identified as Joint Controllers since they have a common purpose in the processing of data, have a joint interest in the outcomes of processing and are agreeing amongst themselves the means and manner of processing for their shared purpose.

An alternative view is that Controllers are separate and distinct because they each define themselves the mode and manner of processing and whilst they are working towards a common goal, in doing so their processing are separately defined and pursuant to specific goals that a programme of work are examining jointly and severally.

In the case of AIDAVA the project is examining the potential of patient controlled health data, the ability to (semi) automatically improve the quality of these data stored within an interoperable format (the PHKG), and the possibility to generate seamlessly from multiple PHKGs secondary datasets supporting decision making and discovery. It is also examining how emerging Health Data Intermediaries can facilitate and improve this process.

To that end all parties considered the need to be able to operate as separate and distinct Controllers. For instance, the HDIs offer citizens, who consent for their services, to join a wider community around data handling amongst other services. Whilst not all of these features are required or mandated by AIDAVA itself, participants in the AIDAVA project will be offered the opportunity to use these other services. The DPIA process has helped to illustrate this point and define the need for additional consent processes outside of the AIDAVA consent process.

The need for distinct Controllers in AIDAVA is also due to the reliance on existing contracts and agreements between clinical partners to access and process data for aggregation purposes and wider research. In effect the processes that AIDAVA has used for defining the data flows have applied the requisite specification of agreements and contracts as per the DPIA. Specifically each HDI will require a bipartite Data Sharing Agreement with each of the clinical sites to receive and share back data as appropriate. These agreements will need to include provision for data to be shared with researcher partners under a separate agreement and for sharing between the Clinical Sites in line with their existing agreements. Given that existing agreements are in place between the Hospital partners, it is anticipated that only DSAs between the HDIs and the Hospitals will be needed.

Arguably a Joint Controller Agreement (JCA) could reduce the need for multiple bipartite agreements. The downside would be that it would be a complex agreement to account for the different flows and existing agreements within AIDAVA and would require negotiation with the different legal departments, which may take months. As the contract drafting is commenced, this consideration can be reviewed but the primary goals will be to ensure all the flows between partners are adequately covered by agreements.

To that end i~HD has been working with partners to assess the requirements for contractual arrangements. The result has been to confirm the set of bipartite agreements between partners - based on their current legal contracts and practices - to cover all levels of processing and to specify the exact mode, manner and limitations of data processing and sharing in line with the Assessment Study Protocol (Annex 1 to Deliverable 1.4 - Assessment) and the Informed Consent Form (Annexe 2 to Deliverable D1.4). A template Data Sharing Agreement, including legal and technical provisions is being agreed across the partners and will form Annex 4 to Deliverable 1.4.

The following agreements are being developed and will be completed within the last quarter of 2023; in some cases, they include only a Data Transfer Specification (DTS) as transfer is within the same entity (hospital to AIDAVA hosted in hospital) and legal/data privacy provisions are already in place; in other cases (HDI to hospital and vice versa) they require both legal and technical provisions. The existing agreements between the clinical sites will be reviewed on a per request basis from the sites themselves; results will be documented in the ROPA.

The AIDAVA Consortium does not have any partners who would act as a Data Processor but as discussed in Section 2, partners need to have the ability to appoint Processors that may be external to the Consortium where needed. To that end, each partner will apply their own existing templates and will inform other partners as needed in line with the DSAs that are in preparation. The Hospitals and

HDIs have agreed to develop bi-directional Data Sharing Agreements. It should also be noted that Hospital Sites may refer to University departments that are closely aligned with Hospitals themselves and have an existing sharing agreement in place (e.g. MUG and UMUC).

DSA per Data Flow (see Figure 1)	1. Hospital systems to AIDAVA in hospital	2. HDI to AIDAVA in hospital	3. AIDAVA in hospital to HDI
ESTONIA			
NEMC internal (Flow 4)	DTS only		
MIDATA to NEMC (Flow 5)		DSA; DTS	
NEMC to MIDATA (Flow 6)			DSA, IPS
AUSTRIA			
MUG internal (Flow 4)	DTS only		
MIDATA to MUG (Flow 5)		DSA; DTS	
MUG to MIDATA(Flow 6)			DSA; IPS
The Netherlands (2 clinical sites: Maastro for BCa nd UMUC for CVD)			
UMUC internal (Flow 4)	DTS only		
Maastro internal (Flow 4)	DTS only		
DME to UMUC (Flow 5)		DSA; DTS	
DME to Maastro (Flow 5)		DSA; DTS	
UMUC to DME (Flow 6)			DSA; DTS
Maastro to DME (Flow 6)			DSA; DTS

3.5 Other Considerations

3.5.1 Ethical Oversight and Advisory

The regulatory requirements relating to the project include the need for independent ethical review. AIDAVA is a complex initiative with multiple points of interaction with partners and participants. Given the number of partners involved, the presence of existing ethical approvals for some aspects of the work that required amendment, each section of the work would be handled under the appropriate ethics committee and would constitute a new application or an amendment to an existing approval.

In each case, Task 4.1 has provided support and guidance on the articulation of the particulars for each committee and will continue to keep a record of the advisory provided and approvals that have been sought along with the outcomes of the committees' reviews. It has been critical to relate the learning from the DPIA processes to ensure that the applications for approvals and associated consent forms are well developed and represent the proposed and actual handling of data.

3.5.2 Strategy for anonymization and pseudonymization

As per the existing ethical approvals and best practice, data used in AIDAVA will be anonymised where possible. The main area where this is applicable is for the use of clinical narratives used for annotation and delivery of the trained data set of the language model developed in Task 5.1.

For the actual prototype, data will be fully identified; indeed fully identifiable data is needed to support correct linkage of these data when integrating multiple data sources. Patients who contributed their data to the system will be duly informed about this and formally consent to it via the Informed Consent Form. However, whenever possible, the AIDAVA prototype will use a pseudonym of the patient but data should be considered as fully identifiable as some users of the systems, like the curators and treating physicians, will be directly in contact with the patient during the evaluation process.

3.5.3 Medical device certification

The AIDAVA will deliver a prototype of a "software as a medical device", not a product. While a prototype is not submitted to the Medical Device Regulation (MDR), the product will be. To avoid potential issues to ensure compliance in the product "to be", it is important to understand the requirements related to MDR certification at prototype level. While defining the business requirements, the project team therefore spelled out - as part of Task 1.3 - the needed requirements at prototype and product level . They are provided in the table below.

System's ability to comply/adhere with/to applicable laws, regulations, standards, and guidelines related to its operation and use. It is important because this ensures that the system operates within the boundaries set by regulatory authorities, and avoids legal or regulatory penalties or other consequences.

Name	*Description*	*Severity*	*Value (min-max)*	
			G1 & G2	*Product*
Compliance	*Extent to which the system complies with applicable laws, regulations, and industry standards related to security.* *Example: System should comply with standards such as PCI-DSS, HIPAA, or GDPR, MDR, ISO27001*	*blocking*	*GDPR (DPIA)*	*HIPAA, GDPR, MDR*
Compliance with Laws and Regulations	*System complies with applicable laws, regulations, and guidelines.* *Example: System should comply with specific regulations, such as data protection laws or safety regulations as well as quality/ safety requirements.*	*blocking*	*See compliance above*	*TBD (based on countries)*
Certification and Accreditation	*System has been certified or accredited by regulatory bodies or industry associations.* *Example: System should be certified or accredited by a specific organisation or agency.*	*minor*	*NTH*	*YES*
Auditability	*The system can be audited or reviewed to ensure compliance with regulatory requirements.* *Example: System should have built-in audit trails (who, what, when, why) or other features that allow for auditing or review.*	*blocking*	*YES*	*YES*

Name	*Description*	*Severity*	*Value (min-max)*	
			G1 & G2	*Product*
PWD	*Periodically check, recall and/or revise issuance of Identification code and password (e.g. Password expiration after 90 days).*	*major*	*NTH*	*YES*

3.5.4 Code of Conduct

In addition to the agreements a Consortium wide Code of Conduct is under development that will ensure that all partners adhere to the highest standards as the project continues. A first draft of the Code of Conduct will be available by M24 and it will include general principles of data management for each of the classes of data flow, each of the data components and their management and access, and each of the parties' key responsibilities in line with the research protocol and wider ethics approvals.

4 Conclusion

This Deliverable has described the key activities to develop an Information Governance Framework. Such a framework is founded on data protection and research governance regulations that exist within an evolving wider regulatory setting that is accounting for larger data processing through AI, but also for further empowerment of the citizenry across Europe.

Any framework requires a Data Management Plan, a Data Protection Impact Assessment (DPIA) - based on an Information Governance Checklist - and a Record of Processing Activity (ROPA) to demonstrate regulatory compliance, but also provides for the requisite transparency, accountability and demonstration of trustworthiness that projects such as AIDAVA must offer to the wider citizenry, especially as the use of AI increases and uses ever increasingly detailed health records to offer better services. This activity will be more tightly regulated by the forthcoming AI Act.

The digital health and wider innovation sectors must also bear in mind that EU regulations are moving more to empowering the citizens to take control of their own data. The Data Governance Act provides for the Data Intermediary which will allow citizens to receive copies of their data for their own uses, including health data. Empowering the citizens is at the forefront of AIDAVA. The involvement of Health Data Intermediaries is a powerful and vital contribution as it ot only allows AIDAVA to research the ramifications and impacts of these important services, but it is also a test case for understanding how best to integrate them in research projects that are tightly governed using approached that did not foresee either AI or Intermediaries necessarily. It is a goal of the AIDAVA Project to propose an effective governance model for HDIs so that they can work as fluidly and effectively as possible without compromising their core mission and founding principles.

