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

Grant Agreement 101057062

AI powered Data Curation & Publishing Virtual Assistant

Deliverable No. D1.4
Description of Assessment Study
(and supporting material)

Contractual Submission Date: 30/11/2023

Actual Submission Date: 30/11/2023

Responsible partner: P8 - NEMC/P2 - b!lo

Funded by
the European Union

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

Deliverable number	**D1.4**
Deliverable title	**Description of Assessment Study (and supporting material)**
Type[1]	R
Dissemination level[2]	PU
Work package number	WP1
Work package leader	P8 - NEMC and P2 - b!lo
Author(s)	Katrin Lepik, Eno-Martin Lotman, Kerli Norak, Mall Maasik, (NEMC) Isabelle de Zegher (b!lo)
Reviewers	Markus Kreuzthaler (MUG) Petros Kalendralis (UMC) Dominik Steiger/Emmanuel Benoist (MID) Hanne Muller (DME) Laura Saucedo-Cuevas (ECPC) Isabella Cinà (EHN)
Keywords	Study protocol, assessment study

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).

The work of MID received funding by the Swiss State Secretariat for Education, Research and Innovation (SBFI), subvention contract 22.00093, REF-1131-52104.

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

Document History

Version	Date	Description
V0.1	15.11.2023	Draft for review
V0.2	23.11.2023	Commented version by all reviewers ready
V1.0	30.11.2023	Final version ready
V2.0	05.07.2024	Answers to reviewers comments ● Comment 1: "*Evaluation study protocol is planned to apply for*

[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

		usability/user acceptance only the SUS (system usability scale), there is need to pay more attention to user acceptance, to the user and use of the system, to the system reliability, safety and security , and if the system is able to access / utilise all relevant patient data/information." This comment was addressed through the sentence added in the Introduction (3rd paragraph) Note: The list of data sources to be used is provided in Section 10.1 of the Annex 1 - Study Protocol. These are the ONLY data that the system will access and utilise; this is strictly constrained by Ethical Committees. ● Comment 2: " *Additionally, one of the main requirements for this reporting period is to provide clear and concrete lists of the data elements that will be collected from the 45 patients (15 per data provider) enrolled in both proposed pilots. These data elements should specify how the data will be codified with interoperability standards. Among the defined data elements, identify those that may require human interaction, specifying whether the human is a patient, data steward, data expert, or healthcare professional.* The list of data DE of the breast cancer registry has been added as Section 10.5 of the Annex 1 - Study Protocol; the list of DE for the CV use case is included in Section 3.2.2 of the same Annex in the table specifying how the SMART risk score should be computed. Both lists include the relevant SNOMED or LOINC code. The objective of AIDAVA is to generate these DE automatically from one or more Personal Health Knowledge Graph as explained in *Deliverable D1.1. - Description of the use cases* (see Figure 5 in Section 4.1 and Figure 6 in Section 4.2)

Table of Contents

List of Definitions

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

List of Abbreviations

CVD - Cardiovascular Disease
DPIA - Data Protection Impact Assessment
DTS- Data Transfer Specification
ELSI - Ethical Legal and Social Issues
G1 - Generation 1 (of prototype)
G2 - Generation 2 (of prototype)
GP - General practitioner
QALY - quality-adjusted life year
HDI - Health Data Intermediary
ICF - Informed Consent Form
MDR - Medical Device Regulation
SIP - Study Information Package
SUS - System Usability Scale

Executive Summary

The AIDAVA prototype will be delivered in 2 generations: Generation 1 in Q3 2024 and Generation 2 in Q2 2026. It will be tested in 4 hospitals and 2 Health Data Intermediaries, with 45 patients respectively per therapeutic area across all sites (90 patients for the 2 therapeutic areas in scope).
This deliverable includes the description of the 4 documents developed to support the execution of this assessment study of the two generations of the AIDAVA prototype in an ELSI compliant way, with a minimum burden for the patients and the sites.

The first document - and the most important one - is the study protocol (Annexe 1); it starts with a synopsis of the study and includes a description of the objectives of the study, the specification of the primary and secondary endpoints, the study schedule with the different activities to take place during the evaluation of the prototype across the 2 generations (including the washout period between the 2 generations), the study population with eligibility criteria, the data points to be collected with associated data collection forms (in RedCap) and the statistical analysis.
Another important document, related to the protocol is the English version of the Study Information Package and Informed Consent Form (Annexe 2) to be translated by each site and provided to patients during the recruitment process.
The third document includes a training plan (Annexe 3) for the patients participating in the evaluation and for the study team. It includes a specification of the different modules and a training program for the participants of the study, based on their role.
The final document is a template Data Sharing agreement (Annexe 4), to be adapted and finalised by each site, including guidance for technical and legal provisions.

The deliverable also includes description of work that was conducted with the help of Health Data Intermediaries (HDI) who helped to identify vendors who would provide a patient app application (to collect Quality of Life information) and a blood pressure medical device to be used during the study; the collected data will be managed by the HDI and provided to AIDAVA for integration in the patient record.

We also provide an overview on the feedback provided by the patients´consultants for the different materials mentioned above, and specify the study design with the schedule of activities as well as the Study Information Package and the Informed Consent Form.

1 Introduction

The AIDAVA intelligent virtual assistant is a prototype medical device[3] that supports patients to curate and publish their personal data in a secure environment, generating an interoperable and reusable personal medical record that can be used in clinical care to improve patient outcomes and to support data driven clinical research. To verify that the resulting prototype meets the requirement, we need to perform a formal assessment study. More specifically, we want to check that the AIDAVA intelligent virtual assistant is:

1. effective in **improving data curation and publishing process** against existing practices,
2. usable and acceptable **by patients** with different levels of health, digital and data literacy **and by expert data curators** whenever their input is required,
3. **valuable for "data users"** such as cardiovascular treating physicians and breast cancer clinical researchers.

The efficacy of AIDAVA virtual assistant will be influenced by the quality of the different data curation tools (supporting automation) and by the explainability module (managing the human computer interaction). The study should help identify how these components impact the effectiveness of the device.

Evaluating a prototype medical device like AIDAVA requires to work with patients who agree to collaborate and share their personal data. In the case of patients recruited through hospitals, this is only acceptable with approval from the local Ethical Committees in each hospital on the basis of a structured research study protocol, with clear description of objectives, and approach including a close monitoring of each patient with study nurses during the evaluation. It goes without saying that the prototype provided to the users needs to be tested following good software development practices including unit test, integration test and mainly user acceptance test; this testing approach is part of the *Deliverable D 3.6 - G1 installed for testing.* System reliability, safety and security are part of the business requirements and included in the testing approach in D3.6.

Task 1.4 "*Details testing scenario to assess performance/acceptance of prototypes*", in scope of this document, is responsible for describing the different steps and activities needed to formally evaluate the prototype with on-site patients using their personal data in real life settings. The assessment study includes Generation 1 and Generation 2 of the AIDAVA prototype. As it was decided to keep as much of the same patient as possible across the 2 generations of the prototype, the assessment study also must cover the period of inactivities for the patients during the two generations.

The first objective of the task was to define an end-to-end testing scenario from data sources, ingestion into the AIDAVA environement, curation and then to publishing based on the use cases defined in Deliverable D1.1 and in compliance with ELSI requirements. An important point in the study was to identify practical and measurement endpoints to measure effectiveness of the prototype and user acceptability. The scenario was developed and agreed upon with key user representatives, hospital staff and Health Data Intermediaries (HDI) representatives; it was also validated by the patient consultants to ensure its feasibility.

[3] A prototype does not need to go through MDR certification ; at the end of the project the idea is to develop a product medical device that will require MDR.

The second objective was to translate this testing scenario into a set of formal documents to be submitted to the different local ethical committees for approval.

While defining the assessment, we needed to ensure that any risks involved with the transfer, storage and disposal of data during the study were properly managed. The third objective was therefore to ensure that the infrastructure supporting the testing of the prototype which is to be deployed at the different sites, would meet local security and data privacy requirements (security, qualification, environment) at each organisation. A thorough overview with details on data streams and responsibilities was created, and a template sharing agreement between parties exchanging data was defined.

The task resulted in 4 documents:

1. Study Protocol following best practices for digital device evaluation
2. Study Information Package (SIP) and Informed Consent Form (ICF); the SIP will be provided to the patients as part of the recruitment process, before they are asked to sign the ICF
3. Training modules and training program for the different participants of the evaluation
4. Data Sharing agreement with legal and technical provision to support data sharing across the different sites

The different activities that took place to develop these documents are presented in Section 2, while Section 3 provides a brief overview of these documents which are attached in full as annexes.

2 Description of Activities

Across the task, the project followed a co-creation approach amongst the different participants: we organised bi-weekly meetings of 2 hours between February and November 2023, held multiple meetings with clinicians and specialists to discuss the specifics of the both use cases (breast cancer and cardiovascular disease (CVD)) and more particularly the secondary endpoints related to medical aspects of the use cases. We also organised one face-to-face workshop with clinicians as a part of the General Assembly meeting on 25-26 October 2023 in Graz to specify details of the study protocol. Finally, we had several meetings with the patient consultants to check the feasibility of the activities to be done by the patients and to verify if the material was acceptable and understandable.

The following table provides an overview of the activities performed during the task.
A more detailed description of the activities related to the different document is provided in the remainder of this section.

Activity	End Date	Lead	Partners involved
Scientific aspects			
Draft Assessment Study Protocol	Apr. 23	NEMC, b!lo	NEMC, UM, MUG, b!lo
Test the protocol with patient consultants	Jun-Aug 23	ECPC, EHN	ECPC, EHN, b!lo
Detailed review of schedule of activities and simulation of testing (see Figure 2)	Sept 23	NEMC	NEMC, MUG, b!lo
Finalise Assessment Study Protocol (with local requirements/translations)	Oct. 23	NEMC	NEMC, UM, MUG, ECPC, EHN, b!lo
Draft Study Information Package (SIP) & ICF	May 23	b!lo	b!llo, NEMC, UM, MUG, EHN, ECPC
Finalise SIP (test with patient consultants)	Aug. 23	NEMC	ECPC, EHN, b!lo
Translate SIP and ICF	Oct. 23	NEMC	NEMC, UM, MUG
HDI agreement	Nov. 23	NEMC	NEMC, UM, MUG, MIDATA, DME, b!lo
Create data collection forms in RedCap	Sep. 23	MUG	NEMC, MUG, b!lo
Create translations for all needed forms in all sites	Oct. 23	MUG	NEMC, MUG, UM
Draft training for patient & other users (non technology aspects)	Nov.23	B!lo, NEMC	
Get Local Ethical Committee´s approval of Assessment Study protocol (by each site)	Nov.23	NEMC	NEMC, MUG, UM
Data Privacy aspects			
Review Assessment protocol + ICF + SIP with Data Protection Officer (IHD)	Sept.23	IHD	IHD, NEMC, other

Activity	End Date	Lead	Partners involved
Template Checklist ready (Information Governance)	Aug.23	IHD	IHD, b!Lo
Information Governance checklist review and Data Privacy Impact Assessment (DPIA) - see Task 4.1 - executed and approved (by each site)	Sept.23	IHD	NEMC, UM, MUG
Data Management/ handling			
Data Transfer Specification (DTS)	Nov.23	b!lo	NEMC, UM, MUG, MIDATA, DME
Definition of data transfer specification (DTS) template; creation of the local DTS for each evaluation site	ongoing and continues in Task 1.5.	NEMC	NEMC, UM, MUG, MIDATA, DME
Draft Sharing Agreement (legal provision and technical provision) - HDI ⇔ AIDAVA	Oct. 23	NEMC, IHD	NEMC, MUG, UM, MIDATA, DME
Selection of blood pressure medical device	Sept 23	NEMC	NEMC, MUG, UM
Third parties app (QALY) s - selection with doctors, introduction, agreement with HDI	Oct. 23	MIDATA	NEMC, UM, MUG, MIDATA, DME
GP data from MUG/UM/NEMC	Oct. 23	MIDATA, DME	NEMC, UM, MUG, MIDATA, DME

Table 1. Overview of activities

2.1 Development of the Study Protocol

B!lo and NEMC created a first draft of study protocol using the Transcelerate eProtocol template ("Common Protocol Template Now Available" 2016) as a basis; this helped the team to follow best practice in study design and ensure that no critical aspects are missed in the conduct of the study. Most specifically we paid attention to the following components:

- Patient data privacy and ELSI, while we are managing fully identifiable data of the patients with their full consent
- Identification of primary and secondary endpoints meaningful to demonstrate the effectiveness of the prototype, and provide measurable outcomes
- Consistency of the different activities
- Effective collection of all data needed during the evaluation, supporting the proposed statistical analysis
- Acceptability of the set of activities by the patients

After the creation of the first draft, clarification discussions with clinicians were held to review, edit and confirm the schedule of activities, eligibility criterias and secondary endpoints of the study.

2.1.1 Patient data privacy and ELSI compliance

The AIDAVA prototype will manage fully identifiable patient data, as we need to link and integrate patients across data sources. To ensure there is no error, it is critical to check the patient's identification before integrating a data source.
Following discussion with the AIDAVA Data Privacy Officer (from IHD), it was considered compliant with data privacy as long as the patient was properly informed of the fact that their data will be fully identifiable and that the prototype will provide appropriate security mechanisms. These aspects were respectively managed in Annex 2 (SIP and ICF) and in Annexe 4 (Data Sharing Agreement including technical description on local solution).

2.1.2 Identification of primary and secondary endpoints

As in any study, the endpoints constitute the cornerstone of the AIDAVA study protocol. The primary endpoints relats to assessing the effectiveness and acceptability of the prototype, while the secondary endpoints check on the validity of the expected output for the 2 use cases. We specifically paid attention to the aspects related to acceptability - as described below.

Assessing usability

The System Usability Scale (SUS) provides a "quick and dirty", reliable tool for measuring the usability. It consists of a 10-item questionnaire with five response options for respondents; from Strongly agree to Strongly disagree. Originally created by John Brooke in 1986, it allows you to evaluate a wide variety of products and services, including hardware, software, mobile devices, websites and applications.

Benefits of using a SUS

SUS has become an industry standard, with references in over 1300 articles and publications. The noted benefits of using SUS include that it:

- Is a very easy scale to administer to participants
- Can be used on small sample sizes with reliable results
- Is valid – it can effectively differentiate between usable and unusable systems

The System Usability Scale

When a SUS is used, participants are asked to score the following 10 items with one of five responses that range from Strongly agree to Strongly disagree:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

The questionnaire and scoring are outlined in the System Usability Scale template. ("System Usability Scale (SUS)" 2013)

Interpreting Scores

Interpreting SUS scoring can be complex. The participant's scores for each question are converted to a new number, added together and then multiplied by 2.5 to convert the original scores of 0-40 to 0-100. Though the scores are 0-100, these are not percentages and should be considered only in terms of their percentile ranking.

Based on research, a SUS score above 68 would be considered above average and anything below 68 is below average, however, the best way to interpret your results involves "normalising" the scores to produce a percentile ranking.

2.1.3 Critical activities and consistency of the different activities

Screening

Screening is the process of active evaluation of potential participants for enrollment in a trial. Screening occurs during the enrollment period to see if they meet the inclusion and exclusion criteria. If they meet the criteria, the subject is eligible to be enrolled in the trial.

In AIDAVA, screening takes place during a regular hospital visit where the physician checks the inclusion/exclusion criteria and briefly introduces the project. If the patient meets the criteria and is interested in participating in the project, the research associate will then:

- share details of the project with patients ;
- introduce the Study Information Package, including use of EQ-5D Quality of Life questionnaire
- introduce Health Data Intermediary (HDI) and answers to questions regarding ownership and future use of the data through HDI
- ask to sign an Informed Consent Form including HDI agreement
- teach CVD patients how to use the medical device to measure blood pressure and send data.

Washout Period

The washout period is the period between development and testing of the two generations (G1 and G2) of the AIDAVA prototypes. This is the period when patients are not expected to use AIDAVA and curate their data (except CVD patients who measure their blood pressure during the washout period and send their data) but during this period it is important to communicate with patients so that they do not withdraw from the study.

Communication every 2 months:

- **Objective**: provide regular information on the project to maintain the patients' interest to decrease dropout during the G2 development/ improvement phase.
- **Medium**: Regular newsletters (send through emails; additional online meetings after 9 months will be explored).
- **Expected planning and content:** n each communication across all newsletters, to include a message highlighting the advantages of using AIDAVA, key milestones and achievements in the project, the value of patients' participation in the evaluation, and the important steps the patients can take to improve the quality of care by using it.

Checking consistency across the activities

Since the important information and schedule related to the study were in several different tables and files, it was important to ensure consistency across the different parts of the Study protocol and align the activities to be performed during the study:

- Overall Schema of the assessment study, (p 11)
- Schedule of Activities (SoA), (p 12-13)
- Study visits overview, (p 30-36) and
- Data collection & entry forms in REDCap.