In conclusion AIDAVA has presented challenges to developing a governance framework that can not only comprehensively address the evolving regulatory landscape, but also implement compliant instruments that are in line with current legislation that has been in place for some years along with new legislation that has little or no precedent and no real world evidence or experience for interoperating with other existing laws. Added to the advent of new regulations in the next eighteen months for AI, AIDAVA has its remit very clearly defined for governance and its activity.

The result of this Deliverable is intended to help provide a basis to navigate these trends and challenges. The details here should hopefully advise other innovation projects in how to address these challenges and ensure that data flows effectively for the benefit of the citizen and wider citizenry, with their satisfaction that it is being adequately protected and its use well explained so that they will continue to reap benefit and support these kinds of initiatives.

5 Next steps

The Governance Framework and its instruments will be reviewed, updated and implemented periodically in line with the needs of the project. The DMP will be updated in M36 and DPIA will be updated every six months, or where there may be a significant change to processing or issue that has arisen. Any requisite changes to other agreements and Codes of Practice as well as independent reviews and amendments will be addressed as and when needed.

The approach will require continued engagement with the consortium partners and involve updates to the various instruments that are described in this deliverable. The engagement points will be at regularly scheduled Consortium Plenaries and ad hoc online workshops.

As the development of the software and infrastructure proceeds, WP4 will continue to update the existing DPIA and DMP as well as develop a risk assessment in line with compliance and where appropriate certification to the ISO 27000 Series on information security management. This will help inform the deployment of policies and measures within the software and infrastructure components in line with risk management practices that conform to internationally recognised best practice.

6 Annexes

6.1 Information Governance Checklist (Data Protection Impact Assessment)

Call: HORIZON-HLTH-2021-TOOL-06
Topic: HORIZON-HLTH-2021-TOOL-06-03
Funding Scheme: HORIZON Research and Innovation Actions (RIA)

Grant Agreement no: 101057062

AI powered Data Curation & Publishing Virtual Assistant

Information Governance Assessment Checklist

Checking the need for a Data Protection Impact Assessment (DPIA)

Annex 1 to Deliverable D4.4

Responsible partner: P5-i˜HD

Funded by
the European Union

Grant agreement no.	101057062
Project full title	AIDAVA - AI powered Data Curation & Publishing Virtual Assistant

Deliverable number	**Annex to D4.4**
Deliverable title	**D4.4 Information Governance Instruments**
Type[4]	R
Dissemination level[5]	SEN
Work package number	4
Work package leader	P6-i˜HD
Author(s)	Nathan Lea (i˜HD) Isabelle de Zegher (b!lo) Dominik Steiger (MIDATA) Remzi Celebi (MU) Maria Christofidou (i~HD) Kerli Norak (NEMC)
Keywords	N/A

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA).

Neither the European Union nor the granting authority can be held responsible for them.

Document History

Version	Date	Description
V0.5	2022-11-28	First version for discussion
V1.0	30/06/2023	Publication Version for D4.4
Vx		

[4] **Type**: Use one of the following codes (in consistence with the Description of the Action):

R:	Document, report (excluding the periodic and final reports)
DEM:	Demonstrator, pilot, prototype, plan designs
DEC:	Websites, patents filing, press & media actions, videos, etc.

[5] **Dissemination level**: Use one of the following codes (in consistence with the Description of the Action)

PU:	Public, fully open, e.g. web
SEN:	Sensitive, limited under conditions of the Grant Agreement

Table of Contents

List of Abbreviations and definitions

The abbreviations and definitions used in the deliverable are based on the AIDAVA Glossary [ref]. Please note that the words *patient* and *participant* may be used interchangeably to refer to patients prior to and after they have been recruited to participate in the AIDAVA studies.

1. Introduction to Information Governance Assessment process

1.1 Principles

Under the General Data Protection Regulation (GDPR), a Data Protection Impact Assessment (DPIA) is only required where proposed data processing is "likely to result in a high risk to the rights and freedoms of natural persons" (Article 35(1)). However, Article 35(3) explicitly requires one where there is 'large-scale' processing of 'special category' (e.g. healthcare) data then a DPIA is required.

One other possibility is that the data being processed is already anonymised (see Recital 26) so falls outside GDPR altogether so that no DPIA is actually required.

However, good project management and Information Governance (IG) suggests that there should be a general approach to risk assessment for any project or business enterprise – if only to determine whether a DPIA might be required.

Ideally, one should work from a simple initial Checklist (this document) which identifies possible areas of information risk and compliance requirements to a 'discussion note' which explores any issues in more depth and may help identify the necessary mitigation methods and mechanisms to offset most if not all risks. **Only if risks are unmitigated or remain 'high' would you move to a formal DPIA report by completing this entire document.**

For the purposes of AIDAVA, the project will conduct a full DPIA. It is designed to be project wide and conducted collaboratively with all partners. The Consortium itself is not a legal entity, rather it is composed of multiple partners who are legal entities and have their own responsibilities to adhere to under GDPR and other regulations. In short, if they are sharing data they are holding already and / or will be processing data that they receive from other partners, each partner will need to conduct their own DPIA using their own templates where they are a Controller. It is their decision as to whether they run their internal DPIA and if they do not, they must be prepared to defend why they have declined to do so. In any event, this DPIA is being developed as a reference tool for all partners so that the project can provide a degree of consistency across any other DPIAs that partners may run, as well as ensure that there is a consistent understanding of the activities within AIDAVA across the partners themselves.

1.2 The Information Governance Assessment approach

Step 1. Information Governance Risk Assessment (responsibility of consortium)

There should be an overview of the proposed project or business change to explain what processing is envisaged as well as the purpose and intended outcome. The 'purpose' is important to establish the legal basis for the processing as well as ensuring that any possible mitigations or counter-measures do not undermine the main rationale for the processing.

The next step is to establish what compliance requirements may apply: GDPR, Data Governance Act, contractual or other regulatory restrictions, consent requirements, or obligations to preserve the data for legal or other reasons (including the benefit of posterity perhaps).

Once the precise range of obligations has been established, then appropriate checks can be made and recorded within the document.

Step 2. Decision on DPIA or nor and justify (responsibility: each partner)

The most obvious of these being GDPR compliance. There must be a 'High Risk' assessment (Appendix A) to determine whether the supervisory authority needs to be informed – generally, it is expected that it will not be necessary; if so, then a formal DPIA report will be needed.

Appendix B has a broader Privacy Impact Assessment that may throw up some broader issues.

Initial conclusions as to next steps or particular countermeasures to be considered should be detailed below.

Step 3. When DPIA is done - share any potential risks on which consortium can/should act and solve and demonstrate compliance at consortium level (responsibility: each partner)

Step 4. DPIA to be kept for auditing (internal, external if asked) to demonstrate compliance at organisation level - optional to share to the consortium (responsibility: each partner)

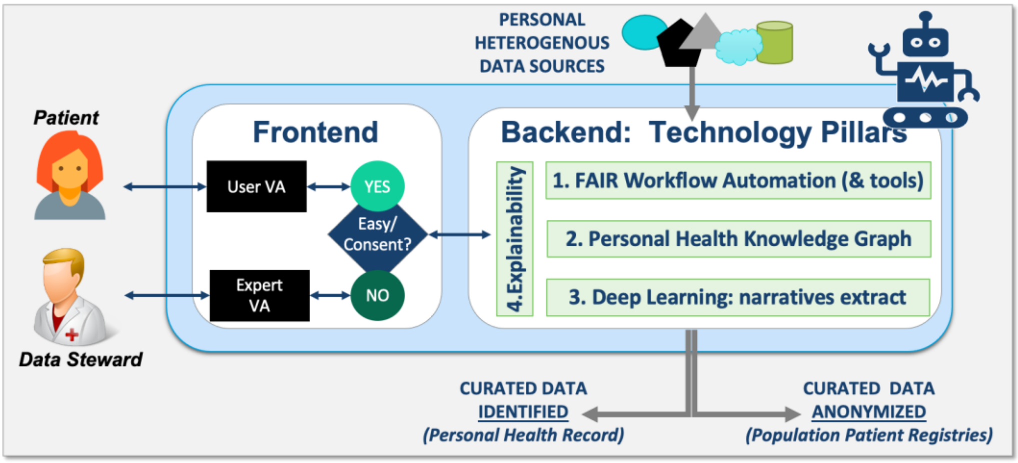
2. Project Background/Overview

2.1. Overview

Project Objectives

Integrated, high-quality personal health data (PHD) represents a potential wealth of knowledge for health care systems, but there is no reliable conduit for this data to become interoperable, AI-ready and reuse-ready at scale across institutions, at national and EU level.

AIDAVA will fill this gap by prototyping and testing an AI powered, virtual assistant maximising automation of data curation & publishing of computable knowledge derived from unstructured and structured, heterogeneous data. The assistant includes a backend with a library of AI-based data curation tools and a frontend based on human-AI interaction modules that will help users when automation is not possible, while adapting to users' preferences. The exceptional interdisciplinary team of the consortium will develop and test two versions of this virtual assistant with hospitals and emerging personal data intermediaries, around breast cancer patient registries and longitudinal health records for cardio-vascular patients, in three languages.