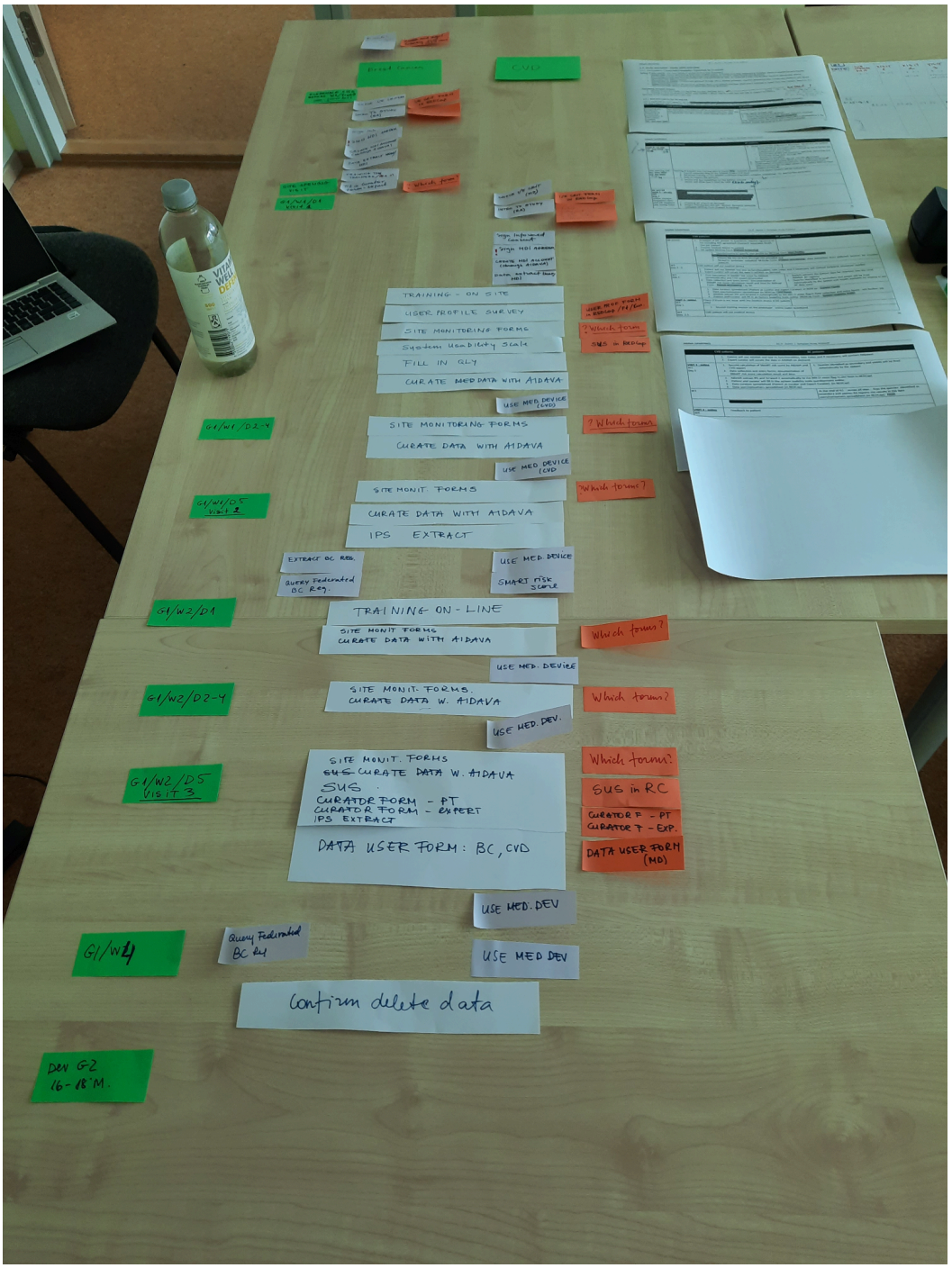


Figure 2. Checking the consistency of the different parts of the protocol

2.2.3 Effective data collection: development data collection and entry forms in REDCap

All the data needed to compute the endpoint must be captured during the study in a similar way across the three evaluation site. We decided therefore to use a formal data collection system, widely used across clinical sites for clinical trials called, REDCap.

REDCap is a secure online platform for building and managing online databases and surveys. It offers a wide range of tools that can be adapted to a wide variety of data collection strategies and allows you to export survey data to Excel and standard statistical packages (SPSS, SAS, Stata, R) and prepare various reports from the collected data in the REDCap environment.

The following REDCap forms were created to enter the data collected during the survey:

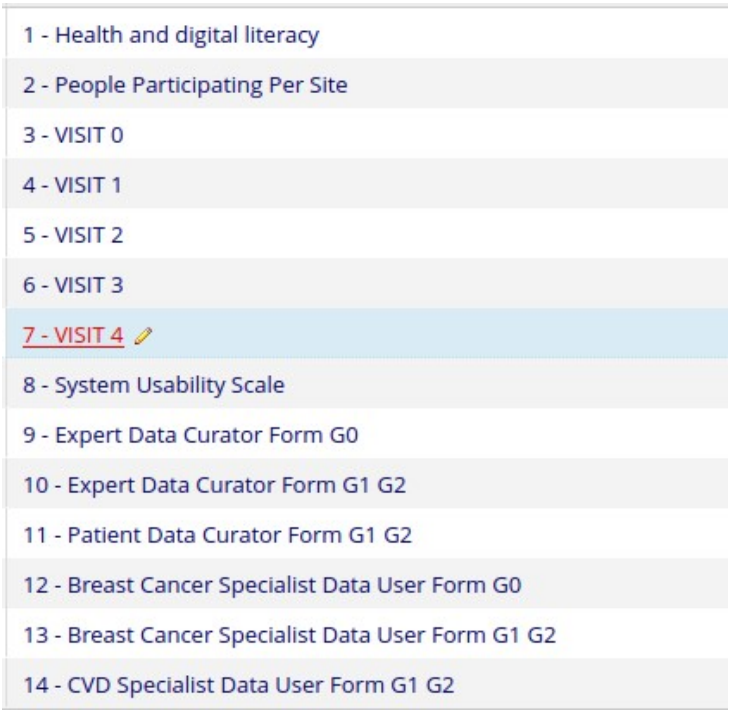


Figure 4. REDCap forms

2.2.4 Acceptability of the set of activities by the patients: Patient consultants´ feedback

After the initial draft of the study protocol, an introduction to materials was given to patient consultants who provided their feedback listed below:

Screening

- Process is generally clear, however recommendations include:
  - Provide a live demo to explain what the medical device and app are and how to use them,
  - Before patients start using the app, an appropriate explanation of its scientific importance should be provided,
  - Use plain language (i.e. no acronyms), repeat information, and explain verbally and in writing,
  - Provide enough time for patients to digest information and ask questions (ensure staff support at screening).

Washout Period

- Most suggested a recurring newsletter by email (i.e. bi-monthly):
    - Highlighting the progress in the project, how AIDAVA can support patients and thank the participants for their contributions

Study Schedule and study activities

- Overall, the schedule is clear
- Recommendations:
    - not to use any abbreviations, acronyms, complex language,
    - explain the schedule orally and visually to site patients
    - For most, 2 weeks seems like a realistic and fair timeframe for site patients to curate data (some considered it too short)

Onsite vs. Online training

- Most consider the distribution of online vs. onsite OK - some prefer onsite, others stressed the time and effort of onsite visits
    - Suggestion to provide the option of both at each stage for participants who cannot make onsite visits

Usability questionnaire

- Overall, the usability questionnaire is clear and easy to answer
- Recommendations:
    - To change the scale (1-5) so that participants cannot answer neutrally and must consider a more positive or negative response
    - To include a free text box where site patients can elaborate on possible areas of improvements or issues encountered
- It is not possible to change the format of the SUS questionnaire as it is an official form.

Other recommendation include:

- Clearly indicate how long some task will take (i.e. task should take approx. x-minutes)
- Consider that some patients may find tasks harder than others and may not be comfortable expressing their issues
    - Suggested solution: Complete exercises/quizzes at the end of the training to check that the patient understands the training and are confident in using the system
- Confirmation of 3rd party app
    - Application to enter Quality of life questionnaire
    - Application to manage medical device
- Check Data Governance aspects with AIDAVA Data Protection Officer (IHD) and ensure alignment with data management plan and with form to perform local DPIA [reference from AIDAVA]
- Develop assessment questionnaires with REDCap
- Translate as needed for Ethical Committee approval

2.2 Development of Study Information Package and Informed Consent

The Study Information Package (SIP) and Informed Consent Form (ICF) form the basic material to be used when recruiting patients into our AIDAVA prototype assessment study.
The first draft of the SIP and ICF was completed in May 2023 and then sent for review by our patient consultants. It was also reviewed by our project´s data privacy officer from partner IHD.

Initial feedback on SIP based on questionnaire from patient consultants highlighted the following:

- "It is very important that words that are rare or hard to understand should be written more simply and more understandable.
  For example *CVD*, some people know what it is but also those kinds of acronyms should be written in full or explained.
  Also the words *curate* and *ingest* are hard to understand. "
- "Good that there is a list of what it is expected from the patients. This is important to be short and clear.
  How much time will it take in these 2 periods to participate in the study? 20 minutes- 8 hours/ day? Please try to specify. "
- "Good to have benefits and risks. It is important to be as clear as possible to the patients so they know what risk also can be so they are not afraid to be in this. The only area of concern was that risks are laid out in a negative format making it sound very risky; rather than saying any questions or complaints it was suggested to rephrase to lighter more positive wording such as any questions or concerns."
- "The text is really on many pages, and it can be hard to read everything for a patient. But all the text is so important that it should be there, and nothing can be taken away …Suggestion was to add in the end of the document also "most important to know" points as a summary ? It could help the patient to summarise everything they read.
  Also text was considered beautifully written but not simple to read in English. This complex writing style could be off putting to participants. It was suggested to make the first paragraph of the section on intro to AIDAVA simpler and easier to read English to get the key points across. "

The consolidated SIP and ICF will be translated into local languages by the sites, as part of the submission to the local ethical committees for approval (if required). It was agreed that a short overview summary of the SIP will be given (in a face-to-face conversation) to the patients by the treating physician during the recruitment phase when introducing the project. The longer version of the SIP will be available on paper for home reading and will be given to the patient if they would like additional information.
Amendments to the document might be made according to the feedback from local ethical committees.

2.3 Development of Training program

To ensure a proper evaluation, it is important to train the different participants: all the participants (including the patient if they are interested) must understand the purpose of the project and the assessment study, the user of the system must be trained to use the system, the research associated must be specifically trained to the different steps to be followed during the assessment, and the IT

supporting team must be trained to provide technical support. This requires a few training sessions that must be developed.
A training plan was therefore drafted; it includes an overview of the different training modules with the objective and content of each module, the estimated time to give the training, what media types are preferred, the target audience as well as the planning for developing the training module (and related materials).

2.4 Development of Data Sharing agreement

The patient data managed within AIDAVA and deployed within each hospital include data coming from the hospital itself as well as data coming from an HDI. It is therefore important to define a data sharing agreement between each hospital and the related HDI, including technical and legal provision. For transferring the data from the hospital to AIDAVA, a technical specification is sufficient.

2.4.1 Work with HDIs

In order to comply with data protection rules for data sharing health data, bilateral agreements were drawn up between HDIs (MIDATA and Digi.me) and hospitals hosting AIDAVA. These contracts specify the roles and responsibilities of parties and data handling to ensure security and compliance to regulations. This process was overseen by our project partner The European Institute for Innovation through Health Data (IHD) responsible for data protection and impact assessment.

In order to allow the possibility for patient to include data from different data sources to be added to AIDAVA, the HDIs and project partners helped to identify suitable candidates for a third party application for patients to fill questionnaires - EQ-5D quality of life questionnaire was selected for the project (“EQ-5D” n.d.). The HDIs helped to negotiate with possible partners, confirm that suitable candidates could be integrated within their system and to hold contractual discussions including technical details.

To provide patients’ the ability to include data available in different formats from other healthcare providers (other than the hospital partners in the project) , the HDIs will include the possibility for patients to upload pdf-s or get data directly from a national health data centre. Different options and solutions were considered and a variety of solutions were chosen to investigate different formats being integrated to the patients data in AIDAVA.

In the cardiovascular use case, the patients will have the opportunity to collect and add their data via a blood pressure monitoring device. The HDIs helped to oversee and confirm which physician approved and CE marked devices (which indicate that a product has been assessed by the manufacturer and deemed to meet EU safety, health and environmental protection requirements) are compatible with their system or which further developments are required to achieve this.

Lastly, it was agreed that after the curation process the HDIs would provide the patient the option to view or own the curated Personal Health Knowledge Graph and display the International Patient Summary of the patient to show a value of AIDAVA from the patient point of view.

All the solutions were also introduced to and reviewed by the patient consultants in the project who then provided their feedback (E.g. on HDI agreement for patients, overview of options that will be

included) and recommendations (E.g. to be more clear in SIP materials on the role of the HDI-s and why they are separate than AIDAVA, etc.) throughout the process.

The work on integrations will continue in Task 1.5 where the HDIs will test all these added features within their systems.

2.4.1 Requirements (security, environment) for testing the prototype in each organisation.

In order to cover all data processing security concerns and legal requirements, separate meetings were held with participating sites and HDIs and the results were compiled into the template Data Sharing Agreement which covers the scope of the data exchange, legal and data privacy provisions and technical provisions. This includes an overview of the data transfer specification principles, processing and security aspects for patient data during testing and initial technical architecture.

3 Results & Discussion

This task delivered 4 documents needed to execute the assessment study and provided in Annex.

- **Annex 1. Study Protocol.**
  The complete overview on how to assess the AIDAVA prototype is provided in the form of a study protocol (Annex 1 of the current document). It gives detailed descriptions of actions and supporting materials from preparation stages to evaluation. The study protocol was developed with project partners together with the clinical sites and their clinical staff, Health Data intermediaries and reviewed by patient consultants. The work on the protocol lasted from February to November 2023 and took multiple iterations to be complete. Changes can still be implemented to the protocol in case local ethical committees request it.
- **Annex 2. Study information package and Informed Consent Form ICF**
  This includes the basic information leaflet to be translated in local language and to be provided to the patient being recruited, before asking them to sign the ICF. The ICF is a one page document with check points.
- **Annex 3. Training plan**
  This document include the different training modules to be developed and provided to the patient and the study team ; it also includes a training program based on user profile
- **Annex 4. Data Sharing agreement**
  This document was developed in collaboration with D4.4; it includes the identification of the different data flows (to and out of the hospitals), link to legal provisions to be agreed across the different partners as well as the description of the data transfer. The detailed Data Transfer Specification are still being finalised by each site and will be completed for insertion in the Data Source Catalogue (D3.5).

4 Conclusion

This deliverable clarifies the development of study protocol, describes the study process and how health data curation and data visualisation will take place. The development included all the parties that take part in the study (clinical sites, HDI-s, Patient consultants, developer of AIDAVA, IHD for data security). The results of this deliverable and its Annexes form the basis for Ethical approval forms in each site to start preparing for testing.

5 Next steps

This deliverable - and the associated material gathered during Task 1.4. - is the basis of the assessment study for generations 1 and 2 of the prototype of the project, as part of Task 1.5, to be executed after approval by the respective Ethical Committees.

We expect that the Ethical Committees will have comments; these comments will be captured across sites, discussed with the project team and adaptation will be made to the common documents provided in attachment with a change log containing the questions of Ethical Committees across site and the answer to these questions. Each site will then be responsible to adapt their local material and resubmit to the local Ethical Committees.

6 Annexes

Annex 1. Study Protocol

Annex 2. Study information package for the patients with ICF

Annex 3. Training modules and training program based on user profile

Annex 4. Data Sharing agreement (developed in collaboration with D4.4) and Data Transfer Specification

7 References

"Common Protocol Template Now Available." 2016. TransCelerate. January 21, 2016. https://www.transceleratebiopharmainc.com/news-center/common-protocol-template-now-available/.

"EQ-5D." n.d. Accessed November 10, 2023. https://euroqol.org/.

"MedMij Framework." 2021. MedMij. December 22, 2021. https://medmij.nl/en/medmij-framework/.

"System Usability Scale (SUS)." 2013. Department of Health and Human Services. October 4, 2013. https://www.usability.gov/how-to-and-tools/resources/templates/system-usability-scale-sus.html.

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

Deliverable No. D1.4
Annex 1 - Study Protocol

Contractual Submission Date: 30/11/2023

Actual Submission Date: 30/11/2023

Responsible partner: P8- NEMC

Funded by
the European Union

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

Deliverable number	**Annex 1 to Deliverable D1.4**
Deliverable title	**D1.4. Description of Assessment Study (and supporting material)**
Type[1]	R
Dissemination level[2]	SEN
Work package number	WP1
Work package leader	P8 - NEMC and P2 - b!loba
Author(s)	Isabelle de Zegher (P2-b!loba) Katrin Lepik, Eno-Martin Lotman, Kerli Norak, Mall Maasik (P8-NEMC)
Reviewers	Nathan Lea (P6-IHD) Monika Kindslehner, Eva Valentina Klocker, Martin Benedikt (P7-MUG) Petros Kalendralis, Bart Scheenstra, Liesbeth Boersma, Rutger van Mierlo (P1-UM) Dominik Steiger (P13-MID) Hanne Muller (P14-DME)
Keywords	Assessment study, use-cases, CVD, Breast cancer registry

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).
The work of MID received funding by the Swiss State Secretariat for Education, Research and Innovation (SBFI), subvention contract 22.00093, REF-1131-52104.