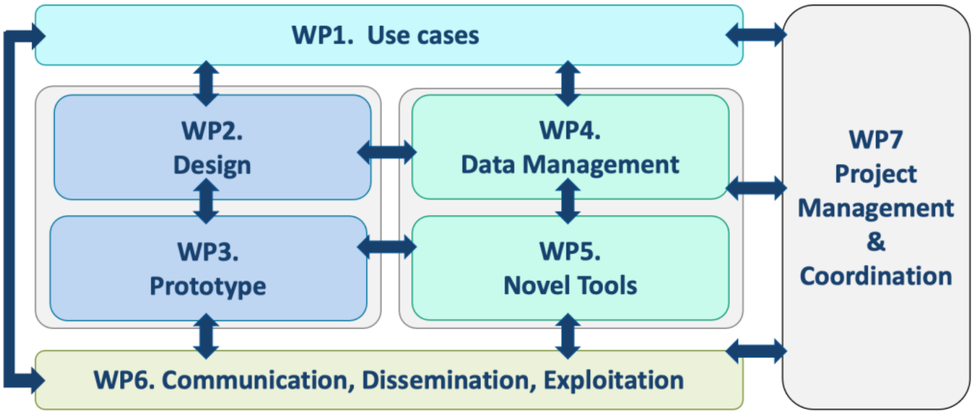
The team will work around four technology pillars:

1. automation of quality enhancement and FAIRification of collected health data, in compliance with EU data privacy;
2. knowledge graphs with ontology-based standards as universal representation, to increase interoperability and portability;
3. deep learning for information extraction from narrative content; and
4. AI-generated explanations during the process to increase users' confidence.

By increasing automation of data quality enhancement, AIDAVA will decrease the workload of clinical data stewards; by providing high quality data, AIDAVA will improve the effectiveness of clinical care and support clinical research. In the long-term, AIDAVA has the potential to democratise participation in data curation & publishing by citizens/patients leading to overall savings in health care costs (through disease prevention, early diagnosis, personalised medicine) and supporting delivery of the European Health Data Space.

Project Organisation

The project will work through a set of interrelated work packages as described in the figure below. We need the support of patients - and patient organisations - around Work Package 1 to ensure we take into account their needs and requirements.


WP1. Use cases
WP2. Design
WP3. Prototype
WP4. Data Management
WP5. Novel Tools
WP6. Communication, Dissemination, Exploitation
WP7 Project Management & Coordination

Partners

Participant	Participant organisation name	Principal Investigator
UM	UM Universiteit Maastricht	Michel Dumontier
b!lo	b!loba	Isabelle de Zegher
EUR	European Research and Project Office GmbH	Vera Schneider
KUL	Katholieke Universiteit Leuven	Marie-Francine Moens
ONTO	Sirma AI EAD	Todor Primov
IHD	The European Institute for Innovation through Health Data	Dipak Kalra
MUG	Medizinische Universität Graz	Andreas Holzinger
NEMC	Sihtasutus Pohja-Eesti Regionaalhaigla	Siiri Heinaru
GND	Gnome Design SRL	Béla Bihari
AVER	Averbis GmbH	Philipp Daumke
ECPC	European Cancer Patient Coalition	Antonella Cardone
EHN	European Heart Network	Birgit Berger
MID	MIDATA Genossenschaft	Dominik Steiger
DME	Digi.me Limited	Dan Bayley

2.2. Process steps

This allows identification of what processing is new or changed through the project:

Step	Current	Proposed
Project initiation, including any amendments or agreements for development as well as approvals	This is brand new processing so no existing processing exists	No change per se, but assessments are underway and the development of protocol, consent and information leaflets has started.
Generation 1 of the AIDAVA prototype which aids automation by triggering relevant data curation & publishing tools and, whenever required, request inputs from users (data stewards	No current processing	Generation1 (G1) will include the framework of the prototype composed of a frontend and a backend. The frontend includes human-AI interaction modules and will be based on an existing bot platform; the backend will contain a

Step	Current	Proposed
or citizens/patients) based on their level of computer/data literacy and medical knowledge.		library of data curation & publishing tools, based on off-the-shelf, preferably open source, tools.
Generation 2 of the AIDAVA prototype	No current processing	Generation2 (G2) will build on G1. The frontend will be expanded with an explainability module to increase usability for users less experienced in interacting with data and medical content; several curation and publishing tools will be replaced by novel tools developed in the project.

2.3. Organisations and roles

Organisation	Role	GDPR impact
Hospital - production environment (including clinical staff)	Data Controller - Use pseudonymized patient data to build ML model (e.g. NLP tool based on annotated text narrative) - Recruit patients - Provide identifiable data - from consenting patients - to AIDAVA - Use patient data generated by AIDAVA for clinical care (e.g. SMART score local) and clinical research (BC anonymized query across sites) - Collect patient data (e.g. health and computer literacy assessment, system usability scale to assess acceptance of the system..) => direct interest in patient data use	Data Sharing Agreement (with Data Transfer Specification) with Hospital AIDAVA *(internal; may not be needed at it may already be covered by existing agreements and/or it is for discovery with anonymized data)*
Hospital - AIDAVA environment (including expert curator and 1st level support)	Data Controller - Process and visualise identifiable patient data to generate PHKG. - Support patient to generate own PHKG by providing input on needed correct/mapping/updates - Extract identified and pseudonymized patient data for data use. - Transfer agreed data to data user, with consent of patient. - First level support of the users of the system, including the patients. => direct interest in accessing patient data (make them reusable)	- DSA with hospital – production (internal see above) - DSA with HDI - production (see below)

Organisation	Role	GDPR impact
HDI - production environment	Data Processor ● Process patient data (acquisition and display) acting on behalf of patient as "intermediary" ● Provide patient data to AIDAVA - with consent of the patient ● Get patient data (IPS and potentially the full PHKG) from AIDAVA for further use by the patient (not by the HDI !) => no interest to access patient data	Data Sharing Agreement (with Data Transfer Specification) with Hospital - AIDAVA ● From HDI to hospital which operates AIDAVA ● From hospital which operates AIDAVA to HDI
Developper (no site support)	Not applicable ● Develop and deliver software to AIDAVA environment ● No access, no interest to patient data	Not applicable - unless access to patient data is needed during development
Developper with site support (2nd and 3rd line)	Data Processor under instructions (exceptionally but not impossible) ● Second and third level support of the users of the system (on request of site 1rd level support) ; this may require access to identifiable data to be able to identify/debug issue	Data Processing Agreement to be signed between Hospital/ Medical University and Developer (who is also constrained by the Consortium Agreement)

3.4. Initial Conclusions

This section includes preliminary conclusions concerning further counter-measures or business viability [possibly tentative].

AIDAVA as an initiative represents a feasibility study to explore the use of an adaptive virtual assistant, which maximises automation in the curation of heterogeneous personal health data (PHD) by triggering relevant data curation & publishing tools and, whenever required, request inputs from users (data stewards or patients) based on their level of computer/data literacy and medical knowledge. Personal data from a patient originates from hospital records and from an Health Data Intermediary (HDI) mandated by the patient to manage his/her personal data such as GP records, personal apps and medical device data .

The study will be applied across both two use cases (Federated "EU"Breast Cancer registry and longitudinal health record of CVD patients) across the three clinical sites, supported each by a HDI. The precise details of inclusion and exclusion criteria, process for approaching and inviting potential participants to join the study, consenting procedures, rights to withdraw from the study and ongoing engagement will be specified in the AIDAVA assessment study protocol that will be used as a basis for conducting the research.

AIDAVA will only succeed if there is maximum agency granted to the participants via informed consent and rights to withdraw, and favourable opinions and advisory from the Research Ethics Committees. The main issues from a data protection perspective are currently around being clear on how much data should be included to ensure completeness of the patient personal health record, specifically whether there are record items referring to sexual health, mental health or other data items that are subject to special regulation. In any event, a list of expected record items must be added to the information leaflets as well as the DPIA, with a justification for their use or exclusion. This will include an explanation as to the need for all records to be included and form part of the informed consent process. The primary goal of AIDAVA is to put the patient in complete control of their data so the bounds of what may be possible or desirable will be assessed as part of the assessments.

Within the project, the roles of each of the partners have been identified (see previous section). While most partners are data processors, only hospitals (NEMC, UM, MUG)are data controllers with an interest in the outcomes of the processing and in using the data. (Note: the patient is not a data controller but has also a clear interest in the outcome of the processing).
For instance HDI (MID, DME) are "facilitators "i.e. they manage data on behalf of the patient, but they have no interest and no intention in using the data of the patient. This holds as well for the developers (UM, GND, ONTO, IHD for data quality rules) if/when they support the system, as well as organisations offering advisory services (i~HD, b!lo).
As a consequence it is suggested

- To confirm if the hospital has already in place a data sharing agreement to process the data internal through AIDAVA
- To sign a Data Processing Agreement - including data sharing/transfer- between each HDI and the relevant hospitals
- To sign a Data Processing Agreement - between each hospital and the developers (GND, UM-DACS)

.

3 Compliance Checks required:

Tick	Requirement	Notes [replace guide text with response]
	Does the project involve processing 'personal data' of any sort?	Yes, the project involves fully identifiable personal health data coming from multiple data sources; these data will be curated and inserted into the patient health record, in the form of a Personal Health Knowledge Graph (PHKG). Personal data will be processed WITH the consent of the patients following explanation on the objectives and type of processing of their data. Participants will be provided a Study Information Package (equivalent to a Patient Information Leaflet) and will be requested to sign Informed Consent Form. Data will be processed either by the patient or by a nominated curator from the hospital; all data processing will be strictly controlled by the study protocol (See Annex 1 to Deliverable 1.4) Participants will also be asked to fill out questionnaires as described in Section 3.3.
	Does the project involve processing 'confidential data' of any sort?	Yes as they belong to the patient health record and should therefore be included in the Personal Health Knowledge Graph (PHKG). See compliance safeguard explained above.
Data Availability requirements		
	Does data need to be held for Good Clinical Practice (GCP) compliance?	Data must be held for data curation, not for GCP compliance. Data will be deleted at the end of the project from the AIDAVA server; the patient PHKG and

		will be proposed to the patient for further personal use.
	Does data need to be held to meet 'Open Data' requirements?	No, as this is about identifiable personal health data and because the reuse outside of AIDAVA may create liabilities and is therefore not allowed Note: We are considering the anonymisation of the annotated narratives and make them available; this will done only following approval by the EC of the hospitals who provided the narratives
	Does data need to be held to meet INternational Council of Medical Journal Editors (ICMJE) requirements or commitments?	Most probably NOT as the core of the publication will not be linked to personal data but to the development of new approaches to data interoperability, and acceptability of the software by the users

3.1. GDPR Compliance Checklist – where 'personal data' is processed:

Tick	Requirement	**Notes** [replace guide text with response]
Article 5: Principles compliance checks		
	a) Is processing lawful, fair, and transparent?	Processing will be pursuant to informed consent in its entirety.
	b) Is the purpose (or purposes) of the processing clearly defined	Yes - this is a research project constituting a feasibility study for the tools and methodology. The primary and secondary endpoints are specified in the Study Protocol (Annex 1 to Deliverable 1.4)
	c) adequate, relevant and limited to what is necessary	As the end objective of AIDAVA is to build a COMPREHENSIVE longitudinal medical record of the patients, all personal health data are relevant and important and are in scope.
	d) accurate and, where necessary, kept up to date	This will follow local guidelines and policies for the clinical health record data and assurances from the developers of the apps that hold participants' data. WP4 also defines processes for data quality that will inform data curation and processing
	e) kept and permits identification of data subjects for no longer than is necessary	This will be in line with European regulations around research governance and integrity, where retention schemes will allow for results reproducibility and potentially eventual Medical Device Regulation certification. Please refer to local site requirements but usually the retention period is 10-20 years and beyond for Medical Device Regulation requirements. The AI Act may also extend this to the 20-5 year mark.
	f) processed securely	Security specifications are available for each component of the project and they will be

Tick	Requirement	Notes [replace guide text with response]
		configured with an ongoing risk assessment approach
	2) can you demonstrate this compliance?	The system will come with a full audit-trail, storing who did what, when and in which context.
Articles 13 & 14 compliance		[See detailed Transparency Checklist below]
	Did the data come from publicly accessible sources?	No, data are expected to come from hospital information systems, GP systems, personal apps, medical devices and national hubs. All these systems have limited access as they contain personnel data.
	Are data subjects informed before processing starts for any new purpose if incompatible with original purpose where the controller wants to use data for a different purpose to the purpose for which they currently hold data	Yes the subject is informed - through information sessions as identified in the Schedule of Activities within Study Protocol (see Annex 1 to Deliverable 1.4) and through a Study Information Package (see Annex 2 to Deliverable 1.4). The subject will be requested to give consent through the Informed Consent form (see Annex 2 to Deliverable 1.4). All the material to be used with the patients are being reviewed with patient representatives from the 2 patient associations, partners in the project. .
	Does the Privacy Notice and/or Study Information Package (SIP) cover this processing?	Yes (see Annex 2 to Deliverable 1.4).
	What patient choices are available? Are these explained?	For the AIDAVA prototype - for evaluation purposes - we will only recruit patients who agreed to be involved in the curation of their data. The local research associate - or principal investigator - will explain the objective of the prototype to the patient who will be left with the choice to participate.
Articles 6 and 9: legal bases		
	What are legal bases under Article 6	Dependent on jurisdiction but likely public task, legitimate interest or where required by local jurisdiction regulations consent. There will be some legal obligation to meet research governance requirements as per any research study.
	What are legal bases under Article 9 (if 'special category' data)	Scientific research in the public interest. Note that commercialisation purposes would need to be handled under a separate arrangement unless consented to by participants but this will still not be permissible as a research project and further approvals will need to be sought after the project completes most likely
	Are Article 6 legitimate interests explained where relevant?	TBC - though it is unlikely LI will be used as a Lawful Basis.

Tick	Requirement	Notes [replace guide text with response]
	Are details of statutory obligations for Article 6 explained where relevant.	They will be.
	Is this proposed processing compatible with the declared purposes?	Yes - it is purely as specified as research with full ethical approval pursuant to participant consent.
Article 89(1) research exemption		
	If for research, do we meet Art 89(1) data minimisation	Yes - all data items will be justified and use of personal data, likely for record integrity and linkage, will also be clearly specified and justified.
Articles 15-23: Data Subject Rights		[See detailed table below]
	Do we support data subject rights?	TBC and part of the particulars for the Data TransferAgreement
	There is no use of automated decision making (e.g. profiling)	There is no automated decision making but automatic computation of risk score in parallel with manual computation by the clinician for comparison purpose (and potential action whenever relevant)
Articles 24-43: Controller-Processor		
	A28 & 29: What measures are there to ensure processors comply?	The evaluation will happen across 3 separates sites 1. Estonia: NEMC (hospital), MIDATA (HDI) 2. Austria: MUG(hospital), MIDATA (HDI) 3. The Netherlands: MUMC+ (hospital), DIGI.me (HDI) Each pair (hospital, HDI) will sign a Data Sharing Agreement building on their current process and/or local certification (e.g. DIGI.me). All processing will be strictly controlled by the Study Protocol (see Annex 1 of Deliverable 1.4), to be approved by the local Ethical Committee before patient recruitment
	A30: Is there an entry for this processing/data held in the register?	A Record of Processing Activity (ROPA) will be held by all project partners (see template in Appendix C).

Tick	Requirement	Notes [replace guide text with response]
	A32-34: Do we ensure appropriate security, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures?	Yes - see Section on Security and Regulatory requirements in *Deliverable 1.3 . User requirements.*
	A37-39: Is there a Data Protection Officer (DPO) and have they been or will they be consulted?	Yes - each partner has their own DPO and Partner i~HD advises on these matters from a project perspective.
Articles 44-50: International transfers		
	What form of data will be transferred to a third country or international organisation	Likely full data sets to an instance of the HDI software, curation tools and KG generation tools run in one of the clinical sites in Estonia, Austria or the Netherlands. There may be an instance used for MIDATA in Switzerland. In any event, data will move within the EU or Switzerland, which holds EU Adequacy for Data Protection since 2000 so special measures will be very limited.
	Are there safeguards for international transfers?	There will be if needed but all recipients are in the EU except MiData which is based in Switzerland, which holds EU Adequacy.
Article 90: Obligations of secrecy		
	Do we meet medical confidentiality requirements?	Yes - no personal data will be shared outside of the hospital or without express consent of the patients.

3.2. Data Subject Rights:

Note if supported and what process/procedure applies; if not, then describe the legal justification for not supporting this right.

Tick	Requirement	Notes [replace guide text with response]
	To be informed: about processing, about choices, about rights, about controller	This will be honoured as part of the information leaflet contained in the Study Information Package (SIP) and Informed Consent Procedures
	the right of access to see or receive a printed copy	This will be honoured as part of the Study Information Package (SIP) and Informed Consent Procedures. Additionally the entire premise of AIDAVA is to allow patients access to their medical records.
	the right to rectification – to correct any material errors in the personal data	This is at the core of the AIDAVA project: to support the patient to correct/update their record

Tick	Requirement	Notes [replace guide text with response]
	the right to erasure – where appropriate, to ask that all personal data is erased	This will be limited to regulatory requirements for the handling of data in a research study. If participants are not satisfied with data processing, they may withdraw from the study where all data processing will be limited to meeting regulatory requirements. Their medical data will be deleted from the AIDAVA platform when regulatory requirements have been met for retention but it will not otherwise be processed further after the participant has withdrawn.
	the right to restrict processing – to ask that some or all processing ceases [see opt-out]	Processing will be strictly limited to the participants identified in the Study Protocol, in each site, with a specific role limiting their access rights and processing rights. Data will only be processed in line with the informed consent and if the participant wishes to withdraw from the study and halt processing they may do so without giving any reason and with no impact on their care or otherwise. .
	the right to data portability – this only applies to data provided directly by individual	The premise of AIDAVA is the provision of data portability so this will be met.
	the right to object to and not to be subject to automated decision-making, including profiling	The profiling element is the core focus of the study so participants may choose not to enrol with the study or withdraw as they deem fit. There is no closed loop decision making. This will be carefully explained in the informed consent procedures, information leaflet and study protocol
	The right to object to a Data Processing Authority (typically the relevant supervisory authority of each Member State)	This remains intact for this study and participants will be given the contact details of the DPOs for the recruiting sites for any questions, exercising their rights or complaints. Note that AIDAVA staff will not answer data protection questions, these will be sent to the DPOs for each recruitment site as per any research project. AIDAVA staff will answer questions about the study. A standard Operating Procedure will be developed as part of the Protocol.
	Where consent is the legal basis, the right to withdraw consent	This remains entirely possible and will be made clear in the consent procedures, information leaflet and protocol, where data will be archived in the event of withdrawal under legal obligation in line with research governance regulatory requirements and practices. Participants will be reminded they may withdraw from the study at any time without giving any reason and will not have their care or any other service impacted in any way.