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

Document History

Version	Date	Description
V0.1	31/06/2023	Final version for discussion by Principal Investigators/ clinicians
V0.2	15/10/2023	Final version for translation
V1.0	30/11/2023	Final version

[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

V2.0	05.07.2024	Answers to reviewers comments ● Comment 2: " *Additionally, one of the main requirements for this reporting period is to provide clear and concrete lists of the data elements that will be collected from the 45 patients (15 per data provider) enrolled in both proposed pilots. These data elements should specify how the data will be codified with interoperability standards. Among the defined data elements, identify those that may require human interaction, specifying whether the human is a patient, data steward, data expert, or healthcare professional.* The list of data DE of the breast cancer registry has been added as Section 10.5 of the Annex 1 - Study Protocol; the list of DE for the CV use case is included in Section 3.2.2 of the same Annex in the table specifying how the SMART risk score should be computed. Both lists include the relevant SNOMED or LOINC code. The objective of AIDAVA is to generate these DE automatically from one or more Personal Health Knowledge Graph as explained in *Deliverable D1.1. - Description of the use cases* (see Figure 5 in Section 4.1 and Figure 6 in Section 4.2)

Table of Contents

Foreword

This document is an annex of *Deliverable 1.4. Definition of assessment study including test scenarios & metrics, and study initiation package.* It is a template, in English, of the Study Assessment to be executed for the 2 use cases of the 2 generations of the AIDAVA prototype across the 4 hospital sites (NEMC in Estonia, MUG in Austria, Maastro and MUMC in The Netherlands) supported by one health data intermediary (MIDATA for Estonia and Austria, DIGI.me for The Netherlands).
It is based on the TransCelerate electronic protocol template[3] and the ICH M11 guideline[4] (clinical study protocol template and technical specifications). It builds on input provided by the breast cancer and cardiovascular physicians and on additional discussion with the AIDAVA project team. It includes in annexes, RedCap based forms that will be used to collect data for further analytics demonstrating the performances of the prototype. The protocol, and more specifically the schedule of activity, has been validated, by the AIDAVA patient consultants.
This template might need to be adapted - including translation - to fulfil the requirements of the local Ethics Committee in each evaluation site.

This template might need to be adapted - including translation - to fulfil the requirements of the local Ethics Committee in each evaluation site.

List of definitions

The definitions used in the deliverable are based on the AIDAVA Glossary [ref].
Key definitions for this document

Name	Definition
SITE Administrator	Responsible for configuring the solution at local level ● Execute onboarding of local data source in data catalogue ● Register and manage local users (based on template user profile) ● First level support
Research Associate (RA)	Can be a nurse, a data steward, a data scientist, a researcher, project manager or a team. Responsible for ● Delegate of the Principal Investigator ● Contact point for the project locally ● Inform and train the patients ● Gather evaluation information from the patient and the curator
Patient	As data curator, the patient ingests their personal data from the different source systems, initiates the curation process and answers questions from the system whenever needed. As data consumers, the patients request to forward their personal data to the local HDI for further visualisation. *Note: Before using the AIDAVA prototype, the patients need first to sign an*

[3] https://www.transceleratebiopharmainc.com/wp-content/uploads/2021/10/CPT_CoreTEE-v009.dotx
[4] https://www.ema.europa.eu/documents/scientific-guideline/ich-m11-technical-specification-step-2b_en.pdf

Name	Definition
	Informed Consent Form (ICF) by which they agree that the system will access and process their personal data, and that curators will also have access to their personal health data. This ICF also allows the patient to stop using the system at any point in time. In this case, their personal data included in the system will be either transferred to them or deleted.
Expert Curator	The expert curator (also called data steward in some site) supports patients in the curation of their data, with the consent/agreement of the patient
Data User	Data Stewards, health care provider, scientific staff having access to extract published from AIDAVA (with patient consent) ● Breast Cancer specialist who can perform analytics on a "Breast Cancer" registry spread across the 3 sites ● Cardiovascular specialist who has access to an automatically computed risk score (instead of having to compute it manually) and can more effectively monitor CVD risks.
Screening	The process of active evaluation of potential participants for enrollment in a trial. Screening occurs during the enrollment period to see if they meet the inclusion and exclusion criteria. If they meet the criteria, the subject is eligible to be enrolled in the trial. [1]
Recruitment	Recruitment is the process used by investigators to enrol people (participants) into a clinical study. Recruitment is based on the inclusion and exclusion criteria that are documented in the study protocol [2]
Enrollment	The process of registering or entering a patient into a clinical trial. Once a patient has been enrolled, the participant would then follow the clinical trial protocol. Clinical investigations are designed to enrol a set number of participants to increase the likelihood of answering the trial questions [3].
Data collection	Systematic gathering of data from various sources, such as surveys, interviews, observations, or existing databases. Data collection focuses on obtaining the necessary information required to address a research question, study, or analysis.
Data capture	Process of entering or recording data into a digital system or physical format. Data capture can be manual, where individuals transcribe data from paper forms or other sources, or automated, where data is directly entered into a computerised system through devices like scanners, sensors, or electronic forms.
Data ingestion	Process of importing, receiving, or acquiring data from various sources - such as databases, files, APIs, streaming platforms, or other data streams - into a data storage or processing system for further analysis or utilisation.

List of abbreviations

AI	Artificial intelligence
ASCVD	Atherosclerotic Cardiovascular Disease
BP	Blood pressure
CABG	Coronary artery bypass graft
CAD	Coronary artery disease
CT	Computed tomography
DM	Diabetes mellitus
ECG	electrocardiogram
EHR	Electronic health records
G1	Generation 1 (of prototype)
G2	Generation 2 (of prototype)
HDL	High density lipoprotein
HDI	Health Data Intermediary
HS-CRP	High sensitivity c-reactive protein
ICF	Informed Consent Form
IPD	International Patient Summary
LDL	Low density lipoprotein
MD	Medical Doctor
MRI	magnetic resonance imaging.
PAD	Peripheral Arterial Disease
PCI	Percutaneous coronary intervention
PEM	Patient experience monitor
PHD	Personal Health Data
PHKG	Personal Health Knowledge Graph
PROM	Patients reported outcomes measures
RA	Research associate
SIP	Study Information Package
STEMI	ST-elevation myocardial infarction

1. Study Summary

1.1. Synopsis

Study Title	Assessment study of the AIDAVA prototype which aims to improve the quality of clinical data, for their reuse in clinical care in patients with established cardiovascular diseases, and in clinical research for breast cancer research.
Rationale	The AIDAVA intelligent virtual assistant is a prototype assisting device[5] that supports patients to curate and publish their personal data in a secure environment, generating an interoperable and reusable personal medical record that can be used in clinical care to improve patient outcomes and to support data driven clinical research. The assessment study will help to determine whether the AIDAVA intelligent virtual assistant is 1. effective in **improving data curation and publishing process** against existing practices, 2. usable and acceptable **by patients** - with different levels of health, digital and data literacy - **and expert data curators**, whenever their input is required, 3. **valuable for "data users"** such as cardiovascular treating physicians and breast cancer clinical researchers. The efficacy of AIDAVA virtual assistant will be influenced by the quality of the different data curation tools (supporting automation) and by the explainability module (managing the human computer interaction). The study will help identify how these components impact the effectiveness of the device. The results of the study will be used to inform the exploitation plan of the prototype, and the needed improvement to bring a final device to the market.
Overall Design	Three arms, prospective cohort-study ● involving 30 patients per site, 15 Breast Cancer (BC) and 15 cardiovascular (CVD) + 3 experienced data curator per site ● based on already collected health data, available across different data sources ● with 3 arms: one for patients assessing the prototype across the 2 generations[6], one for the patients assessing the prototype only for a single generation, and one for the expert curators (assessing the prototype across the 2 generations) ● across 4 epochs (see Figure 1) ○ Epoch 1: 1 week duration, takes place just before deployment and evaluation of Generation 1 (G1) of the prototype; it will allow to gather information from the curator on the current situation. ○ Epoch 2: 2 weeks duration per patient during period of July - September 2024: assessment of G1, by ingesting personal health data into the prototype and curating them

[5] A prototype does not need to go through MDR certification ; at the end of the project the idea is to develop a product medical device that will require MDR.

[6] The AIDAVA prototype will be delivered through 2 generations or releases: Generation 1 (G1) by end Q2 2024 and Generation 2 (G2) by end Q4 2025.

	○ Epoch 3: 16 to 18 months period, "wash out" period during which Generation 2 (G2) of the prototype will be developed; regular information will be provided to the subjects ○ Epoch 4: 2 weeks duration (in February/March 26); assessment of G2, by performing the same task than in Epoch 2 to enable comparison of the performance of the 2 generations of the prototype After enrollment, CVD subjects will receive configuration of the blood pressure medical device and receive initial training; additional training updates and filling in assessment questionnaires will take place through video conference or on site depending on the patients´ preference.
Number of Participants	● Approximately up to 50 patients per site/ per therapeutic area will be screened for eligibility and given a short information session introducing the study; 15 eligible patients who agree to participate will be enrolled ● 3 expert data curators will be assigned per site, based on the current roles
EC and DMC	Each site has a responsible Ethical Committee that will review the protocol, adapted to local requirements and translated. Taking into account that the device has no direct medical impact, there will not be any data monitoring committee. However the device to be used will be subject to a data protection impact assessment and codes of practice developed by AIDAVA and overseen by its Ethics Advisory Board.

End points

Objective	Endpoints
Primary: Comparison of efficiency without AIDAVA (G0), with G1, and with G2, assessed by the following metrics	
Metric 1. Evaluate acceptance by end users	System Usability Scale (user answer to questionnaire)
Metric 2. Measure impact on workload for quality enhancement of data	● Average time spent in curating needed data elements (user entry in dedicated form) ● Amount of additional data curated in the average time (automatically computed by system)
Metric 3. Measure quality of data resulting from the data curation & publishing - in terms of their interoperability and reusability	Score (computed automatically) measuring components related to ● volume (number of data items curated) ● completeness (number of missing items) ● consistency (number of consistency checks that fail) ● availability of context information (metadata)
Metric 4. Assess the quality/quantity of concepts that can be extracted from clinical narratives with different technologies	● Precision (M4.1): fraction of correct gold standard instances among the retrieved instances ● Recall (M4.2): fraction of the gold standard concepts that are successfully extracted

Secondary - Breast Cancer use case	
Demonstrate the ability to perform the exact same computation in each site and to consolidate across sites, showing interoperability of data curated from different sources and ability to build a federated registry.	• Metric 1: Percentage of patients - across sites - treated with breast conserving therapy and whole breast radiotherapy that receive a boost to the tumour bed. • Metric 2: In patients - across sites - undergoing surgery and radiotherapy, without adjuvant chemotherapy: % of patients with time between surgery and radiotherapy > 8 weeks. • Metric 3: Percentage of In patients - across sites - receiving nodal radiotherapy out of any radiotherapy.
Secondary - Cardiovascular use case	
Demonstrate the value of having interoperable and reusable data in a standard format to provide information of value in clinical care - without additional burden to the physician	• Metric 1. Compare scoring time between AIDAVA and CVD experts • Metric 2 . Compare scoring accuracy between AIDAVA and CVD experts

1.2. Schema

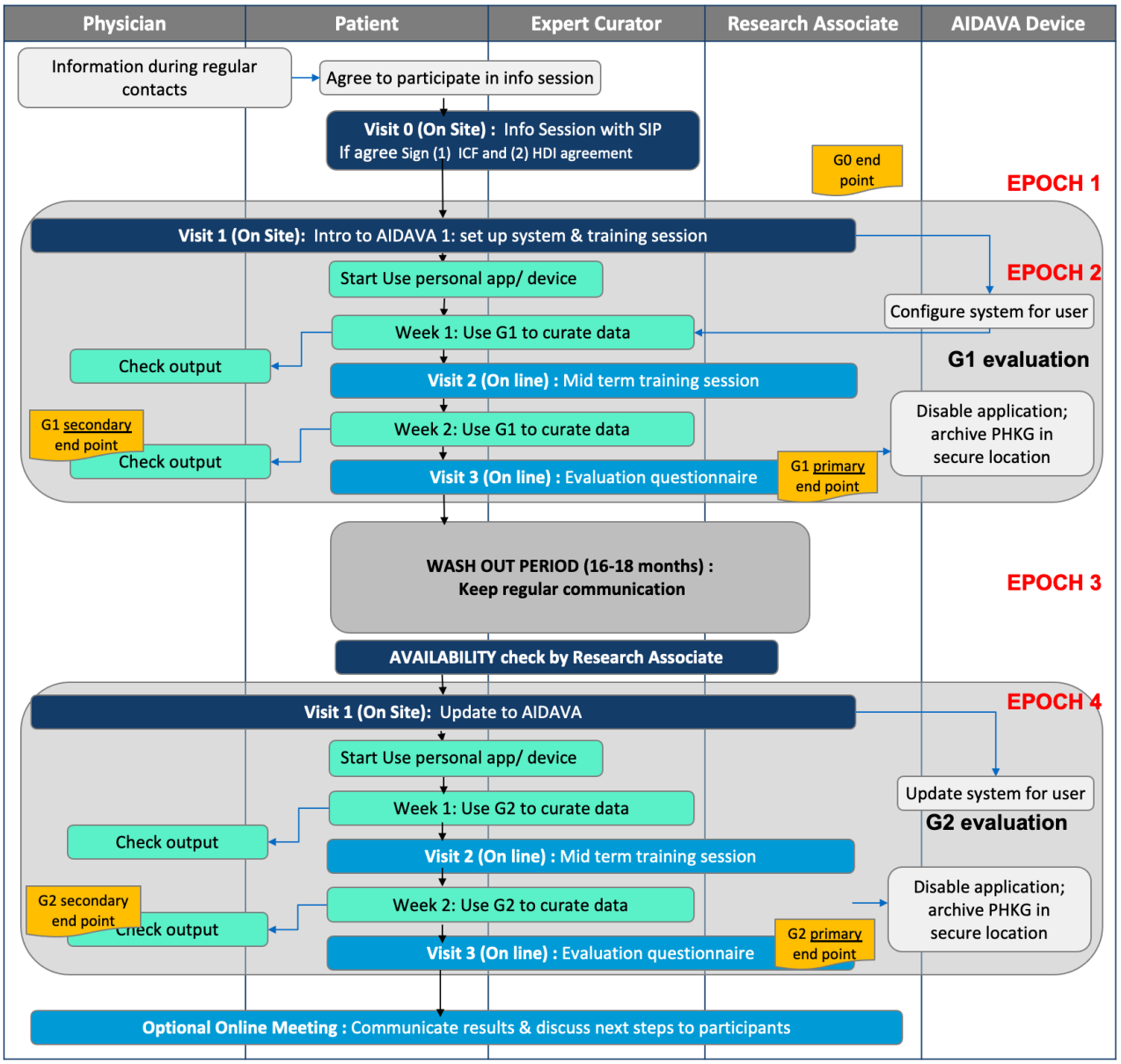


Figure 1. Schema of the assessment study

1.3. Schedule of Activities (SoA) - see more details in Section 4.

Procedure	Screen (< 15 d before W1)	Intervention Period Weeks (with a washout of 16 to 18 months)																		Notes
		Generation 1								Develop G2	Rescreen (< 15 d before W1)	Generation 2								
		W1 D1 V1	W1	W1 D5	W2 D1 V2	W2	W2 D5 V3	W3	W4 V4			W1 D1 V1	W1	W1 D5	W2 D1 V2	W2	W2 D5 V3	W3	W4 V4	
Check I/E criteria (MD)	xBC	xCVD								(16 to 18 months)	(x)	xCVD								REDCap Form 3
Intro to study (RA)	xBC	xCVD									(x)									SIP (Annex 2)
Sign Informed consent	xBC	xCVD									(x)									On paper
Sign HDI agreement	xBC	xCVD									(x)									On paper
Create HDI account (through AIDAVA)	xBC	xCVD									(x)									
Data extract Hosp/HDI	xBC	xCVD																		After signing ICF
Training the research team	x																			Hospital contact persons, RA, expert curator, MD
Training - on site		x										x								
Training - on line					x										x					
User profile - Health and digital literacy		x										x								REDCap Form 1 (Patient, Curator)
Site monitoring forms		x	x	x	x	x	x	x	x			x	x	x	x	x	x			REDCap Forms 4,5,6,7 (RA)
System Usability Scale							x										x			REDCap Form 8 (Patient, Curator)
Fill in QALY		x										x								Personal Apps
Use BPM device		xCVD	xCVD		xCVD	xCVD				xCVD	xCVD	xCVD		xCVD	xCVD					See Section 6.2

Procedure	Screen (< 15 d before W1)	Intervention Period Weeks (with a washout of 16 to 18 months)																	Notes
		Generation 1								Develop G2	Rescreen (< 15 d before W1)	Generation 2							
		W1 D1 V1	W1	W1 D5	W2 D1 V2	W2	W2 D5 V3	W3	W4 V4			W1 D1 V1	W1	W1 D5	W2 D1 V2	W2	W2 D5 V3	W3	W4 V4	
Curate data with AIDAVA		x	x	x	x	x	x					x	x	x	x	x	x			
Curator form - patient							x										x			REDCap Form 11 (Patient)
Curator form - expert	G0						x										x			REDCap Form 9, 10 (Curator)
IPS extract				x			x							x			x			
Extract BC registry				xBC			xBC							xBC			xBC			
Query Federated BC reg. /sec. endpoints				xBC				xBC						xBC				xBC		
SMART risk score (CVD)				xCVD			xCVD							xCVD			xCVD			Manual & auto
Data user form: BC, CVD							x										x			REDCap Form 12,13 (BC), and 14 (CVD)
Feedback to Patient (results + regular info)									x	every 3 mth									x	
Confirm delete data (if dropping out)									x										x	

Figure 2. Schedule of Activities

x = all patients

(x) = only for patients who did not participate to G1 evaluation

xBC = Breast Cancer patient only

xCVD = Cardio-vascular patients only

2. Introduction

The **AI**-powered **Da**ta Curation **V**irtual **A**ssistant (AIDAVA) project is a 4 year Horizon Europe project (www.aidava.eu). It includes 14 partners (12 EU partners and 2 associated partners). It started in September 2022 and will conclude in August 2026.
Its objective is to develop and test a prototype virtual assistant that will help patients to **curate & publish**[7] their personal health data, coming from hospital systems and from Health Data Intermediary (HDI)[8] mediating transfer of GP data, patient health apps and wearables (see detailed list of data sources in Section 10.1. Data Sources). The prototype will maximise automation in the curation and publishing process of personal health data by orchestrating multiple AI technologies; it will request input from the patient - at their level of health and digital literacy - or from expert data curators when automation is not possible. The system will direct the question to the patient or to the expert curator, based on the type of question and the level of experience of the patient.