3.3. Detailed Transparency Checklist[6]

Does privacy information provided to data subjects include:

Tick	Requirement	Notes [replace guide text with response]
	The name and contact details of our organisation	All DPO and contact details will be listed for all partners.
	The name and contact details of our representative (if applicable)	As above.
	The contact details of our data protection officer (if applicable)	As above
	The purposes of the processing	Listed in the information leaflet.
	The lawful bases for the processing	Listed in the information leaflet and informed consent procedures in line with local jurisdiction regulatory directions.
	The legitimate interests for the processing (if applicable)	TBC - these will be listed if needed in the information leaflets (Study Information Package - Annexe 2 of Deliverable 1.4) and informed consent procedures.
	The categories of personal data obtained (if the personal data is not obtained from the individual it relates to)	Listed in the information leaflet and informed consent procedures in line with local jurisdiction regulatory directions.
	The recipients or categories of recipients of the personal data	See above
	The details of transfers of the personal data to any third countries or international organisations (if applicable)	See above
	The retention periods for the personal data.	See above
	The rights available to individuals in respect of the processing	See above
	The right to withdraw consent (if applicable)	See above
	The right to lodge a complaint with a supervisory authority	See above

[6] Taken from UK Information Commissioner's Office template as an example

Tick	Requirement	**Notes** [replace guide text with response]
	The source of the personal data (if the personal data is not obtained from the individual it relates to)	See above
	The details of whether individuals are under a statutory or contractual obligation to provide the personal data (if applicable, and if the personal data is collected from the individual it relates to)	They are not but wIll be listed in the information leaflet and informed consent procedures in line with local jurisdiction regulatory directions.
	The details of the existence of automated decision-making, including profiling (if applicable)	WIll be listed in the information leaflet and informed consent procedures in line with local jurisdiction regulatory directions.
	We provide individuals with privacy information at the time we collect their personal data from them – or where e obtain personal data from a source other than the individual it relates to, we provide them with privacy information	Yes - via informed consent procedures and information leaflet.
	within a reasonable of period of obtaining the personal data and no later than one month	Yes
	if we plan to communicate with the individual, at the latest, when the first communication takes place	The patient is a key user in the evaluation of the prototype; he/ she will be in continuous communication with the research associate during this evaluation.
	if we plan to disclose the data to someone else, at the latest, when the data is disclosed	This will all be specified within the consent procedures and information leaflet, and unless required due to unforeseen circumstances, this will be set for the project.

	We provide the information in a way that is:
	☐ concise;
	☐ transparent;
	☐ intelligible;
	☐ easily accessible; and
	☐ uses clear and plain language.
		All patient related information is assessed with the patient representatives from the 2 partners Patient Association Organisation

	When drafting the information, we: ☐ undertake an information audit to find out what personal data we hold and what we do with it. ☐ put ourselves in the position of the people we're collecting information about.	See above.

Tick	Requirement	Notes [replace guide text with response]
	☐ carry out user testing to evaluate how effective our privacy information is	
	When providing privacy related information across the evaluation to individuals, we use a combination of appropriate techniques, such as: ● a layered approach; ● dashboards; ● just-in-time notices; ● icons; and ● mobile and smart device functionalities.	We will work in line with advice from the Supervisory Authorities. This will be included in the SIP (as part of the recruitment and consenting process) and the Privacy Notice online on the AIDAVA website.

3.4. Security & Access Control Checklist

Controls need to be appropriate to the level of risk: identified special category data needs more protection against potential misuse than non-personal data.

Tick	Requirement	Notes [replace guide text with response]
	Data Security classification (above Official)	- Official-Sensitive - **Secret** - Top Secret - Public Domain
	Personal Data involved [GDPR]	Yes
	Special Category of personal data involved [GDPR]	Yes
	Electronic Communications (inc. cookies) e.g. Privacy and Electronic Communications Regulation [PECR] (https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32002L0058)	Yes
	Credit Card data	NOT IN SCOPE
	Legal enforcement [LED2018]	Unlikely - unless part of criminal investigations in line with any research project
	Financial data	NOT IN SCOPE
	Intellectual Property (detail owner)	Part of Consortium Agreement

Tick	Requirement	**Notes** [replace guide text with response]
	Commercial in confidence (detail owner)	Part of Consortium Agreement
	Data Location (storage or processing) (include any back-up site(s))	- UK - **EU/EEA** - EU White-list - USA - **Other: Switzerland**
	Is data held in secure data centre?	Either within Hospital Network or Certified infrastructure for Netherlands Regulatory Specification (including ISO 27000 Series)
	Is this new supplier, location, or system?	The AIDAVA solution includes 2 systems • The Virtual Assistant (VA), running within a dedicated testing environment within the hospital. This is a new system in the hospital. • The related HDI which transfers data with the VA. This is a new supplier for the hospital. - wHDIs are a new supplier bit for an established service
	Is all user access subject to 2-factor authentication?	2-factor (e.g. password & key generated by the system)
	Are there established JML procedures?	This will be in line with partner policies but a SOP will need to be defined.
	Are there checks that passwords are robust and secure enough?	In line with partner policies and will be checked
	Are all administrator & user accounts routinely monitored?	In the audit trail, every action of the accounts will be tracked
	Are systems protected against malware and other attacks?	The AIDAVA system will be placed in a firewall protected environment in the clinical partner's system, the API endpoint which will be available from outside are protected by authentication.

[Need some aspect of CIA/impact-likelihood assessment]

3.5 Information Asset (IA) Register Checklist

Tick	Requirement	**Notes** [replace guide text with response]
	Are there new IAs being created?	Yes - mainly the personal health knowledge graph (PHKG), as a formal representation of the patient longitudinal health record.
	Are old IAs being retired?	No

	Have Information Asset Owners (IAO) and InformatioN Asset Administrators IAC) been consulted?	Yes - as part of the agreement to participate in AIDAVA
	Has the Information Asset Register (IAR) been updated/amended?	The Data Catalogue will fulfil this requirement and each site will update their own existing IARs. See A30 under page 12.
	Data Retention classification & period	In line with research governance regulations - PHKG of the patients will be retained potentially up to 25 years depending on local Member State law and any investigations that may be required. This may not be the case for derived data sets (IPS of the patient, extract to compute CVD score of the patients and BC registry extracts). Data might possibly be reused.
	Data retention procedure/functionality in place	Provided by each hospital site, HDI and Developer within the AIDAVA consortium.

Appendices

Appendix A – Supervisory Authority ‘High Risk’ Check

If the DPIA shows ‘high risk’ processing which cannot be mitigated, then the DPIA should be sent to the relevant authority for review before any processing starts. Note that their review may take several weeks to process. A ‘High Risk’ assessment represents a ‘risk to the rights and freedoms of individuals’ – so may extend beyond GDPR consideration, including Human Rights.

GDPR Article 35(3) provides three examples:

a) a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;

b) processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 1013; or

c) a systematic monitoring of a publicly accessible area on a large scale

As a case in point the UK Information Commissioner’s Office (ICO) cites the following but an different Supervisory Authority definition can be provided if needed:

1. Systematic and extensive profiling with significant effects
2. Large scale use of sensitive data [viz. ‘special category’ in GDPR terms]
3. Public monitoring

These being the same as (a)-(c) above. They further identify:

1. **New technologies**: processing involving the use of new technologies, or the novel application of existing technologies (including AI).
2. **Denial of service**: Decisions about an individual’s access to a product, service, opportunity or benefit which is based to any extent on automated decision-making (including profiling) or involves the processing of special category data.
3. **Large-scale profiling**: any profiling of individuals on a large scale.
4. **Biometrics**: any processing of biometric data.
5. **Genetic data**: any processing of genetic data, other than that processed by an individual GP or health professional for the provision of health care direct to the data subject.
6. **Data matching**: combining, comparing or matching personal data obtained from multiple sources.
7. **Invisible processing**: processing of personal data that has not been obtained directly from the data subject in circumstances where the controller considers that compliance with Article 14 would prove impossible or involve disproportionate effort.
8. **Tracking**: processing which involves tracking an individual’s geolocation or behaviour, including but not limited to the online environment.
9. **Targeting of children or other vulnerable individuals**: The use of the personal data of children or other vulnerable individuals for marketing purposes, profiling or other automated decision-making, or if you intend to offer online services directly to children.
10. **Risk of physical harm**: Where the processing is of such a nature that a personal data breach could jeopardise the [physical] health or safety of individuals.

'High Risk' assessment using ICO criteria:

[The assessment can be one of N/A (not applicable), Low, Medium, or High. The comments should explain how the assessment is justified.]

Criterion:	**Assessment**	**Comments**
New technologies	Y	A Virtual Assistant and research tool
Denial of service	N	
Large-scale profiling	N	At least not yet - no more than 90 participants across the two Use Cases and three sites
Biometrics	N	
Genetic data	N	
Data matching	Y	
Invisible processing	N	
Tracking	N	
Targeting of children or other vulnerable individuals	N	Unless participants are otherwise considered vulnerable due to age and / or health condition
Risk of physical harm	N	

Appendix B – Broad Privacy Risk Assessment:

1.	Data accuracy and timeliness	Yes - in line with clinical sites, registries, and participant held apps. For the onboarding, curation, and KG generation and use data quality checks will be implemented.
2.	Differential treatment of patients/data subjects	No - participation will not affect care provision and this will be made clear.
3.	Data Accuracy and identification	To be covered in development specification and data quality work
4.	Holding / sharing / use of excessive data within systems	Potentially but all data use will be justified and subject to Research Ethics Committee Review
5.	Data held too long within systems	As above
6.	Excessive range of access in terms of users to personal data (consider new users/change of access privileges)	The access requirements will be put to an Ethics Committee, PAOs and PCs.
7.	Potential for misuse of data, unauthorised access to systems	There is a requirement to work to high security standards on assured systems and these will be implemented in line with codes of practice based on ethics committee approvals and partner risk assessments.
8.	New sharing of data with other organisations, including new or change of suppliers	HDIs and research organisations will receive data from the hospitals to achieve the purposes of the study and will be bound by data sharing agreements and security requirements as is standard.
9.	Variable and inconsistent adoption / implementation	The implementation is in line with a research protocol as per D1.4 and will be approved by an ethics committee(s). This will ensure consistency in implementation and adoption for the study conduct.
10.	Legal compliance, particularly DP transparency requirements and support for data subject rights	We anticipate a high standard of compliance given the risk management approach and independent ethics committee approvals.
11.	Medical confidentiality	Medical records will be shared with participant consent for their own use. Ethics committees will approve the project and it will be subject to informed consent and with security infrastructures that are compliant or certified to ISO 27001.

Appendix C – Description of the ROPA

See spreadsheet [ref]

6.2 Joint Controller Agreement - draft template

Call: HORIZON-HLTH-2021-TOOL-06
Topic: HORIZON-HLTH-2021-TOOL-06-03
Funding Scheme: HORIZON Research and Innovation Actions (RIA)

Grant Agreement no: 101057062

AI powered Data Curation & Publishing Virtual Assistant

Joint Controller Agreement
Annex 2 to Deliverable D4.4

Funded by
the European Union

Document History

Version	Date	Description
V1.0	2023-05-02	First version based on i~HD Joint Controller Agreement Template, version 1.0, 2022-20-12
Vx		
Vx		

Table of Contents

List of Abbreviations and definitions

The abbreviations and definitions used in the deliverable are based on the AIDAVA Glossary [ref].

Joint Controller Agreement

BETWEEN:

UNIVERSITEIT MAASTRICHT (UM), established in MINDERBROEDERSBERG 4-6, 6211 LK MAASTRICHT, the Netherlands | PO Box 616 6200 MD, Maastricht, the Netherlands, the Coordinator

b!loba [b!lo], established in JAGERSLAAN 23, TERVUREN 3080, Belgium,

KATHOLIEKE UNIVERSITEIT LEUVEN [KUL], for the purposes of this Agreement represented by KU Leuven Research and Development, with offices in Waaistraat 6 box 5105, LEUVEN 3000, Belgium,

SIRMA AI EAD [Onto], established in 135 TSARIGRADSKO SHOSE BLVD, SOFIA 1784, Bulgaria,

THE EUROPEAN INSTITUTE FOR INNOVATION THROUGH HEALTH DATA [IHD], established in OUDE MECHELSESTRAAT 165, STROMBEEK-BEVER 1853, Belgium,

MEDIZINISCHE UNIVERSITÄT GRAZ [MUG], established in AUENBRUGGERPLATZ 2, GRAZ 8036, Austria,

SIHTASUTUS POHJA-EESTI REGIONAALHAIGLA [NEMC], established in J SUTISTE TEE 19, TALLINN 13419, Estonia,

GNOME DESIGN SRL [GND], established in STR DAKO 2A, SFANTU GHEORGHE 520014, Romania,

AVERBIS GMBH [AVER], established in SALZSTRASSE 15, FREIBURG IM BREISGAU 79098, Germany,

And the following Associated Partners:

DIGI.ME LIMITED [DME], established in 7 BOWER ROAD, WRECCLESHAM, FARNHAM, GU10 4ST, United Kingdom,

MIDATA Genossenschaft [MID], established in c/o EvalueScience AG, PFINGSTWEIDSTRASSE 16 8005 ZURICH, Switzerland,

hereinafter, jointly or individually, referred to as “Joint Controllers” or “Parties” relating to the project entitled AI powered Data Curation & Publishing Virtual Assistant,
in short AIDAVA, hereinafter referred to as ‘Programme’.

Whereas:

A. pursuant to the Consortium Agreement of 1st January 2021, (hereinafter referred to as the "Consortium Agreement") and the Grant Agreement 101057062 signed between the European Commission and the AIDAVA Consortium, or an Associate Member's Agreement where a Party is not a member of the AIDAVA consortium, the Joint Controllers have entered into cooperation the subject of which is (1) to generate synthetic data from patient health records for the purpose of developing the AIDAVA virtual assistant, and (2) to exchange patient health record for the purpose of testing the AIDAVA virtual assistant;

B. the Cooperation requires that the Joint Controllers process personal data, whilst they jointly determine the purposes and means of processing of personal data;

C. the processing of personal data by the Joint Controllers requires that a transparent manner of determining their respective responsibilities be established as regards their compliance with the obligations under the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (hereinafter referred to as "General Data Protection Regulation" or "the GDPR"[7]) and other generally applicable laws as well as relations between the Joint Controllers and the data subjects;

D. on concluding this Agreement, the Parties, seek to regulate the terms of processing of personal data in such a way that they meet the provisions of the GDPR, and

E. with regard to the data they process, the Joint Controllers act as controllers for the purposes of Article 24 et seq. of the GDPR referred to in D,

the Parties whose roles and responsibilities are further outlined in Appendix 1, decided to enter into the following Agreement:

§ 1.Definitions

For the purposes of this Agreement, the Parties agree that the following terms shall have the following meaning:

1. "Controller/Joint Controller" means any natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data;
2. "Coordinator of the Program" means <<COORDINATOR NAME>> who will act pursuant to the roles defined as Coordinator in the Consortium Agreement.
3. "Personal Data" means any information relating to an identified or identifiable natural person (hereinafter referred to as "data subject");
4. "Third Country" means any country that is not a member of the European Union or the European Economic Area;
5. "Processor" means any natural or legal person, public authority, agency or other body which processes Personal Data on behalf of the Controller;
6. "Data Protection Law" means the GDPR as well as other provisions of EU Member States or third country's national law applicable to a relevant Party, passed in relation to personal data protection, including in particular the provisions of the given Controller's national law;

[7] https://eur-lex.europa.eu/eli/reg/2016/679/oj

7. “Processing” means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;
8. “General Data Protection Regulation”, “GDPR” means the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC; wherever this Agreement refers to specific Articles of GDPR, it shall also apply to the corresponding provisions in national legislation guaranteeing a similar level of safety;
9. “Information System” means a group of cooperating devices, programs, information processing procedures and program tools used for the purpose of data processing;
10. “Cooperation” means the cooperation between Controllers defined in Recital A;
11. “Agreement” means this Agreement on Joint Control of Personal Data;
12. “Consortium Agreement” means the agreement referred to in Recital A;
13. “Consortium” means the AIDAVA Consortium;
14. “Data Provider” means the party to this agreement who is Processing data to provide it to Data Recipients to achieve the purposes of the Programme, where Providers of Data may also be Recipients;
15. “Data Recipient” means the party to this agreement who is receiving data from the Data Provider(s) and will Process it to achieve the purposes of the Programme, where Recipients of data may also be Providers.

§ 2. Subject-matter of the agreement

1. This Agreement regulates mutual relations between the Parties as regards the joint control of Personal Data, and in particular it determines in a transparent manner the Joint Controllers’ responsibilities for compliance with the obligations under the GDPR; it also defines the representation of the Joint Controllers in contacts with the data subjects and their relations with those data subjects.
2. For the purpose of proper implementation of this Agreement, the Joint Controllers shall:
   1) cooperate on performing the obligations of the Joint Controllers of Personal Data;
   2) process the Personal Data with which they have been entrusted with regard to the Cooperation pursuant to this Agreement, GDPR, where applicable the Consortium Agreement or the Associate Member agreement and other generally applicable laws and
   3) refrain from any legal or factual actions which might in any way undermine the security of Personal Data or threaten the other Joint Controller with civil, administrative or criminal liability.
3. Categories of data subjects and personal data, the purposes and means of processing, including the participation of Joint Controllers in those processes, as well as the categories of recipients of the Personal Data shall be defined in Appendix 2 to the Agreement.

§ 3. Controllers' rights and obligations

1. The Joint Controllers declare that they have the means enabling them to process and protect Personal Data they are processing, including information systems meeting the requirements of the appropriate level of security, as stipulated by the GDPR. They will each fully adhere to the applicable Data Protection Law(s) with respect to obligations and responsibilities of controllers.
2. In particular, the Joint Controllers shall:
   1) exercise due diligence in processing Personal Data and process Personal Data pursuant to the Agreement, the GDPR and other provisions of Data Protection Law(s), including the appropriate provisions of each Controller's national law;
   2) restrict access to Personal Data only to persons who need the access to Personal Data for the purposes of the Agreement and Cooperation, provide those persons with relevant authorisations, offer relevant training on personal data protection and ensure confidentiality of Personal Data processed thereby, both during and after their employment or other cooperation with a Joint Controller;
   3) assist the other Joint Controller, where possible, in meeting its (i) obligation to respond to requests from data subjects and (ii) obligations laid down in Articles 13, 14 (if applicable) and 32 through 36 of the GDPR;
3. The Joint Controllers shall provide each other with the necessary assistance in carrying out the obligations referred to in section 2 point 3) above, in particular in the notification of a personal data breach, by:
   1) providing, at the request of a Controller, information concerning the processing of personal data immediately upon receipt of such request as soon as possible;
   2) notifying the other Joint Controllers of any breach as soon as possible but not later than 48 hours of its discovery. The notification should include all the information referred to in Article 33 (3) of the GDPR. If - and to the extent that - the information cannot be provided at the same time, they can be given successively without undue delay;
   3) providing to the other Joint Controllers all information necessary for the communication of a personal data breach to the data subject;
   4) informing the other Joint Controllers of inquiries, requests or demands from data subjects and other individuals, national or European Union public administrations, including relevant supervisory authorities and courts, as well as any controls or inspections by such authorities in connection with the joint controllership of Personal Data; information shall be provided promptly and in such a way as to enable the other Joint Controllers to comply with the obligations set out in sections 2 and 3, without undue delay but not later than 7 calendar days after receipt of an inquiry, request or demand or after the start of a control or inspection.
4. The Joint Controllers shall abide by the provisions outlined in Appendix 3 relating to information Security Management and Appendix 4 outlining the Data Protection Impact Assessment and Joint Controller Responsibilities Matrix.
5. The Joint Controllers shall abide by the provisions of any Standard Operating Procedures or Codes of Conduct such as may be required by national legislation or the Parties (such that these do not prejudice or overall any national legislation) to achieve the purposes of the Cooperation.