The end goal is to support patients - or their delegates - to manage all their health data into a longitudinal personal health record, that is interoperable and reusable for multiple purposes, clinical care as well as clinical research. AIDAVA is also intended to support expert curators - within hospitals - to clean and curate hospital's patient data, for further reuse.

There will be 2 generations of the prototype.

1. Generation1 (G1), to be deployed by June 2024, will be orchestrating the use of existing tools, open source tools or licensed products.
2. Generation2 (G2), to be deployed in December 2025, will improve G1 by integrating novel tools developed in the project. This includes NLP tools - extracting structured data from clinical narratives in local language - and an human-AI module to increase understanding and usability by users with limited health and digital literacy.

To demonstrate the value and the performance of this prototype, we will install these 2 generations of the prototype in a testing environment within the hospital and will perform the assessment study described in this document.
As the system is a prototype, no part of it (including any AI algorithm included in the system) will be used for clinical decision making and cannot be used either as a decision support system for clinicians or as the basis of clinical research.

2.1. Study Rationale

The purpose of the assessment study is to verify that the AIDAVA prototype virtual assistant meets the objective set for this device at the onset of the project.

1. Improve the quality, FAIRness, and portability of heterogeneous, Personal Health Data (PHD)
2. Increase value and reusability of PHD through integration and semantic enrichment within a common standard representation - called the Personal Health Knowledge Graph (PHKG).

[7] By "***data curation & publishing***" we mean the integration, harmonisation and quality enhancement (curation) and the transformation into a target format (publishing) of multimodal data, collected from various sources, to make it more usable by humans and machines.

[8] Health Data Intermediaries are emerging organisations, regulated under the European Data Governance Act (DGA), enabling personal health data sharing on behalf of each consenting patient.

3. Optimise the data curation & publishing process.
4. Ensure compliance with EU ethical & data privacy requirements when processing personal data.
5. Demonstrate the value of the developed novel tools for more effective treatment (Use Case 1 - breast cancer registries federated across hospitals) and for preventive care (Use Case 2 - longitudinal health record of cardiovascular patients at risk of sudden cardiac arrest).

2.2. Background

The AIDAVA virtual assistant prototype developed during the project will be deployed in a secure testing environment (see light green box below) within the 3 participating hospitals; each hospital will be working with a dedicated Health Data Intermediary (orange box on the left).

The primary objective is to support patients and expert curators[9] in managing the patient longitudinal health record, and to allow patients to visualise their record. The secondary objective is to demonstrate that interoperable and reusable longitudinal health records can be reused (with consent of the patient) to effectively support clinical care and clinical research.

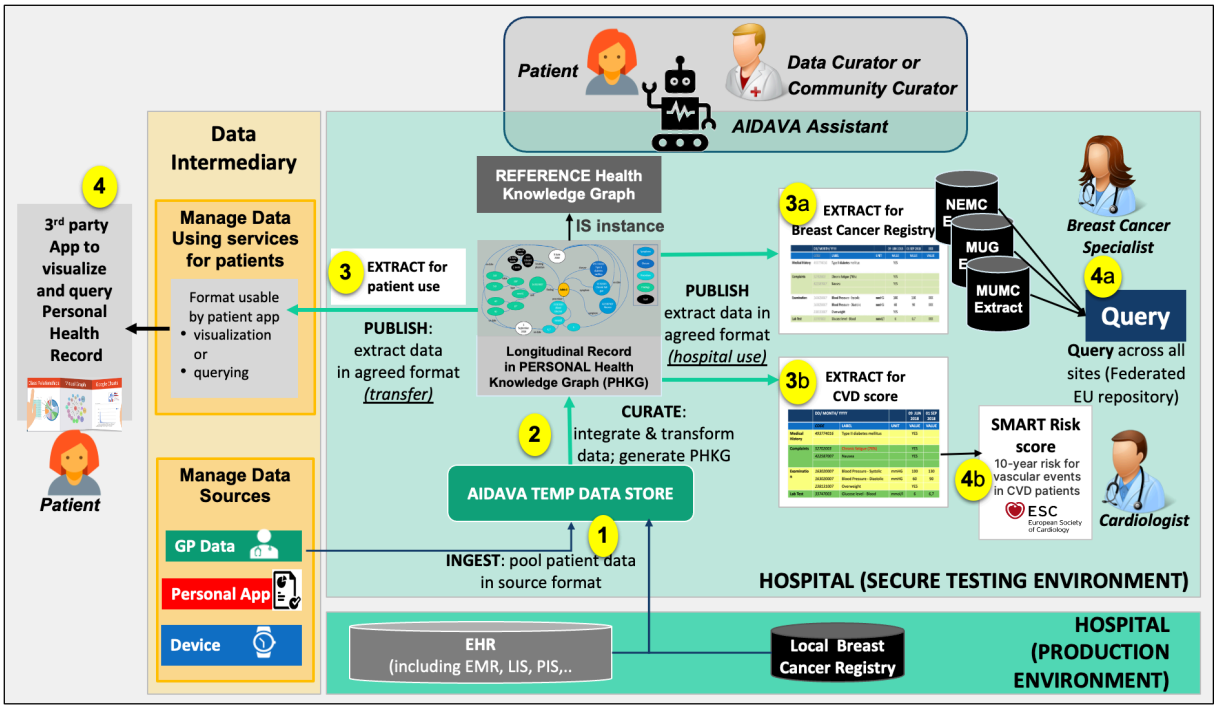


Figure 3. Components of the AIDAVA prototype and data flows

- *Flows in greens represent internal flows within AIDAVA (see below)*
- *Flow and bullet in yellow represent external flows to AIDAVA, subject to Data Sharing/Processing Agreement with Data Transfer Specifications (see Section 11.4 on Data Protection)*

Patients and data curators are expected to work through the following steps.

- **Step 1**. **Ingestion**. The patient - or data curator working as deputy of the patient - identifies data sources from the hospital EHR and from the patient HDI, and requests transfer of their

[9] See definition in Study Population - Section 5.2

personal data from these data sources into the AIDAVA TEMP DATA STORE available within the hospital testing environment. These data are stored in their original format. Extraction from the data sources can be done with the support of the local registry steward, or automated whenever possible; this will be further specified in the Data Transfer Agreements signed between the Health Data Intermediaries (HDI-s), managing the non-hospital data of the patient, and the hospital.

- **Step 2**. **Curation.** The patient requests the AIDAVA virtual assistant to transform, integrate, potentially correct, and complete ingested data to generate a standardised representation of its individual's health record, the Personalised Health Knowledge Graph (PHKG). The AIDAVA virtual assistant will maximise automation in curating the data by orchestrating execution of relevant data curation tools for each curation step, and requesting input only when needed from the patient - at their level of health and digital literacy - or from the data curator when it is expected that the patient will not be able to answer.
- **Step 3. Publishing.** The HDI responsible for managing the data of a patient, specifies the data that should be extracted from the PHKG to support visualisation and potential other processing through the selected 3rd party app. At the moment of writing the document, we intend to extract the data elements supporting the International Patient Summary (IPS) based on the HL7 FHIR profile. The specification of the extract can be done once and executed many times, whenever there is a refresh of the data sources and the PHKG. Once the extract's specification is confirmed, the AIDAVA virtual assistant extracts the data from the PHKG, performs the needed transformation to generate the target format required and transfers the data to the HDI.
- **Step 4. Use.** The 3rd party application, selected by the respective HDIs, will display the patient data and support querying. The fact that the applications used for visualisation are different, while working on the same underlying PHKG structure will demonstrate interoperability and reuse of the underlying data.

Clinical care providers and clinical researchers also benefit from the PHKGs curated by the patient and/or the data curator.

For breast cancer clinical researchers.

- **Step 3a. Publishing data to simulate delivery of a "EU" Breast Cancer registry.** To simulate how a tool like the AIDAVA prototype can support delivery and maintenance of a "EU" wide Breast Cancer registry, a list of data elements to be extracted from each participating centre has been established by the breast cancer specialists (see ref). This list is a first draft to demonstrate the concept; to deliver a real EU Breast Cancer registry, the list should be updated after validation by an international team of breast cancer specialists.

  The AIDAVA prototype will generate the same data extract - from heterogeneous data sources - in the specified format across the 3 participating sites.
- **Step 4a. Analytics of the extract across different sites.** To demonstrate interoperability of the resulting extracts, the prototype will execute a set of "federated" queries (see Section 3.2.1) i.e. queries run with local data across all 3 sites with consolidation of the results into one single answer. This enables the build a EU wide registry while avoiding transfer of data outside of the hospital.

For cardiovascular specialists, following patients.

- **Step 3b. Publishing data to compute the SMART risk score** . The AIDAVA Virtual assistant supports automatic extraction and transformation from the patient PHKG into the data

elements and format needed for computation of the SMART risk score (see Section 3.2.2). The prototype virtual assistant will generate the same data extract in the specified format across the 3 participating sites.

- **Step 4b. Analytics of the extract across different sites.** To demonstrate interoperability of the PHKG and of the resulting extracts, we will use the same SMART risk algorithm across the 3 sites and provide results to the cardiovascular specialists.

The fact that the system can generate 3 different extracts from the same PHKG (visualisation for the patient and potentially additional processing as specified in the HDI collaboration agreement, extract for breast cancer registry, extract to compute the SMART risk score), demonstrates a key principle of AIDAVA "**curate once, use many times".**

2.3 Benefit/Risk Assessment

Taking into account that the study is assessing a "software as a device" prototype system with no impact on clinical care or clinical research, the risks are limited while the anticipated benefits are major. The table below provides a list of benefits and risks for the different actors involved.

	Benefit	Risk
Patient (Breast Cancer and CVD)	• Manage/control their own data and are able to visualise and query their record • Thanks to high quality health record, access to better clinical care (preventive and personalised medicine) • Contribute to the greater good by sharing high quality data for research	If data are not secured properly[10], risk of fraudulent access to highly sensitive data (this risk will be mitigated when deploying the prototype) If the patient does not answer properly to a question of AIDAVA, this could introduce errors in the medical record (but there will full traceability of the information entered)
Expert data curator[11]	Decrease workload when preparing data sets for secondary use (registry or clinical care score)	Time to learn another system, with too much workload upfront Curator may introduce errors (does not answer properly to a question of AIDAVA), this could introduce errors in the curated record in AIDAVA (i.e. the personal health knowledge graph or PHKG) **but the original record will NOT be modified** and there will be full traceability of the information entered.
Breast Cancer Specialist (Use Case 1)	Faster and more efficient access to high quality data to support breast cancer research across EU	The risk that the curated data are not sufficient/of sufficient quality to make decisions (being mitigated by data quality module). *Note: As the system is a prototype, it cannot be used for clinical decision making or as a*

[10] Each site will perform a Data Protection Impact Assessment; requirements for the prototype being developed include data security principles in alignment with GDPR and EU best practices.

[11] See definition in Study Population - Section 5.2

	Benefit	Risk
		decision support system for clinicians.
Cardiovascular Specialist (Use Case 2)	Faster and more efficient access to clinical score for their patients, supporting more effective monitoring	The risk that the curated data are not sufficient/of sufficient quality to make decisions (being mitigated by data quality module) *Note: As the system is a prototype, it cannot be used for clinical decision making or as a decision support system for clinicians.*

3. Objectives and Endpoints

3.1. Primary endpoints

The **primary objective** of the assessment study is to compare the acceptance and the performance of the AIDAVA prototype versus existing practice, and between the 2 versions of the prototype. To this end, we propose to measure the following endpoints that will be recorded across sites.

Objective	Primary Endpoints
Metric 1. Evaluate acceptance by end users	Metric definition: Structured questionnaire (adapted from a System Usability Scale with additional questions such as ● I would recommend AIDAVA to my friend, colleague, or family member ● I am interested / ready to work with AIDAVA when available on the market ● I understand the purpose of data curation ● I am ready to spend the needed time to ensure proper data curation ● The curation process was perfect/acceptable/too long for me Measurement score (**M1**) through structured questionnaire (System Usability Scale - form) provided in REDCap
Metric 2. Measure impact on workload for quality enhancement of data (including quality of the different curation tools)	Metric definition ● Compare time spent in curating data in ○ G0 (current situation) ○ G1 (generation 1 of the prototype) ○ G2 (generation 2 of the prototype). ● Check the amount of data curated in the average time: are there more data being curated together with the ones needed to support the use case Measurement ● For estimated time spent in curating ○ (**M2.1**) User entry in dedicated "Curator form" (RedCap) For curators, when requesting an estimate for G0, it is important to ensure that curated data set have the same level of complexity as for G1 and G2 complexity: ■ Extract a sample (synthetic data ?) to be curated including several data sources; ■ Request existing registry stewards to provide a time to curate data up to insertion into registry ○ (**M2.2**) System check "active" time of user to answer a query ● Amount of additional data elements are curated automatically computed ○ **(M2.3)** How many data elements have been curated automatically by the systems ○ **(M2.4)** How many data elements have required human interventions and been successful answered by the patients (i.e. answer provided to the question)

Objective	Primary Endpoints
	○ **(M2.5)** How many data elements have required human interventions can not be answered by the patients and are sent to the curators ○ **(M2.6)** How many data elements send to the curators did not get appropriate answer to finish the curation process
Metric 3. Measure quality of data resulting from the data curation & publishing - in terms of their interoperability and reusability	Metric definition: Data quality score measuring components related to the quality of the personal health knowledge graph and supporting its re-use. This includes dimensions such as ● volume (number of data items curated), ● completeness (number of missing items), ● consistency (number of consistency checks that fail), ● availability of context information (metadata). Measurement: automatically computed by system at G1 and G2: ● System computes scores automatically ; formula are being defined in D4.6/Task 4.2 (due date Feb 2024)
Metric 4. Assess the quality/quantity of concepts that can be extracted from clinical narratives with different technologies	Metric definition: ● Computed quality of the NLP concept extraction against the manually annotated data (“gold standard”) ○ Precision (M4.1): fraction of correct gold standard instances among the retrieved instances ○ Recall (M4.2): fraction of the gold standard concepts that are successfully extracted Measurement 0. Requires a manually annotated gold standard 1. Measure at G0 (current situation): measure # of concepts extracted for registry, manually – in limited time frame (few minutes) 2. Measure at G1 and G2: Precision/Recall for three different NLP systems a. The minimal-effort translation-based system developed for G1 b. The tools with novel DL models developed for G2 c. Measuring generalisation ability to new languages (at G2) The expected outcome is that (b) > (c) > (a). We will measure how well models trained of any two of the three languages (German, Dutch, Estonian) can generalise to the third language as a proxy for how well the overall system can adapt to new languages.

3.2. Secondary endpoints

The **secondary objective** of the clinical evaluation study is to assess the value of a tool such as AIDAVA to support reuse of data for clinical research, by delivering a EU wide federated clinical registry, and for clinical care, by providing useful information to the treating physicians while decreasing their workload (time saving). The secondary endpoints will be recorded in each site in a dedicated spreadsheet (see model in Section 10.6) shared across sites.

3.2.1 Breast Cancer Use case

Objective. The measures described below have no scientific rationale; their intent is to demonstrate that heterogenous data curated across different sites and different countries have been "FAIRified" thanks to AIDAVA i.e. they can produce comparable - and interoperable - data sets across different sites, supporting de facto delivery of a federated cross site registry.

The metrics will be based on assessing time needed and accuracy for 3 queries to be performed on the BC registry. **Important to note that the queries need to have a consolidated answer ACROSS sites** to demonstrate interoperability**.**

Queries	Technical Specification of the query (all codes are SNOMED code matched with the list of Data Elements in the BC registry)
Metric 1: Percentage of patients - across sites - treated with breast conserving therapy and whole breast radiotherapy that receive a boost to the tumour bed.	**BC1 = P2/P1 WHERE** **P1** = Total number of patients 404684003 Clinical finding (finding) WITH 64368001 |Excision of part of breast (procedure)| AND 428923005 |Radiotherapy to breast (procedure)| **P2** = Total number of patients 404684003 Clinical finding (finding) WITH 64368001 |Excision of part of breast (procedure)| AND 428923005 |Radiotherapy to breast (procedure)| AND 445232009 |Boost radiation therapy (procedure)|
Metric 2: Percentage of patients - across sites - undergoing breast surgery and radiotherapy, without adjuvant chemotherapy AND with time between surgery and radiotherapy > 8 weeks	**BC2 = P2/P1 WHERE** **P1** = Total number of patients 404684003 Clinical finding (finding) WITH 392090004 Operation on breast (procedure)| AND 108290001 Radiotherapy (procedure) WITHOUT 367336001 |Chemotherapy (procedure)| BEGINTIME OF (367336001 |Chemotherapy (procedure)|) >= BEGINTIME OF (392090004 Operation on breast (procedure)|) **P2** = Total number of patients 404684003 Clinical finding (finding) WITH 392090004 Operation on breast (procedure)| AND 108290001 Radiotherapy (procedure) WITHOUT 367336001 |Chemotherapy (procedure)| BEGINTIME OF (367336001 |Chemotherapy (procedure)|) >= BEGINTIME OF (392090004 Operation on breast (procedure)|) AND TIME OF 108290001 Radiotherapy (procedure) >= TIME of (108290001 Radiotherapy (procedure) + 8 weeks)
Metric 3: Percentage of in	**BC3 = P2/P1 WHERE**

Queries	Technical Specification of the query (all codes are SNOMED code matched with the list of Data Elements in the BC registry)
patients - across sites - receiving nodal radiotherapy out of any radiotherapy	**P1** = Total number of patients 404684003 Clinical finding (finding) WITH 108290001 Radiotherapy (procedure) **P2** = Total number of patients 404684003 Clinical finding (finding) WITH 168522007 Radiotherapy for lymphatic irradiation (procedure)

The actual metrics/ endpoints can then be defined as follows.

Objective	(Secondary) end point
Metric 1. Time required to compute different queries of interest within the BC registry	Metric definition: Time to have an answer to the 3 queries (BC1, BC2 and BC3) with information from the BC registry - across sites Note 1: should include time of the BC specialist and supporting team (spending most of the time) Note2: should be the same between G1 and G2) Measurement: ● **TBCx.** User entry in dedicated "Data User form" (RedCap) for each of the queries
Metric 2: Accuracy of measurement	Metric definition: accuracy of the answer for each of the parameter between AIDAVA and the BC specialist Measurement: For each of the query (BC1, BC2, BC3), ● **HBCx**. Ask the value from the human (through RedCap Form) ● **ABCx**: get the value computed by AIDAVA ● **ACCx**: accuracy of calculation by AIDAVA (ABCx) compared to BC specialist (HBCx)

3.2.2 Cardiovascular Use case

Objective. The objective of the endpoints below is to demonstrate the value of having interoperable and reusable data in a standard format to generate valuable and high-quality information in clinical care, without additional burden to the treating physicians.

Objective	(Secondary) end point
Metric 1. Compare scoring **time** between AIDAVA and CVD expert	Metric definition: Time to compute the SMART score (see definition below) Measurement:

Objective	(Secondary) end point
	TCVD. Time required for the treating physician to compute the SMART risk score (collect + analyse + enter data into a standalone web page calculator) - per estimate of user in dedicated entry form - compared to the time needed to calculate the same score by AIDAVA, automatically (G1 and G2)
Metric 2: Compare scoring **accuracy** between AIDAVA and CVD expert	Metric definition: Accuracy of automatic calculation of the U-Prevent SMART risk score compared to cardiovascular expert-entered score calculation (% difference in the score output - per estimate of user in dedicated entry form) (G1 and G2) Measurement: ● **CVD1**. U-Prevent SMART score computed by treating physician (https://u-prevent.com/calculators/smartScore) ● **CVD2**. U-Prevent SMART score computed by AIDAVA

The SMART Risk Score [4] is a tool to estimate 10-year risk for recurrent vascular events in patients with manifest cardiovascular disease. The calculator is located at the U-Prevent site [5], a clinical decision support platform, for the application of risk prediction models in clinical practice; the calculator has been recreated and optimised by ORTEC for personalised cardiovascular risk management. The U-Prevent site is CE marked according to the Medical Devices Directives (MDD); marking for MDR compliance is in progress. The pilot site was built by UMC Utrecht. Using the calculators is free. The U-Prevent team develops extra services for premium users, optimization of data management, e.g. by FAIR and GDPR-compliant data exchange from electronic health records. Technical specification to compute Model A, the SMART Risk Score [6] can be found below.

Formulas	SNOMED / LOINC Code (available in PHKG)	Population[12] value if missing
Recalibrated SMART risk score for patients with CVD: 10-year cardiovascular disease risk (%) = (1- 0.71840 exp[A + 1.933]) x 100% where A =		
-0.0349602236 x age in years + 0.0005510715 x (age in years)^2	Compute from SNOMED: 184099003 |Date of birth (observable entity)|	Mandatory
+ 0.2876587433 [if male]	Answer is Y in case of presence of SNOMED: 248153007 |Male (finding)|	Mandatory
+ 0.3455832714 [if current smoker]	A= last occurrence of smoker SNOMED: << 77176002 |Smoker (finding)| B = last occurrence of Quit-Smoking (<< 160617001 |Stopped smoking|) OR Ex-Smoker (<< 8392000 |Non-smoker|) IF time of A is after B,	Mandatory (default = N)

[12] Numbers from Dutch population (baseline characteristics from the SMART study); it is valid for male/female and all ages. The same number can be used across the EU population.

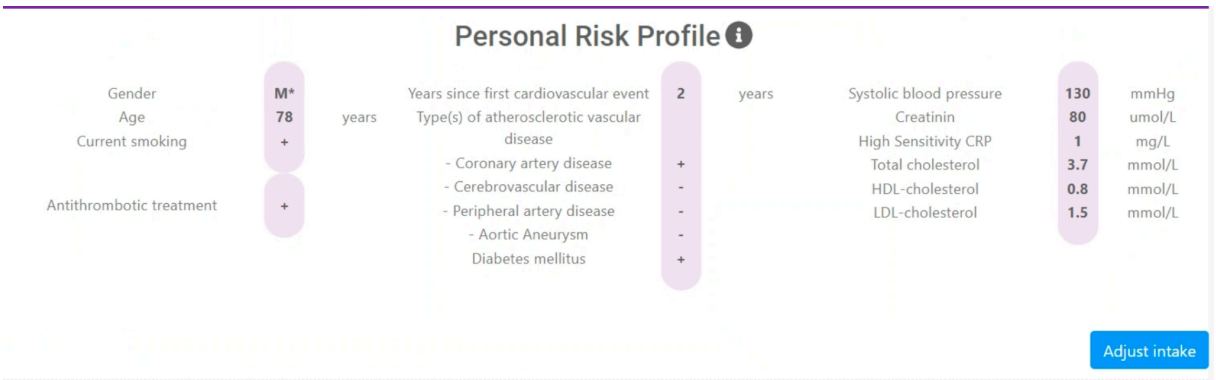
Formulas	SNOMED / LOINC Code (available in PHKG)	Population[12] value if missing
	THEN value is Y ELSE value is N	
+ 0.0018913154 * systolic blood pressure in mmHg	Take last value in PHKG ; SBP= LOINC: 8480-6 Note: in G2, we could improve with longitudinal BP (average value across a period of time)	Mandatory
+ 0.3181706587 [if diabetic]	Answer is Y in case of presence of SNOMED: << 73211009 |Diabetes mellitus (disorder)|(***and all children codes***)	VAL = 17% * 0.223 (average)
+ 0.2947019539 [if history of coronary artery disease, of acute coronary syndrome, myocardial infarction or coronary revascularization]	Answer is Y in case of presence of ● SNOMED: << 53741008 |Coronary arteriosclerosis (disorder)| (***and all children codes***) ● **OR** SNOMED: << 394659003|Acute coronary syndrome (disorder)| (***and all children codes***) ● **OR** SNOMED: << 22298006|Myocardial infarction (disorder)| (***and all children codes***) ● **OR** SNOMED: <<415070008 | Percutaneous coronary intervention (procedure)>> (***and all children codes***) ● **OR** SNOMED: <<232717009 | Coronary artery bypass grafting (procedure) >> (***and all children codes***)	Mandatory (default = N)
+ 0.3483178604 [if history of cerebrovascular disease, TIA, brain infarction, amaurosis fugax, retina infarction or carotic surgery]	Answer is Y in case of presence of ● SNOMED: << 230690007 |Cerebrovascular accident (disorder)| (***and all children codes***) ● **OR** SNOMED: << 266257000|Transient ischemic attack (disorder)| (***and all children codes***) ● **OR** SNOMED: << 432504007 | Cerebral infarction (disorder)| (***and all children codes***) ● **OR** SNOMED: << 88032003 | Amaurosis fugax (disorder)| (***and all children codes***) **OR** SNOMED: << 175362007 | Carotid and/or cerebral and/or subclavian artery operations (procedure)| (***and all children codes***)	Mandatory (default = N)
+ 0.3303566308 [if abdominal aortic aneurysm, history or presence of supra- or infrarenal aneurysm of the aorta > 3 cm, or aortic surgery]	Answer is Y in case of presence of ● SNOMED: << 233985008 |Abdominal aortic aneurysm (disorder)| (***and all children codes***) ● **OR** SNOMED: << 67362008 |Aortic aneurysm (disorder)| (***and all children codes***) ● **OR** SNOMED: << 32907006 | Operation on aorta (procedure) | (***and all children codes***)	Mandatory (default = N)
+ 0.2244665798 [if peripheral artery disease, history or presence of claudicatio intermittens, lowered ankle-arm index or surgical intervention (like angioplasty,	Answer is Y in case of presence of ● SNOMED: << 840580004 |Peripheral arterial disease (disorder)|(***and all children codes***) ● **OR** SNOMED: << 63491006 | Intermittent claudication (finding)| (***and all children codes***) ● **OR** SNOMED: << 446841001 | Ankle brachial	Mandatory (default = N)

Formulas	SNOMED / LOINC Code (available in PHKG)	Population[12] value if missing
stenting, endarterectomy, vascular bypass or amputation)]	pressure index (observable entity)| (***and all children codes***) ● **OR** SNOMED: << 387713003 | Surgical procedure (procedure) (like <418285008 | Angioplasty of blood vessel (procedure)> OR <233434005 | Insertion of stent into vein (procedure)> OR <392031002 | Endarterectomy (procedure) > OR <116360008 | Arterial bypass graft (procedure)> OR <81723002 | Amputation (procedure)>| (***and all children codes***)	
+ 0.0476995851 x years since first diagnosis of vascular disease	years (YRS) = current year - Y1 (year of first diagnosis of any of the above 4 diagnosis groups) ● Y1 = smallest of A,B,C,D where ● A = year of first occurrence of coronary artery disease ● B = year of first occurrence of history of cerebrovascular disease ● C= year of first occurrence of abdominal aortic aneurysm ● D = year of first occurrence of peripheral artery disease	VAL= 1 year
+ 0.0016497342 x years since first diagnosis of vascular disease ^2	years = YRS as above	
+ 0.5403642493 x log(nonHDL)	nonHDL = total-cholesterol - HDL-cholesterol WHERE ● total-cholesterol in mmol/L = Last value LOINC: 14647-2 OR if not available VAL = 4.9 mmol/l ● HDL-cholesterol in mmol/L = Last value LOINC : 14646-4 OR if not available VAL = 1.2 mmol/l	For all lab value,i f the unit of measurement is different in the ontology that the one in the formula transform with unit needed
+ 0.0396752081 * eGFR² (in mL/min/1.73m² + 0.0002186126 * eGFR² (in mL/min/1.73m²)	Last value LOINC: 62238-1 If not available, derive from creat(SerumCreatine in micromol/l (LOINC = LP145994-2) (*per paper Based on Heart. 2013 Jun;99(12):866-72 & Circulation. 2016 Nov 8;134(19):1419-1429* eGFR in mL/min per 1.73 m2 = If patient is male (SNOMED: 248153007) THEN = 32788 * $creat^{(-1.154)}$ * $age^{(-0.203)}$ ELSE 32788 * $creat^{(-1.154)}$ *$age^{(-0.203)}$ * 0.742	VAL creatine = 92 micromol/l (to be transformed)
+ 0.1517601731 * log(High Sensitivity-CRP in mg/dL)	If the last value of LOINC: 30522-7 is above 15 then use the previous value (up to the time the value is below 15). If there is no value below 15, then take default	VAL = 2.2mg/l = 22mg/dL

Formulas	SNOMED / LOINC Code (available in PHKG)	Population[12] value if missing
+ 0.2107210313 * usingAnticoag	Answer is Y in case of presence of ● one or more drugs of category Platelet aggregation inhibitors excl. Heparin (ATC = B01AC) OR ● A combination of One or more drugs of category Platelet aggregation inhibitors excl. Heparin (ATC = B01AC) + one or more drug of category Direct factor Xa inhibitors (ATC = B01AF) Example: can be ● Ascal, aspirin or an other equivalent (AAS, carbasalate calcium) ● Dual antiplatelet therapy : ○ Ascal / aspirin + a P2Y12 inhibitor ○ Direct oral anticoagulants (DOAC) P2Y12 inhibitor ○ Ascal / Aspirin + low dose direct working oral anticoagulation (DOAC)	VAL = Y

In case of missing data, some parameters can use population average value as indicated in the last column. If a mandatory value is missing, the default value indicated in the last column will be used. Note: it is expected that birthdate, gender and systolic blood pressure will be available. The score will be computed by AIDAVA and by the treating (junior) physician. It will then be checked by the senior physician; in case of discrepancy between the score computed by AIDAVA and by the physicians, the senior physician will require explanations on how AIDAVA came to the different answers/codes needed to compute the score and decide the reason for discrepancy.

The senior physicians will be presented with a table with all the parameters used to compute the score (as displayed in the figure below) and be able to get information on how AIDAVA came to a certain value (which data source, which transformation).



The senior physician will then be able to decide on one of the following reasons for discrepancy.

Health record incomplete: the information was not available in the medical record, while it should have been – and therefore AIDAVA could not identify the right value.

- Error in AIDAVA: the information is rightly included in the medical record, but AIDAVA did not include this information properly
- Error from physician: AIDAVA is right and the junior physician evaluation is not correct.

To assess the longitudinal aspects of AIDAVA, the SMART score will be calculated 2 times (after G1 and ± 16 to 18 months later, at the end of G2 use year follow-up). The aim is to see if modifiable risk factors can be captured and risk can be calculated by AIDAVA automatically and reliably.

4. Study Design

4.1 Overall Design

The assessment study will be a cohort, 3 arms, prospective[13] study involving 30 patients (15 breast cancer and 15 CVD) + 3 expert curators in each site. It will be based essentially on available data, collected during clinical care (see Section 10.1 for detailed description of data sources), but will also include some prospective data from medical devices and Quality of Life questionnaire

There will be 3 arms that will allow comparison across these different populations of users.

- "G1 and G2" patients who contributed to evaluation of G1 and G2.
- "Single Generation" patients who contribute to the evaluation of only one generation of the prototype (G1 or G2)
- Expert data curators

There will be 4 epochs

- Epoch 1 will last 1 week and take place just before deployment and evaluation of G1; it will allow to gather information from the expert curator on the current situation/curation process and to provide a dedicated training to the curators.
- Epoch 2 will last for 2 weeks, and is expected to take place in the period of July - September 2024. Patients will be requested to ingest their data with the support of the prototype, and to curate them. Whenever needed, the data expert curator will be able to support them.
- Epoch 3 will last for 16 to 18 months: "wash out" period during which G2 of the prototype will be developed;
- Epoch 4 will last for 2 weeks and is expected to take place in the period February/March 26. Patients and curators will be requested to assess Generation 2 (G2) of the prototype, by performing the same task as in Epoch 2 to enable comparison of the performance of the 2 generations of the prototype.

4.2. Patient input in Study Design

The development of the AIDAVA prototype is taking place with the support of "patient consultants" selected by the ECPC (European Patient Cancer Coalition) and EHN (European Heart Network) patients associations. The selected patient consultants are people with a genuine interest in being more in control of their own medical record (or the health record of their loved ones) to improve the quality of their own health and support clinical research; during the AIDAVA project, they do not contribute their personal data but act as representatives of patients.
Four patient consultants were selected from each organisation; each patient consultant is supporting the project for about 42 person days - including participation in co-creation workshops - and is paid for this effort as consultant based on a fee (350 EUR/day) recommended by the European Patient Forum at the time of writing the proposal.
One of the tasks of these patient consultants was to review the enclosed protocol; they also will be asked to review interim versions of the prototype in development and to pre-test the AIDAVA prototype.

[13] This is a prospective study as we enrol patients, check existing data and collect new data (QALY and medical device data) and see how the endpoints develop. A portion of the study is however retrospective as we use existing data available within the medical dossier of a patient.

4.3. Enrollment

Patient screening and enrollment

- The recruitment period will last from Jan to June 2024.
- Patients will be identified by the local research associate - with support from responsible clinicians - in each hospital, based on the eligibility criteria specified in Section 5.2.
- A trained research associate will then approach the patient (see Section 4.4 for details)
    - to explain the study objectives and approach based on the study information package (SIP, see ref) and a structured information session; this will include the expected benefits for the patients to contribute to the study and the expected workload during the assessment.
    - to discuss the Informed Consent Form (ICF - see Annex 2) and check if the patient understands the form and is ready to sign it
    - to explain to the patient the need to sign a - temporary - collaboration agreement with the identified Health Data Intermediary (HDI) to pool non hospital data
    - to check if the patient is ready to evaluate the tool himself/herself.
- Following the information session, the patient who agrees to participate will be requested to
    - start to use personal apps (BrightFish[14]- see Section6), with a quality of life questionnaire,and explain why it is important and how to use the app
    - (for CVD patients) start using the provided medical device (Withings Blood Pressure Monitor, a CE marked device - see Section 6); training on how the device should be used will be provided
    - sign the ICF
    - sign a collaboration agreement with the HDI collaborating with the hospital, for the duration of the study
- Once a patient has signed in for the evaluation study, the research associate will keep supporting him/her during the evaluation phase.

Expert curators enrollment

- Expert data curators are existing staff within the hospital that perform curation regularly; some of them contributed to the requirement of the AIDAVA prototype solution.
- Curators will be following the same information sessions and training given to the patients.