§ 4. Data subjects' rights

1. The Joint Controllers shall inform, in any way they deem appropriate and where they are required to do so by their existing approvals or legislative requirements, the data subjects of the essences

of this Agreement and shall provide them the information referred to in Appendix 2 in accordance with Article 26 and Article 12 of the GDPR.

2. The information referred to in section 1 shall be primarily provided to the data subjects by the Controller who collects the personal data.
3. Data subjects may contact any of the Joint Controllers about the rights granted to them by Articles 15 - 22 of the GDPR. The contacted Controller shall identify the responsible Controller and forward the request internally to this Controller. The originally contacted Controller shall carry out all necessary communication with the data subject.
4. The responsible Controller shall be determined as follows: If the data of the data subject is part of a set of data which can be attributed to a Controller, this Controller shall be responsible. In all other cases the Controller contacted by the data subject shall be the responsible Controller.
5. The Joint Controllers undertake to comply with the data subjects' rights and shall assist one another with the execution of data subjects’ requests.

§ 5. Transfers of Personal Data to third countries

Controller and/or its Processor(s) that transfer(s) personal data in the scope of the execution of the Agreement to a Controller and/or Processor and/or other entity situated in the third country that does not present adequate safeguards under the GDPR shall ensure that such transfer complies with the GDPR (e.g. pursuant to Article 45 of the GDPR – on the basis of an adequacy decision - or Article 46.2.c GDPR)– on the basis of standard data protection clauses adopted by the Commission in accordance with the examination procedure in Article 93.2 or pursuant to Article 49 of the GDPR (EU Standard Contractual Clauses). A copy of standard data protection clauses referred to in the preceding sentence shall be provided when so requested by a data subject.

§ 6. Entrusting Processors with processing of Personal Data

1. The Controllers jointly consent to <<COORDINATOR NAME>> appointing Data Processors to Process the data and achieve the purposes of the Programme.
2. <<Processors>> will act as Processors with processing of Personal Data subject to this Agreement on terms and to the degree defined by this Agreement and Article 28 of the GDPR.
3. Each Controller may entrust <<Processors>> with processing of Personal Data under this Agreement only for the purposes of this Agreement, Consortium Agreement and the Cooperation.
4. Processors can only carry out specific Personal Data processing activities on behalf of <<COORDINATOR NAME>> once the <<COORDINATOR NAME>> has entered into a contract with <<COORDINATOR NAME>> laying down the obligations of the latter related to Personal Data protection in a manner ensuring sufficient guarantees of technical and organisational measures for the processing to meet the requirements of the GDPR.
5. Processors may carry out specific Personal Data processing activities on behalf of a Controller without entering into the contract referred to in section 3 as long as it is possible pursuant to another legal instrument under EU law or national law, which binds the Processor and the Controller.
6. This Paragraph shall apply in the case of any intended modifications regarding adding processors or replacing processors with other processors.
7. Categories of processors are listed in Appendix 2. The Joint Controllers shall provide detailed information on its Processors on request to the data subject.
8. For the avoidance of doubt, Processors will also include agencies recruited to process Human Samples for the purposes of metabolomics and Genome Wide Association Studies in line with the Cooperation purposes.

§ 7. Controllers' liability

The liability of the parties is governed by the legal regulations, in particular Article 82 of the GDPR with regard to the processing activities that they are in charge of as defined in regard to the Controller's role in the Collaboration and as stated in Appendix 1.

Each Party is subject to an obligation of result concerning the respect of the security, confidentiality and integrity of the Data.

The Parties's liability shall be limited to direct damages resulting from a breach by another Party or its subcontractors. Damages related to the security, confidentiality and integrity of Data shall not be considered as consequential damages.

No limitation of liability shall apply in the event of a breach of confidentiality, security and integrity of Data, in the event of fraud or breach of essential obligations by a Party.

§ 8. Collaboration of the Parties

The Parties shall cooperate in supervising the implementation of this Agreement.

The Parties agree that at the time of the implementation of the Agreement they shall cooperate closely, informing one another of any circumstances that have or may have effect on processing of Personal Data.

Each Party designates a contact point to coordinate the collaboration of the Parties in connection with the implementation of the Agreement, disclosing their personal data in Appendix 1.

Amendments to the Appendices shall not require an amendment of the Agreement, however all Parties shall have to be notified in advance thereof either in writing or electronically by the Coordinator. All Parties will have 14 business days to approve or object to the amendment. If after 14 business days a Party or Parties have not responded, the Coordinator may proceed as if these Parties have agreed.

The Parties shall allow each Data Provider, prior with proper notice in advance, at all reasonable times to inspect and review the steps being taken by it to comply with the terms of the Agreement and will give each Data Receiver any assistance which it reasonably requires with that inspection and review. Any inspection or audit initiated by a Data Provider will be paid by the Data Provider.

§ 9. Docking clause

1. An entity that is not a Party to these Clauses may, with the prior agreement of the Parties, accede to this Agreement at any time, either as a data provider or as a data recipient or where appropriate both, by completing the Appendix 1 and 2 and signing the Agreement.
2. Once it has completed the Appendices and signed, the acceding entity shall become a Party to this Agreement and have the rights and obligations of a data exporter or data importer in accordance with its designation in Appendices 1 and 2.
3. The acceding entity shall have no rights or obligations arising under these Clauses from the period prior to becoming a Party.

§ 10. Liability limitation

1. Each Party shall be liable to the other Party/ies for any (direct) damages it causes the other Party/ies by any breach of GDPR.
2. Each Party's aggregate liability towards the other Party under this Agreement shall not exceed each Party's total share of budget in the Collaboration.
3. Parties may not invoke the conduct of a processor or sub-processor to avoid its own liability.

4. The Parties agree that to the maximum extent permitted by applicable laws and regulations, in no event shall any Party (including its affiliates and subcontractors, and their respective directors, officers, and employees) be liable to the other Party in contract or otherwise howsoever arising or whatever the cause thereof, for any indirect damages, meaning any loss of profit, business, goodwill, reputation, contracts, revenues or anticipated savings or business opportunities, which arise directly or indirectly from any default on the part of the other Party.

§ 11. Term and termination of the Agreement

The Agreement will take effect as of the Effective Date, which is the date of the last signature to this Agreement.

The Agreement shall be concluded for the period of implementation of the Cooperation and as long as and until, after the termination of the Cooperation, obligations still have to be fulfilled.

§ 12. Final provisions

The Parties hereby agree that the Controllers shall process Personal Data pursuant to this Agreement free of charge, and neither the conclusion of this Agreement nor the processing of data pursuant thereto shall entitle any Controller to seek from other Parties, on whatever legal basis,

a) remuneration,

b) reimbursement of any costs or expenses incurred for the purpose of due performance of the Agreement,

c) exemption from any obligations contracted to that end or advances on such costs or expenses,

even if at the time of entering into Cooperation or concluding this Agreement, despite exercising due care, the Controller was unable to foresee the circumstances justifying such rises, costs, expenses or obligations.

Should any provision hereof become invalid or ineffective, the Parties shall adopt all measures possible to replace it with a valid and effective provision reflecting the goal and meaning of the invalid or ineffective provision to the extent of applicable law. Should any provision hereof be or become invalid or ineffective at any time, it shall not restrict the validity or effectiveness of the remaining provisions of the Agreement.

In the event of any discrepancies between the provisions of the Agreement and the terms of Cooperation agreed by the Parties, the provisions of this Agreement shall prevail.

4. Any amendments hereto must be in writing on sanction of invalidity, subject to § 8 (4).
5. Any disputes arising under the Agreement shall be resolved by amicably or by a common court with jurisdiction over the registered office of the Controller sued and pursuant to the laws applicable in its country.
6. The Effective Date will be the date of the last signature to this agreement.

Signatures of the Parties (to be handled by certified signature service, e.g. VeriSign)

For Each Party:

Name and role of the signatories on behalf of the Party

Date(s) and location(s) of signature

Party: ______________________________

Signature: ______________________________

Date: ______________________________

Name: ______________________________

Role: ______________________________

Appendices

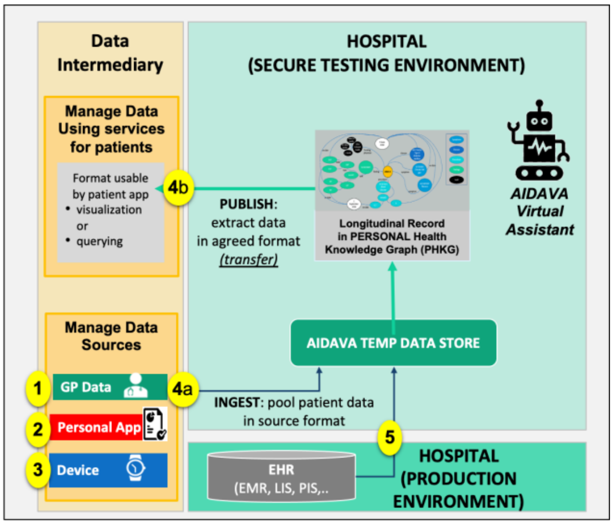
Appendix 1: Roles, Activities and Responsibilities of the Parties

Organisation Name	Country		
UNIVERSITEIT MAASTRICHT	Netherlands	Data Recipient and Data Provider	
b!loba	Belgium		
EURICE EUROPEAN RESEARCH AND PROJECT OFFICE GMBH	Germany		
KATHOLIEKE UNIVERSITEIT LEUVEN	Belgium		
SIRMA AI EAD	Bulgaria		
THE EUROPEAN INSTITUTE FOR INNOVATION THROUGH HEALTH DATA	Belgium		
MEDIZINISCHE UNIVERSITAT GRAZ	Austria		
SIHTASUTUS POHJA-EESTI REGIONAALHAIGLA	Estonia		
GNOME DESIGN SRL	Romania		
AVERBIS GMBH	Germany		
EUROPEAN CANCER PATIENT COALITION	Belgium		
EUROPEAN HEART NETWORK AISBL	Belgium		
MIDATA Genossenschaft	Switzerland		
DIGI.ME LIMITED	United Kingdom		

Appendix 2: Data Processing Particulars

Introduction

Based on use cases identified in Deliverable D1.1, there is a need for potentially 5 exchanges as displayed in the figure on the right and further described in the table below This document is aimed at covering Data Sharing at hospital level (i.e. numbers 4a, 4b and 5). Data sharing agreements between other Data Providers and the Health Data Intermediaries (HDI) are out of scope of this document; it is expected that each HDI will use their own existing agreements. If relevant they could use the enclosed template as a basis.