[14] https://brightfish.com/solutions/ehealth/

4.4. Study execution : Study visits overview

4.4.1. Prerequisite (± 1 month before evaluation, respectively for G1 and G2)

Testing of data transfer (see more details in Section 11.4 on Data Protection)

1. Hospital data: Patient data (testing data) from hospital system to AIDAVA prototype (deployed in hospital) based on predefined data transfer specifications and signed data sharing agreement (see template in Annex 4 - Data Transfer Specifications)
2. Personal Data to HDI: Data from Device/app API, and potentially GP data, to Health Data Intermediary based on agreed data sharing agreement, with patient consent
3. Personal Data to AIDAVA: Personal data gathered through HDI to AIDAVA prototype based on predefined data transfer specifications and signed data sharing agreement
4. Patient IPS to HDI: patient IPS (based on synthetic data) from AIDAVA prototype (deployed in hospital) to Health Data Intermediary, based on HL7 data format (with agreed version) and signed data sharing agreement

Pretraining of expert curators

1. Request to answer the health and digital literacy questionnaire to include in the user profile (in REDCap , User Profile Form)
2. Training on what the project is about, what AIDAVA is about, how to curate data and how to support the patient

Purchase of the medical devices in each site

4.4.2. First time patient (for G1 and G2)

This is applicable for new patients, for G1 (assessment period: July /Sept 2024) and G2 (assessment period: Feb/March 2026) of the prototype.
See data collection and entry forms in Section 10 to collect the information on the patient visit

	CVD patients	BC patients
Screening and enrollment (Checking I/E criteria and participation to information session) **G1 - Jan-June 2024**	Not applicable: CVD patients will be screened and enrolled after an infarction event (acute myocardial infarction - MI) W1, D1 during the execution of the study starting August 2024	Approximately up to 50 patients screened, 15 patients enrolled During regular hospital visit 1. Checking I/E criteria (In REDcap, Patient Screening FORM) 2. Project introduction (10 min, doctor). Research associate complete REDCap Form (in REDCap, Patient Screening FORM) *If the patient meets the criteria and is interested in participating in the project, he/she will move on to the research associate*

	CVD patients	BC patients
<u>VISIT 0 - on site</u> **G2 - Sep 2025 - Feb 2026**		3. Under supervision of Research Associate - 30-60 min per patient ○ Share details of the project with patients ; ○ Provision of SIP and explanation - including use of EQ-5D QALY quality of life questionnaire several time ○ Introduction of HDI and questions regarding ownership and future use of the data through HDI ○ Answer to questions ○ Request to read and sign ICF and HDI agreement 4. Research associate update REDCap Form (in REDCap, Patient Screening FORM) 5. Confirmation of patient enrollment;
Site opening visit **G1: June 2024** **G2: Feb 2026**	Training of trainers (see training plan in Annex 3) : research associate and expert curators will be trained for: a. Overall understanding on the AIDAVA project and tools b. Advanced training on the AIDAVA prototype (ingestion, curation, publishing) - for appropriate generation c. How to support the patients d. Device and app installation for the patient: present device and show how it must be used	
G1 and G2 **<u>VISIT 1 - on site</u>** **W1** **Day 1** (either elective outpatient clinic visit OR inpatient stay during index MI event),	Approximately up to 50 inpatients screened, 15 inpatients enrolled (Note: Day1 and Day2 may seem busy for a patient recovering from MI; most patients are however recovering well and find difficult to remain in hospital during the monitoring period) 1. Checking I/E criteria (In REDcap) 2. Project introduction (10 min, doctor). Research associate complete REDCap Form (in REDCap, Patient Screening FORM) *If the patient meets the criteria and is interested in participating in the project, he/she will move on to the research associate* 3. Under supervision of Research Associate - 30-60 min per patient	

<table>
<tr><th></th><th>CVD patients</th><th>BC patients</th></tr>
<tr><td rowspan="2"></td><td>a. Share details of the project with patients ;
b. Provision of SIP and explanation - including use of EQ-5D QALY - quality of life questionnaire several time
c. Introduction of HDI and questions regarding ownership and future use of the data through HDI
d. Answer to questions
e. Request to read and sign ICF and HDI agreement
f. Provide medical device to patient
4. Research associate update REDCap Form (in REDCap, Patient Screening FORM)
5. Confirmation of patient enrollment;</td><td></td></tr>
<tr><td colspan="2">1. Create HDI account (through AIDAVA)
2. Questionnaire to specify user profile (In REDCap, User Profile FORM)
3. Training session for the patient on the prototype by the research associate
4. Patient fills EQ-5D form through HDI provided app
5. Research associate complete REDCap Form (in REDCap, Patient monitoring FORM) data extraction from different sources for enrolled patients
6. Patient and expert curator curate medical data with AIDAVA</td></tr>
<tr><td rowspan="2">W1
Day 2 - 5</td><td>CVD patient will use medical device</td><td></td></tr>
<tr><td colspan="2">1. Patient will use AIDAVA and test its functionalities, take notes and if necessary, will contact Helpdesk whenever needed
2. Expert curator will curate the data in AIDAVA on demand.
3. RA fill the patient monitoring form (in REDCap, Patient monitoring FORM)</td></tr>
<tr><td rowspan="2">W1
Day 5</td><td>1. First calculation of SMART risk score by AIDAVA (automatically) and CVD expert (manually).
2. Data collection and entry forms : Documentation of SMART risk score calculation result and time (in REDcap, Patient Monitoring FORM) by RA</td><td>1. System extracts the patient data for insertion into the local AIDAVA BC registry
2. Queries identified as secondary end points will be fired automatically by the system and results available for display to the BC data users</td></tr>
<tr><td colspan="2">1. Data curators spreadsheet (Patient as curator and Expert Curator), (in REDCap, Patient Curator FORM; Expert Data Curator Form)</td></tr>
</table>

	CVD patients	BC patients
	2. Data user/consumers spreadsheet (In REDCap Data user/consumer FORM) 3. AIDAVA extract IPS and forward it automatically to the HDI + enter flag in Data collection and entry forms (REDCap, Patient monitoring FORM)	
VISIT 2 - online W2 Day 1	CVD patients use medical device Check if there is any issue with the medical device (CVD patient only)	
	1. Second training session on the prototype - online (open questions) 2. Patient and expert curator curate medical data using AIDAVA 3. RA fill the patient monitoring form (in REDCap, Patient monitoring FORM)	
W2 Day 2-5	CVD patient will use medical device	
	1. Patient will use AIDAVA and test its functionalities, take notes and if necessary, will contact helpdesk 2. Expert curator will curate the data in AIDAVA on demand. 3. RA fill the patient monitoring form (in REDCap, Patient monitoring FORM)	
VISIT 3 W2 Day 5	1. Second calculation of SMART risk score by AIDAVA and CVD expert 2. Data collection and entry forms: documentation of SMART risk score calculation result and time	
	1. AIDAVA extract IPS and forward it automatically to the HDI (+ enter flag in REDCap, Patient monitoring FORM) 2. Patient and curator will fill in the system usability scale questionnaire online (in REDCap, System Usability Scale - SUS FORM) 3. Data curators spreadsheet (Patient and Expert Curator). (in REDCap, Patient Curator FORM; Expert Data Curator Form) 4. Data user/consumers spreadsheet (In REDCap Data user/consumer FORM)	
W3		At the end of G1 - across all sites - fires the queries identified as secondary end points; RA reports the results in the data users/consumers spreadsheet (in REDCap Data user/consumer FORM).
End of evaluation	Delete personal patient data and PHKG (after finalisation of evaluation report), but keep evaluation score from G1.	
VISIT 4 - online W4	Feedback to patient	

4.4.3. Development of G2 (improvement of G1)

<table>
<tr><th></th><th>CVD patients</th><th>BC patients</th></tr>
<tr><td rowspan="2">August 2024 /Sept 2025</td><td colspan="2">Development of G2
Patient will receive information letters every 3 months</td></tr>
<tr><td>CVD Patients will continue using the Medical device (see Section 6) and will continue sending these data to Data Intermediary.
In case of problems patient can contact the research associate</td><td></td></tr>
</table>

4.4.4. Returning patients (for G2 only)

This is applicable only for patients who have participated to the assessment of G1 and are also assessing G2

<table>
<tr><th></th><th>CVD patients</th><th>BC patients</th></tr>
<tr><td>2 weeks for start
G2: Jan 2026</td><td colspan="2">Refresh training of Expert curators and Research associates</td></tr>
<tr><td>G2
<u>VISIT 1 - on site</u>
W1
Day 1</td><td colspan="2">1. Upload of all data (starting from scratch)
2. Patient will be called in for outpatient visit by the research associate
3. Questionnaire to specify user profile (In Redcap, User Profile FORM)
4. Training session for the patient on the prototype - G2 version - by the research associate (2,5h)
5. Patient fills EQ-5D QALY - quality of life questionnaire through HDI provided app
6. Research associate complete REDCap Form (in REDCap, Patient monitoring FORM) for enrolled patients
7. Patient and expert curator curate medical data with AIDAVA</td></tr>
<tr><td rowspan="2">W1
Day 2-5</td><td>CVD patient use the medical device</td><td></td></tr>
<tr><td colspan="2">1. Patient will use AIDAVA and test its functionalities, take notes and if necessary, will contact helpdesk
2. Expert Curator will curate the data in AIDAVA, on demand
3. RA fill the patient monitoring form (in REDCap, Patient monitoring FORM)</td></tr>
</table>

	CVD patients	BC patients
W1 Day 5	1. First calculation of SMART risk score by AIDAVA (automatically) and CVD expert (manually). 2. Data collection and entry forms documentation of SMART risk score calculation result and time (in REDCap, Patient monitoring FORM)	1. System extracts the patient data for insertion into the local AIDAVA BC registry 2. Queries identified as secondary end points will be fired automatically by the system and results available for display to the BC data users
	1. AIDAVA extract IPS and forward it automatically to the HDI (+ enter flag in REDCap, Patient monitoring Form) 2. Data curators spreadsheet (Patient as curator and Expert Curator). (in REDCap, Patient Curator FORM; Expert Data Curator Form) 3. Data user/consumers spreadsheet (In REDCap, Data user/consumer Form)	
<u>VISIT 2 - online</u> W2 Day 1	CVD patients use medical device Check if there is any issue with the medical device (CVD patient only)	
	1. Second training session on the prototype - online (open questions) 2. Patient and expert curator curate medical data using AIDAVA 3. RA fill the patient monitoring form (in REDCap, Patient monitoring Form)	
W2 day 2-5	CVD patient use the medical device	
	1. Patient will use AIDAVA and test its functionalities, take notes and if necessary, will contact helpdesk 2. Expert curator will curate the data in AIDAVA on demand. 3. RA fill the patient monitoring form (in REDCap, Patient monitoring Form)	
<u>VISIT 3</u> W2 Day 5 March 2026	1. Second calculation of SMART risk score by AIDAVA and CVD expert 2. Data collection and entry forms: documentation of SMART risk score calculation result and time	1. Queries identified as secondary end points will be fired automatically by the system
	1. AIDAVA extract IPS and forward it automatically to the HDI (+ enter flag in REDCap, Patient monitoring Form) 2. Data curators spreadsheet (Patient as curator and Expert Curator). (In REDCap, Patient Curator FORM; Expert Data Curator Form) 3. Data user/consumers spreadsheet (In REDCap, Data user/consumer form) 4. Patient and curator will fill in an System Usability Scale online (in REDCap, System Usability Scale - SUS Form) 5. RA fill the patient monitoring form (in REDCap, Patient monitoring Form)	

	CVD patients	BC patients
	6. End of study	
End of evaluation	Delete personal patient data and PHKG (after finalisation of evaluation report)	
VISIT 4 - online W4	Feedback to patient	

4.2. End-of-Study

The end of the study is defined as the date of Week 2/ Day 5 in Generation 2, for the last participant in the study. A participant is considered to have completed the study if the participant has completed all periods of the study including the last scheduled procedure shown in the SoA in Section 1.3.

At the end of the study, the final report with consolidated metrics and statistical analysis will be issued.

The personal data that have been curated and generated in the form of Personal Health Knowledge Graph (PHKG) will be maintained within the AIDAVA instance in the hospital up to the end of the project.
At the end of the project, the patients will be asked if they want to keep their PHKG, and if yes where; the data will be transferred to the requested data store. If the patients do not wish to keep the PHKG, a confirmation will be requested to the patient ; in case of confirmation the PHKG will be deleted.

5. Study Population

5.1. Overview

For testing the 2 generations of the AIDAVA prototype, we focused on **2 Therapeutic Areas (TAs) in 3 countries/languages**, within **2 types of healthcare organisations.**

We wanted to evaluate 2 different **TAs** to cover a broader scope of type of data collected in these 2 different medical conditions, while limiting the workload for training AI tools. Considering that we want to solve a technical problem that is critical to solve medical issues, we used the following criteria to select the Therapeutic Areas (TAs): 1) clinical areas in need of further clinical research and of high quality, FAIR data; 2) population prevalence that includes a wide potential base of AIDAVA users; 3) patient cohorts and/or registries are available via clinical and HDI partners in the AIDAVA consortium, providing real-life examples of the problems we want to solve. This led us to the selection of the following TAs.

- **Breast cancer.** Half of breast cancer patients are younger than 64 years. They present an interesting population in the context of AIDAVA, in that they and their loved ones may be highly motivated to do curation tasks and will have sufficient computer literacy to interact with data. While there is no EU-wide breast cancer registry[15]; screening registry exists at national level and would benefit from more detailed information[16] and consolidation at EU level. The 3 academic hospitals in the AIDAVA consortium have operational breast cancer registries and/or a large number of patients.
- **Cardiovascular disease, specifically for patients at risk of myocardial infarction.** Cardiac arrest is 3 times more frequent in 50+ years elderly males than in females; they constitute a nice counterbalance to the females' population in breast cancer. In addition, this is a TA where availability of a high quality personal longitudinal record is extremely valuable for clinical practitioners to predict the cardiovascular risks. All 3 participating centres have a wide cohort of CVD patients.

We focused on 3 **languages** (Estonia/Estonian, Austria/German and The Netherlands/Dutch) as we needed to limit the workload related to annotating text-based documents, required as training sets for the AI algorithms while providing a foundation for demonstrating the approach and showcasing its potential expansion across all EU languages.

In terms of **organisation**, we work with academic hospitals and personal Health Data Intermediaries (HDIs) supporting personal health data sharing on behalf of each consenting patient, and regulated under the emerging European Data Governance Act (DGA). MIDATA and DIGI.me are partners of the AIDAVA project and contributed to the development of the prototype

Country	Academic hospital	Health Data Intermediary
Estonia	North Estonia Medical Center (NEMC)	MIDATA (MID)
Austria	Medical University Graz (Graz)	MIDATA (MID)
The Netherlands	(CVD) Maastricht University Medical Center (MUMC) (BC) Maastro Clinic (Maastro)	DIGI.me (DME) - certified MedMij

[15] https://www.karger.com/Article/Fulltext/503715

[16] https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-019-6846-6

5.2 Study Population

The study participants include

- **Patients** who met the inclusion/exclusion criteria specified below
- **Expert data curators,** i.e. staff from the hospital, responsible for the maintenance of the local AIDAVA registry, extracting data from the patient medical record and inserting them into the registry
- **Breast Cancer specialists**: staff from the hospital responsible for treating patients and active in clinical research with the local breast cancer registry.
- **CVD practitioners:** staff from the hospital responsible for treating CVD patients

The staff from the hospital that will collaborate on the study is specified in Section 10.4.

The patients will be selected based on the following inclusion/exclusion criteria.

Inclusion criteria (MUST)	• Data available in EHR within related medical centre • Owner and user of a smartphone • Consent to give access to their personal data (based on scope provided in the Informed Consent Form) to the data curator identified in the hospital • Agree to sign a collaboration agreement with the relevant HDI - for the duration of the study • Agree to test the AIDAVA prototype (G1 and G2) and to delegate to a representative data curator in case of issues
Inclusion criteria for BC patients (MUST)	• Gender: All • History of a confirmed diagnosis of breast cancer treated with surgery and radiotherapy
Inclusion criteria for CVD patients	• Gender: All • Confirmed diagnosis of symptomatic type 1 acute Myocardial Infarction
Inclusion criteria (NICE TO HAVE)	• Health and digital literacy
Exclusion criteria	• Data not accessible (for instance for technical reason) • Vulnerable persons • Children/minors • Inability/unwillingness to use smartphone applications • Does not use a smartphone • Russian only as a language (in Estonia)

5.3. Lifestyle Considerations

No specific requirements

5.4 Screen Failures

A screen failure occurs when a participant who has consented to participate in the study is not subsequently participating in the study.

A minimal set of screen failure information is required to ensure transparent reporting of screen failure participants. This includes

- demography: age, gender, education level, user profile
- eligibility criteria
- screen failure details: known explanation for failure
- and any concern related to study

As mentioned above, approximately up to 50 patients will be screened - per site/ therapeutic area - to ensure we can enrol 15 patients in the study.
Individuals who do not meet the criteria for participation in the study (screen failure) will not be rescreened.

6. Study Intervention(s)

The objective of collecting these data is to demonstrate that AIDAVA is capable to integrate - and curate - data generated by the patients, as well as data generated by the Health Care Providers (hospital staff and GP) and therefore ensure a complete view of the patient records. The data collected through the personal apps and medical devices will not be used for clinical decision making.

6.1 Personal Apps

Participants will be requested to use a specific personal app to capture personal information that could be of value in a longitudinal health record.

	Breast Cancer & CVD
Name	BrightFish
Data Captured	QALY questionnaire (local language): EQ-5D standard
Period of use/ Frequency	Once during Week 1 of the evaluation (for G1 and for G2)
Provider	BrightFish
Link	https://brightfish.com/solutions/ehealth/
Pricing	(in negotiation - maximum of 16 K for whole projects)
Instructions for use	Provided via HDI in training materials
Monitoring of compliance of use	Not needed as the value of the information collected by the personal app has not be identified as a key data element for the registry

6.2. Medical Devices (CVD patient only)

Device/app	Notes
BP monitor with Afib detection	Withings BPM Connect
Data Captured	Blood Pressure
Period of use/ Frequency	During the whole evaluation : 3 times a week / 3 times a day Afterwards - once a month/ 3 times a day
API Integration	API Guide
Google Fit app	For Android users
Apple Health app	For iOS users
Instruction for use	In case of issue with the value of BP : go to the GP In case of issue with the app or the device: go the RA
Instructions in case of defect	Request support from Research Associate
Disposal after end of study	Patient can keep it Note: patients withdrawing from the study should return the device to the Research Associate/ site
Monitoring of compliance of use	Research Associate

All device deficiencies (including malfunction, use error and inadequate labelling) shall be documented and reported by the investigator throughout the clinical investigation.

7. Participant Discontinuation/Withdrawal

7.1 Participant Discontinuation/Withdrawal from the Study

- A participant may withdraw from the study at any time at the participant's own request for any reason (or without providing any reason).
- A participant may be withdrawn at any time at the discretion of the investigator for behavioural, or compliance reasons.
- At the time of discontinuing from the study, the personal health data from the patient will be deleted in the same way it is done at the end of each generation. The endpoints entered by the patient as part of the evaluation, will be used, if there is consent form the patient (per Informed Consent Form).
- The participant will be permanently discontinued from the study intervention and the study at that time.
- If the participant withdraws consent for disclosure of future information, the sponsor may retain and continue to use any data collected before such a withdrawal of consent.
- If a participant withdraws from the study, the participant may request destruction of any samples taken and not tested, and the investigator must document this in the data collection and entry forms (REDCap).

7.2 Lost to Follow up

A participant will be considered lost to follow-up if the participant repeatedly fails to return for scheduled visits and is unable to be contacted by the study site.
The following actions must be taken if a participant fails to return to the clinic for a required study visit:

- The site must attempt to contact the participant and reschedule the missed visit as soon as possible, counsel the participant on the importance of maintaining the assigned visit schedule, and ascertain whether the participant wishes to and/or should continue in the study.
- Before a participant is deemed lost to follow-up, the investigator or designee must make every effort to regain contact with the participant (where possible, telephone calls, and if necessary, a certified letter to the participant's last known mailing address or local equivalent methods). These contact attempts should be documented in the participant's medical record.
- Should the participant continue to be unreachable, the participant will be considered to have withdrawn from the study.

In the AIDAVA assessment study, there is a risk of patient discontinuation because of the 16 to 18 months time lag between EPOCH 1 and EPOCH 3. To decrease the risk and keep the momentum / interest of patients we will issue regular newsletters (one page every 3 to 4 months). In case of an important discontinuation we might need to ask sites to recruit new patients for EPOCH 3.

In case of withdrawal, patients will be requested what they wish to do with the PHKG containing their data and available in the system; they will be able to receive these data or to request deletion of the PHKG in the same way that at the end of the study (see Section 4.4). In case of lost to follow-up, the PHKG of the related patient will be deleted at the end of the study.

8. Study Assessments and Procedures

- Study procedures and their timing are summarised in the SoA. Protocol waivers or exemptions are not allowed.
- Adherence to the study design requirements, including those specified in the SoA, is essential and required for study conduct.
- All screening evaluations must be completed and reviewed to confirm that potential participants meet all eligibility criteria. The investigator will maintain a screening log to record details of all participants screened and to confirm eligibility or record reasons for screening failure, as applicable.
- Procedures conducted as part of the participant's routine clinical management and obtained before signing of the ICF may be utilised for screening or baseline purposes provided the procedures met the protocol-specified criteria and were performed within the timeframe defined in the SoA.
- In the event of a significant study-continuity issue (eg, caused by a pandemic), alternate strategies for participant visits, assessments, medical device distribution and monitoring may be implemented by the investigator, as per local health authority/ethics requirements.

The study is using already collected data, except for personal apps and data collected through the medical devices.

Adverse Events (AEs) or Serious Adverse Events (SAEs) related to the AIDAVA prototype are not expected during this assessment study, as this is a Class 1 medical device supporting patients in managing and curating available data and there are no medical interventions during the assessment apart from collecting additional data.

For CVD patients, in case of medical device deficiency for the CVD use case, the potential risk is that a SMART risk score will not be properly computed. Physicians will calculate the score as well manually and can compare the calculations with the score that AIDAVA has provided; in any case no medical decisions will not be made based on AIDAVA score alone.
In case of a high SMART risk score detected during the assessment, the treating physician should use measures that include promoting a healthy lifestyle through behaviour changes, including nutrition, physical activity, relaxation training, weight management and smoking cessation programmes for resistant smokers. These would ideally help cope with the illness and improve adherence and cardiovascular outcome.

9. Statistical Considerations

The analysis and reporting will be done on all data from all participants at 3 points in time: before the evaluation starts (curator arm only for G0), after G1 evaluation (G1 report) and after G2 evaluation (G2 report).

9.1 Statistical Hypotheses

The primary objective is to demonstrate that G1 of the AIDAVA prototype is superior to the current situation in achieving data curation and publishing of personal health data; and that G2 is superior to G1. Improved efficiency will be measured through 4 endpoints i.e. M1 (usability of the tool), M2 (decreased time to curated data), M3 (high quality of resulting data), M4 (extraction of structured data from text).

- **The 2 critical metrics are M2 (time) and M3 (quality).**
- While important, **M1 (usability)** should be possible to improve (more training, increased explanation, ..) and the results of the evaluation will provide directions on how to improve.
- M4 (NLP performance) will improve with the introduction of new technologies (Large Language Models like ChatGPT and BARD) and is therefore less critical for the future of AIDAVA.

Thus, the null hypothesis to be tested in relation to the **primary estimand**[17] is as follows.

- Null hypothesis (H0): The efficiency of data curation and publishing of personal health data is not different between G1 or G2 of the AIDAVA prototype and the current situation.
  There is no impact due to training on data curation.

vs.

- Alternative hypothesis (H1):
  - G1 of the AIDAVA prototype is more effective than the current situation with respect to the efficiency of data curation and publishing of personal health data and G2 is even more effective than G1.
  - The efficiency improvement is due mainly to the tool; some improvement is related to the training on data curation.

The null and alternative hypotheses corresponding to the secondary estimands are as follows:

Secondary objective for Breast Cancer Use Case:

- Null hypothesis: There is no difference with AIDAVA, to get an answer to specific queries on the BC registry across sites, in terms of accuracy and time.
- Alternative hypothesis (H2-BC): AIDAVA G1 provides correct answers on queries on the BC registries across sites (with an accuracy 90% or higher), while avoiding BC specialists (and supporting team) to spend time. G2 has similar or better results.

[17] estimand (trial objective) is the target of estimation. In AIDAVA, the primary estimand is improved efficiency in data curation

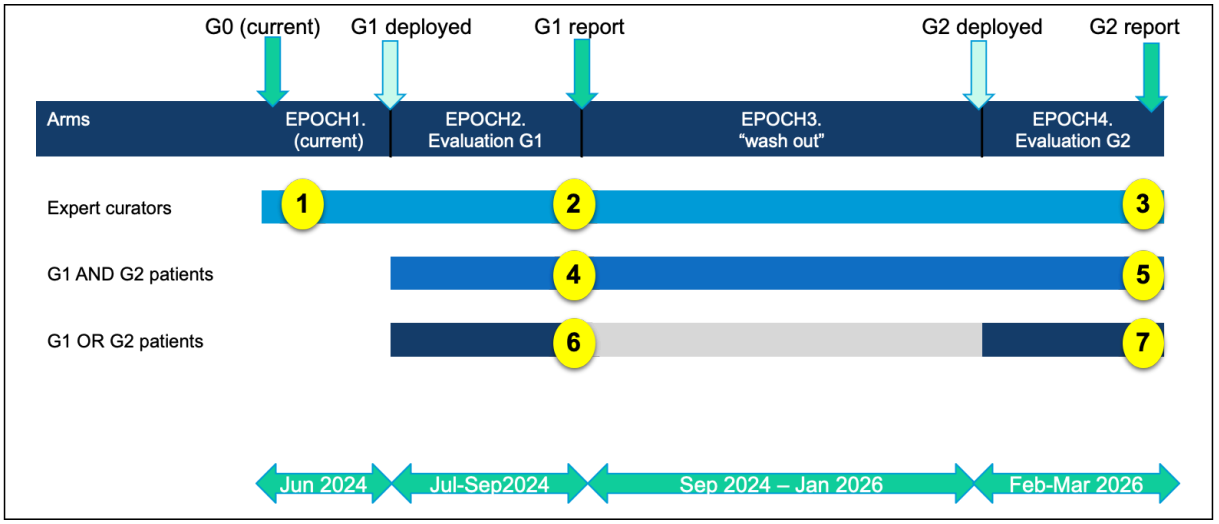
Secondary objective for CVD specialists :

- Null hypothesis: There is no difference to compute SMART score for patients across sites, in terms of accuracy and time with AIDAVA.
- Alternative hypothesis (H2 - CVD): AIDAVA G1 provides an accurate SMART score (accuracy of 80%+) on patients, across sites, without any time spent by CVD specialists. G2 has similar or better results.

9.2 Analysis Sets

Primary endpoint components

The analysis set will include all expert curators and patient participants and will take into account the primary data points (PDP) identified in yellow in the figure below across the 3 arms.

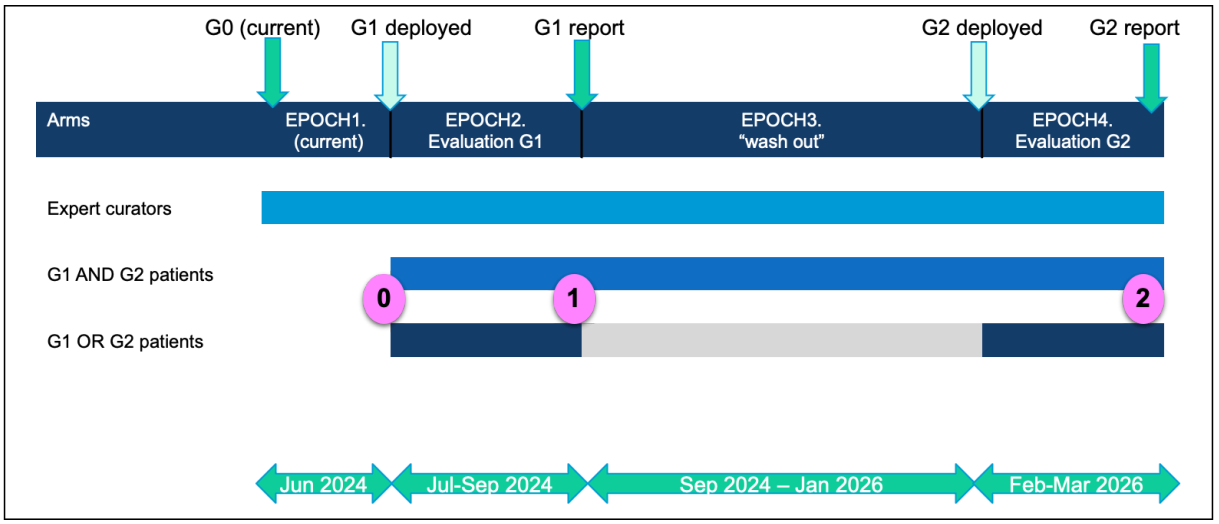


Primary Data Point	Generation	Arm	Description of measures (through RedCap Form and computed)
PDP1	G0	Expert Curator	**M2.1** : Estimated time spent in curation per patient record (min)
PDP2	G1		**M1**: System Usability Score **M2.1**: Estimated time spent in curation per patient record (min)
PDP3	G2		Same than for G1 - measured for G2
PDP4	G1	G1 **AND** G2 users	**M1**: System Usability Score **M2**: ● Time spent in for curation in min ○ Estimated time in RedCap form (**M2.1**) ○ Computed "active" time to answer queries (**M2.2**) ● Computed amount of data elements ○ curated automatically by system (**M2.3**) ○ requiring human intervention ■ with answer from the patient (**M2.4**)

Primary Data Point	Generation	Arm	Description of measures (through RedCap Form and computed)
			■ without answer the patient and requiring curator input (**M2.5**) ■ without answer requested from curator but not successful (**M2.6**) **M3**: Computed data quality scores of curated record in terms of completeness, correctness, consistency and trustworthiness (as detailed in Task 4.6). **M4**: Computed quality of the NLP concept extraction against the manually annotated data ("gold standard") ● Precision (**M4.1**) which measures the fraction of gold standard instances among the retrieved instances ● Recall (**M4.2**) which measures the fraction of the gold standard concepts that could successfully be retrieved
PDP5	G2		Same than PDP4 - for G2
PDP6	G1	G1 **OR** G2 users	Same than PDP4 - for G1
PDP7	G2		Same than PDP5 - for G2

Secondary endpoints

The analysis set will include all data users (BC specialists and CVD experts) and will take into account the secondary data points (SDP) identified in orange in the figure below across the patient arms. As indeed this is about reuse of data, after the curation process; there is no effect expected from the fact that patients did use G1 before G2 or not.



Secondary Data Point	Generation	Arm	Description of measures (through RedCap Form and computed)
Breast Cancer			
SDP0	G0	With ± 15 patients on site	• **EBCx**: estimated total effort (person min, person hours or person days) for BC specialist and supporting local team to answer the queries (BC1, BC2, BC3 described in Section 3.2.1) across sites • **TBCx**: estimated duration (min, hours or days) for the BC specialist to have an answer to the queries (BC1, BC2, BC3) across sites. Duration = time between the first request and the time the answer is provider
SDP1	G1	All patients	• **EBCx**: total effort for BC specialist (and local team) to answer the queries (BC1, BC2, BC3) across sites • **TBCx**: duration (min, hours or days) for the BC specialist to have an answer to the queries (BC1, BC2, BC3) across sites. • **HBCx**: answer to the query as provided by human BC specialist • **ABCx**: answer to the query as computed by AIDAVA • **ACCx**: accuracy of calculation by AIDAVA compared to BC specialist
SDP2	G2		Same than SDP1 - for G2
CVD			
SDP1	G1	All patients	• **CVD**. SMART risk score computed by AIDAVA • **TCVD1.1/TCVD1.2.** Time required for the treating/ senior physician to compute the SMART score (collect + analyse + enter data into a standalone web page calculator) • **CVD1.1/CVD1.2**. SMART score computed by treating / senior physician • **ACVD1.1/ACVD1.2**. Accuracy of automatic calculation of the score compared to treating/senior physician-entered score
SDP2	G2		Same than SDP1 - for G2

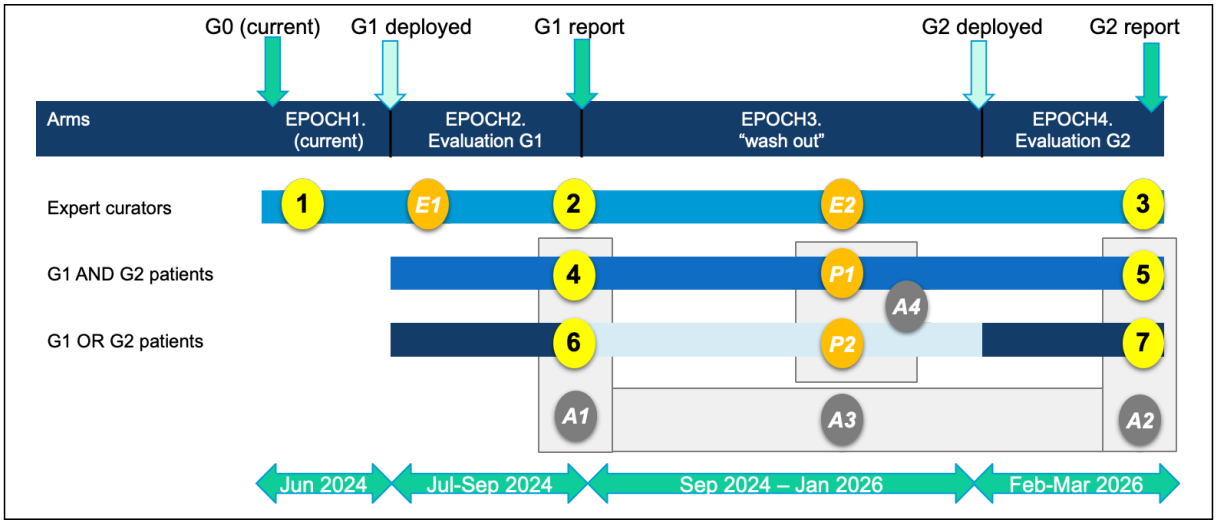
9.3 Statistical Analyses

Three tests have be retained

- **Paired t-test**: statistical test that compares the means of two related groups, where each subject is measured at two different time points or under two different conditions. It is used to determine if there is a significant difference between the two groups.
- **Mann-Whitney U test (unpaired t-test)**: non-parametric statistical test used to compare the medians of two independent groups. It is often used when the data are not normally distributed or when the sample sizes are small.
- **Wilcoxon signed-rank test:** non-parametric statistical hypothesis test used to compare two populations with two matched samples.