Data Intermediary
HOSPITAL (SECURE TESTING ENVIRONMENT)
Manage Data Using services for patients
Format usable by patient app
• visualization
or
• querying
4b
PUBLISH: extract data in agreed format (transfer)
Longitudinal Record in PERSONAL Health Knowledge Graph (PHKG)
AIDAVA Virtual Assistant
Manage Data Sources
AIDAVA TEMP DATA STORE
1
GP Data
4a
INGEST: pool patient data in source format
2
Personal App
3
Device
5
EHR (EMR, LIS, PIS,..
HOSPITAL (PRODUCTION ENVIRONMENT)

#	Data Provider	Recipient	Data Shared	Comments
1	Holder of GP data	HDI	GP Data	May not be possible for all evaluation sites
2	Holder of data from QLY questionnaire	HDI	QLY data	
3	Holder of medical device data	HDI	Medical device data	Only for CVD patients
4a	HDI	Hospital (AIDAVA)	All patient data , in scope, from HDI	
4b	Hospital (AIDAVA)	HDI	Patient IPS in HL7	
5	Hospital (production environment)	Hospital (AIDAVA)	All patient data in scope of AIDAVA from hospital system	May not require legal provisions, only legal provisions

Categories of data subjects

Categories of data subjects
All patients enrolled for the AIDAVA evaluation study, based on the eligibility criteria identified in Ethically Approved Study protocol.

There will be 15 Breast Cancer patients and 15 CVD patients, meeting the following eligibility criteria

- Health record available within the medical centre.
- Confirmed interest in being in more control of your personal health data and to to increase the quality of these data by working with the AIDAVA prototype.
- Owner of a smartphone and/or a tablet and comfortable to work with digital apps
- Confirmed diagnosis of Breast Cancer, aged 30 to 65 years and female OR confirmed diagnosis of symptomatic type 1 acute ST-Elevation Myocardial Infarction, aged 18 - 65 years.

Vulnerable persons and children are explicitly excluded.

Categories of personal data

Health data related to the patient condition, including personal identifiable information.

- Identification information (name, identifier) will be pseudonymized.
- Sensitive information such as ethnicity, birthdate, gender as well as health data and biometric data will be included in their source format

Purpose of processing

The AIDAVA project in scope of this Joint Controller Agreement will deliver 2 generations (G1 and G2) of a prototype at TRL5/6.

1. Generation1 (G1) will include the framework of the prototype composed by a backend and a front end part. The curation & publishing tools included in the backend will be based on open source tools; the front end will be based on an existing chatbot/typebot platform.
2. Generation2 (G2) will build on G1. The data curation tools will be replaced by the novel curation tools developed in the project. In addition, the front end of G2 will be expanded with an human-AI module to increase usability for users less experienced in data and medical content.

To develop and assess the AIDAVA virtual assistant we need to access patient data in 3 phases; each phase requiring EC approval.

1. (ongoing, already approved by the responsible Ethical Committee) Phase 1. De-identified documents for annotation of clinical narratives. To develop AI based novel curation tools in G2, that extract information from clinical narratives, we need training data sets. The training data sets are composed by pseudonymized extracts of clinical narrative that are then annotated with the concept to be extracted. These narratives should come from any patients - and preferably excluding the patients that will be assessing the prototype in Phase 2.
2. (in scope of this Data Sharing Agreement) Phase 2: Synthetic data. To support the development of the virtual assistant, we need to have test data which are representative of the problem to be solved so that the development team can verify if the system works properly before it is being used by patients on site. The extraction and data transfer process of these data will be the same as for Phase 3 described below. These data will then be undergoing a manual anonymization process to generate synthetic data, that will be then transferred to the development sites. See detailed processing below
3. (in scope of this Data Sharing Agreement) Phase 3: Pseudonymized personal data with reidentification when linking data across sources with patient consent. To demonstrate the value and the performance of the virtual assistant, we will install the prototype in a simulated environment, for use by consenting patients with their personal health data, and will perform a structured and formal evaluation study defined in the Study protocol being developed and to be approved by the local Ethics Committees. See detailed processing below

Processing of the data - Phase 2. Synthetic data

Purpose	Extract a set of data representative of real life to support 1. development of the prototype in the different development sites , outside of the hopsital (P-GND in Romania, P-UM in The Netherlands, ..) and 2. testing of the system before it is actually deployed. The main objective is to minimise discomfort for the patients that will evaluate the system.

Number of subjects/ site	Data can be scattered across multiple patients if needed to decrease any data privacy risks. A set of 3 patients per type of data source is deemed sufficient
Means of processing	Extract and pseudonymization of the data ● Extract of sample data based on technical specification mentioned in Annex III ● For pseudonymization / de-identification purpose, ○ replace with fake information, names and identification numbers (passport, social security) ○ remove email addresses, phone numbers Anonymization for generation of synthetic data (so that data can be used outside of the hospital) ● Apply statistical techniques [which ones/ Remzi] to the de-identified dataset to generate a synthetic dataset that has similar properties to the original data. ● Validate the synthetic data to ensure that it does not contain any PII or other sensitive information that could be used to re-identify individuals. ● Keep the resulting synthetic data on local server and forward a copy to the agreed development centres (who signed the Joint Controller Agreement and are bound to confidentiality)
Who is responsible for data collection and processing?	
Recipients/Categories of recipients	
Categories of Processors	
Duration of Processing	± 2 weeks (June 2023 ?)
Location of Processing	Local evaluation sites ● Hospitals: NEMC, MUC, UM ● HDI: MID, DME

Processing of the data - Phase 3. Pseudonymised data

Purpose	To formally evaluate the AIDAVA prototype (G1 and G2) with real patient and real data; the evaluation study is strictly constrained by an evaluation protocol that will be submitted for approval by the local Ethics Committee.
Number of subjects/ site	15 Breast Cancer patients 15 CVD patients
Means of processing	Extract and pseudonymization of the data ● Extract of sample data based on technical specification mentioned in Annex III ● For pseudonymization / de-identification purpose, ○ replace with fake information, names and identification numbers (passport, social security) ○ remove email addresses, phone numbers

	● Keep in a separate file, the mapping between the fake information and name, identification numbers for temporary re-identification when linking data from different data sources Transfer the data to the hospital secure server (see technical details in Annex III) ● For Hospital: data will not leave the hospital premisses but will be transfer from the productive environment to the secure testing environment when AIDAVA will be deployed and tested ● For HDI ????
Who is responsible for data collection and processing?	
Recipients/Categories of recipients	
Categories of Processors	
Duration of Processing	± 2 weeks (before deployment of respectively G1 and G1 - see planning)
Location of Processing	Local evaluation sites ● Hospitals: NEMC, MUC, UM ● HDI: MID, DME

Appendix 3: Information Security Management and Confidentiality

Confidentiality

The Parties shall:

Where possible, render anonymous or pseudonymised data in accordance with their local approvals and requirements to achieve the goals of the Cooperation where it is recognised this is not possible for Genetic Data;

Provide on request to each other details of their approaches to anonymous and / or pseudonymised data and in the case of genetic data ;

Ensure that they provide assurance of sufficient anonymisation and / or pseudonymisation prior to sharing the anonymous or pseudonymised data;

Put in place organisational and technical measures to assure the ongoing confidentiality of the data, including all processing systems and services;

Ensure that any transmission of data is authorised and encrypted using industry standard encryption protocols;

Routinely monitor, test and audit the systems and services, making improvements as may be necessary.

Personnel

The Parties undertake to:

adequately train its personnel on data protection and security;

inform personnel of the confidential nature of the Shared Data and ensure that personnel understand their obligations;

maintain a data protection and security training programme for relevant personnel and monitor completion of training.

Hardware, Software, Malware and Business Continuity management

The Parties shall:

only use current, up to date and supported versions of software and industry standard hardware for the purposes of the cooperation;

ensure that its suppliers of hardware and software adhere to industry recognised standards in the procurement of hardware and software.

Ensure the ability to restore the availability of and access to systems and services used to achieve the purposes of the Cooperation in a timely manner in the event of any incident;

regularly back-up copies of all data and store backups securely and separately from the live systems.

provide and routinely use industry recognised virus and malware protection software and techniques to prevent infection by viruses and malware

routinely update software and virus definitions for anti-virus software

Access controls

The Parties shall:

provide access to data only in line with the requirements to achieve the goals of the Cooperation and in accordance with the requirements of the approvals and any consents that have been received from Data Subjects, preventing any access by unauthorised parties;

allow for access to all personnel engaged with the Cooperation, including those personnel involved with the running and maintenance of any software, hardware and services;

provide at the very least password controlled access to systems and manage the privileges of users within the software and services.

Appendix 5: Data Transfer Technical Specifications

(see Annex 4 to the Study Protocol defined in Deliverable 1.4).