Primary Endpoint(s) Analysis

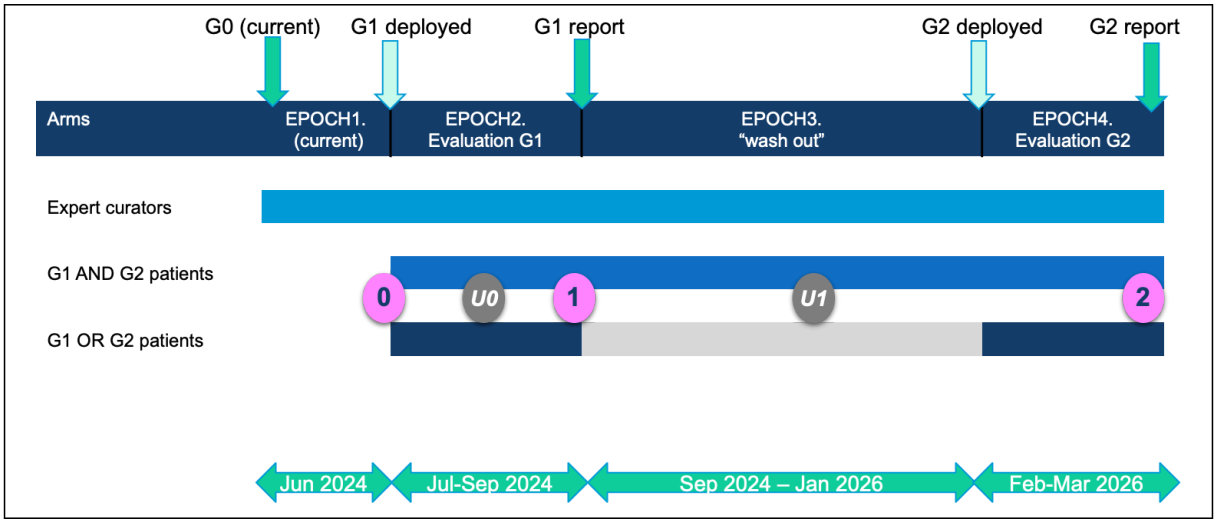
The endpoints are defined in Section 9.2 will be used to support the analysis



Computed	Description of measures	Test & meaning (per H1)
Population: Expert Curators		
E1	*Difference between data points 1 and 2*: **E1.M2.1Δ =** M2.1@PDP1 - M2.1@PDP2 **E1.M2.2Δ =** M2.2@PDP1 - M2.2@PDP2	Paired t-test H1: less time spent in curation in G1 than in G0 for expert curators
E2	*Difference between data points 2 and 3:* **E2.M1-Δ** = M1@PDP2 - M1@PDP3	Paired t-test H1: G2 is more usable than G1 for expert curators
	E2.M2.1Δ = M2.1@PDP2 - M2.1@PDP3 **E2.M2.2Δ =** M2.2@PDP2 - M2.2@PDP3	H1: less time spent in curation in G2 than in G1 for expert curators
Population: Patients		
P1	*Difference in data points 4 and 5 (i.e. difference between G1 and G2 for arm where patients used G1 AND G2)*	Paired t-test ***(effect of system + training/experience from G1)***
	P1.M1-Δ = M1@PDP4 - M1@PDP5	H1: G2 more usable than G1
	P1.M2.xΔ = M2.x@PDP4 - M2.x@PDP5	H1: G2 requires less time for curation than G1 H1: G2 enables curation with less human intervention than G1
	P1.M3-Δ = M3@PDP4 - M3@PDP5	H1: G2 provide same or higher quality than G1
P2	*Difference in data points 6 and 7 (i.e. difference between G1 and G2 for arm where patients used G1 OR G2)*	Unpaired t-test ***(effect of system only)***
	P2.M1-Δ = M1@PDP6 - M1@PDP7	H1: G2 more usable than G1
	P2.M2.xΔ = M2.x@PDP6 - M2.x@PDP7	H1: G2 requires less time for curation than G1 H1: G2 enable curation with less human intervention than G1
	P2.M3-Δ = M3@PDP6 - M3@PDP7	H1: G2 provide same or higher quality than G1
A1	*Sum across patient arms for G1* A1.M1 = M1@PDP4 + M1@PDP6 A1.M2.x = M2.x@PDP4 + M2.x@PDP6 A1.M3.x = M3.x@PDP4 + M3.x@PDP6 A1.M4.x = M4.x@PDP4 + M4.x@PDP6	Descriptive Statistics Performance in G1 of the AIDAVA system.
A2	*Sum across patient arms for G2*	Descriptive Statistics

Computed	Description of measures	Test & meaning (per H1)
	A2.M1 = M1@PDP5 + M1@PDP7 A2.M2.x = M2.x@PDP5 + M2.x@PDP7 A2.M3.x = M3.x@PDP5 + M3.x@PDP7 A2.M4.x = M4.1@PDP5 + M4.x@PDP7	Performance in G2 of the AIDAVA system.
A3	*Difference in A1 and A2 (i.e. difference between G1 and G2 across all patients)*	Unpaired t-test
	A3.M1-Δ = A1.M1 - A2.M1	H1: G2 more usable than G1
	A3.M2.xΔ = A1.M2.x - A2.M2.x **A3.M3.x-Δ** = A1.M3.x - A2.M3.x	H1: G2 requires less time for curation than G1 H1: G2 enable curation with less human intervention than G1
	A3.M4.X-Δ = A1.M4.x - A2.M4.x	Wilcoxon signed-rank test H1: The NLP system developed in G2 yields a higher precision than the NLP system from G1. H1: The NLP system developed in G2 yields a higher recall than the NLP system from G1.
A4	*Difference in P1 and P2 (difference between the 2 patient arms)*	Unpaired t-test (effect on training/experience) There is a difference in performance due to training/experience with the system : patients having used G1 are more/not more effective in G2 than patients who work with only one generation of the prototype (G1 or G2)
	A4.M1-Δ = P1.M1-Δ - P2.M1-Δ	H1: G2 more usable than G1 (due to experience during G1)
	A4.M2.x-Δ = P1.M2.x-Δ - P2.M2.x-Δ **A4.M3.x-Δ** = P1.M3.x-Δ - P2.M3.x-Δ	H1: G2 requires less time for curation than G1 as patients are more experienced H1: There are less questions of the system unanswered during G2 than in G1 as the patient acquired experience in G1
	A4.M4	Not useful

Secondary Endpoint(s) Analysis



Computed	Description of measures	Test & meaning (per H1)
Breast Cancer		
U0	*Difference in data points 0 and 1 (i.e. difference between G0 and G1 across a sample of 15 patients)* **U0.TBCx-Δ** = TBCx@SDP0 - TBCx@SDPA **U0.EBCx-Δ** = EBCx@SDP0 - EBCx@SDP1 **U0.ACCx-Δ** = ACCx@SDP0 - ACCx@SDP1	Unpaired t-test H1: G2 has more (or same) accuracy than G1 while keeping the same time saving and duration
		Descriptive Statistics around performance in G1 of the AIDAVA system (e.g. mean,..)
U1	*Difference in data points 1 and 2 (i.e. difference between G1 and G2 across all patients*) **U1.TBCx-Δ** = TBCx@SDP1 - TBCx@SDP2 **U1.EBCx-Δ** = EBCx@PSP1 - EBCx@SDP2 **U1.ACCx-Δ** = ACCx@SDP1 - ACCx@SDP2	Paired t-test H1: G2 has more (or same) accuracy than G1 while keeping the same time saving and duration
Cardiovascular		
		Descriptive Statistics around performance in G1 of the AIDAVA system (e.g. mean,..)
U1	Difference in data points 1 and 2 (i.e. difference between G1 and G2 across all patients) **U1.TCVD-Δ** = TCVD@SDP1 - TCVD@SDP2 **U1.ACVD-Δ** = ACVD@SDP1 - ACVD@SDP2	Paired t-test H1: G2 has more (or same) accuracy than G1 while keeping the same time saving

9.4 Sample Size Determination

We expect a minimum of 15 patients per site and per use case, for a total of 90 patients and 12 data curators across the 2 use cases.

The main objective of the study is to assess performance and acceptance of the AIDAVA prototype and provide evidence that the overarching approach - and underlying tools - can open to new possibilities in data quality enhancement and FAIRification of the data. For this proof of concept study, it is assumed that between 50 to 80 subjects will be sufficient. As we expect some drop out between the 2 epochs of the study, for G1 and for G2 of the prototype, we targeted a higher number than needed.

	Use case 1 Health record of cardiac patients	Use case 2 Breast Cancer Patient registry
NEMC - MIDATA	15 patients	15 patients
	+3 expert data curators	
MedUniGraz - MIDATA	15 patients	15 patients
	+3 expert data curators	
MUMC+/Maastro - DIGI.me	15 patients (MUMC)	15 patients (Maastro)
	+3 expert data curators	
TOTAL	**45 patients**	**45 patients**
	+ 9 expert data curators	

10. Supporting Documentation and Operational Considerations

10.1 Data sources to be curated

The following data sources are expected to be used to support the use cases.

Source	Subcomponent	Short description on information use	Breast Cancer	Cardio
EHR	Discharge Summary/Discharge letter	A handover document that explains to other healthcare professionals why the patient was admitted, what has happened to them in hospital, and other information needed to continue care	YES	YES
	Medical history	A record of information about a person's health. A personal medical history may include information about allergies, illnesses, surgeries, immunizations, and results of physical exams and tests.	YES	YES
	Progress notes	Progress notes are intended to provide an updated analysis of a patient's condition and progress during treatment.	YES	YES
	Prescribed medications	A prescription drug is a pharmaceutical that requires a medical prescription to be dispensed and their full list is provided usually at the end of a therapeutic encounter.	YES	YES
	Medical imaging reports	Contains the interpretations of images with the main goal of presenting the outcomes of the imaging procedure (e.g. for BC - MRI, Mammography, Ultrasound, PET CT/Tomography and for CVD: Cardiac CT, cardiac echo, coronagraphy, scintigraphy of the patients to physicians)	YES	YES
	Medical images	Confirmed critical ones : Mammography (20 MB/patient for 4 images) and Cardiac echo (1-2 GB in DICOM format per patient)		
	Pathology reports	Contains morphological description, diagnosis, predictive and prognostic factors of tumour and pTNM. Includes reports of cytological specimens, biopsies and surgical specimens	YES	NO
	Surgical procedure descriptions	Detailed description of surgical procedure contains type(s) and extent of surgery, specimens removed and intraoperative pathology consultation report	YES	NO
	Multidisciplinary meeting reports	Multidisciplinary meeting reports contain information about staging, treatment indications and decisions.	YES	NO
	TNM staging	TNM staging describes stage of tumour	YES	NO
	Patient referral document	Patient referral documents contain information about current symptoms and disease, medical imaging. In some cases it can be on paper.	YES	NO

Source	Subcomponent	Short description on information use	Breast Cancer	Cardio
	Laboratory reports	Describe the results of laboratory tests performed for the patient samples	NO	YES
	Echocardiography report	Echocardiography is a diagnostic tool for diagnosis and follow-up of heart disease and its report provides the interpretation of the medical imaging procedure	NO	YES
	Coronary angiography report	Coronary angiogram is a procedure that uses X-ray imaging of the heart's blood vessels and it report provides the interpretation of the medical imaging procedure	NO	YES
	Ambulance record	A record of care provided during the ambulance stage of treatment.	NO	YES
	Emergency department record	A record of care provided during the emergency department stage of treatment.	NO	YES
	ICU/CICU progress notes	Provide an updated analysis of a patient's condition and progress during the ICU stay.	NO	YES
Health Data intermediary	GP record	Contains notes and information from the GP on conditions, lab results, allergies, prescriptions. It is less detailed than any information contained in the EHR. As GP records typically cover the lifetime of a patient, they are offering a truly longitudinal perspective.	YES	YES
	Personal App	Quality of Life Questionnaire (EQ-5D 9 QALY) running on a smartphone -	YES	NO
	Connected medical device	Certified digital devices that help gather vitals (and potentially many other parameters) directly from the patient. One approved medical device will be identified for this use case.	NO	YES

10.2 Informed Consent for participation in the evaluation of the AIDAVA prototype tools

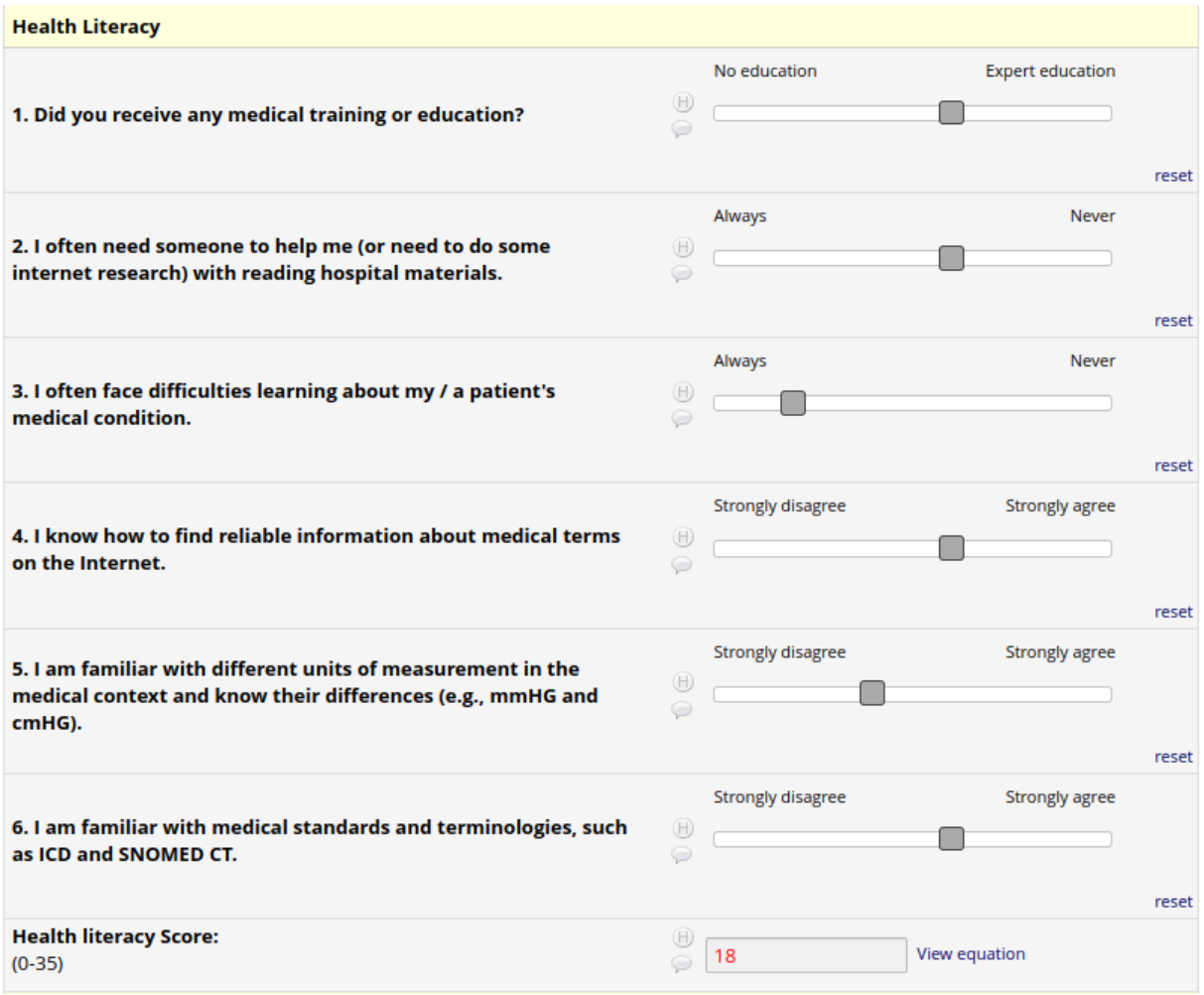
(see Annex 2)

10.3 Questionnaires and data entry forms:

10.3.1 Questionnaire to specify the user profile:

These questionnaires will be used to set up the user profile of each subject that will work with the system; it will help the system to interact with the user at the right level taking into account context information and the level of data, health and digital literacy of the user.
Considering that there is an interval of 16 to 18 months between the evaluation of G1 and of G2, users will be requested to fill the question before each evaluation.

Questions to assess medical/ health literacy



Question to assess computer/digital literacy

Digital Literacy		
1. Did you receive any computer science or data science training or education?	No education … Expert education	reset
2. I know how to access the metadata of an electronic file.	Strongly disagree … Strongly agree	reset
3. I am always keen on understanding the technical details of innovative solutions.	Strongly disagree … Strongly agree	reset
4. When I use a search engine, I can take advantage of its advanced features.	Strongly disagree … Strongly agree	reset
5. I know how to protect myself from unwanted and malicious online encounters and materials.	Strongly disagree … Strongly agree	reset
6. When I face a technical problem, I am able to find solutions on the internet.	Strongly disagree … Strongly agree	reset
Digital literacy Score: (0-35)	16	View equation

10.3.2 Site monitoring forms

List of people participating per site (in addition to patients)

This includes information on the following participants

- Research associate coordinating the study locally
- Expert curator supporting the patient in the curation process (one of them can be the same person than the research associate
- Breast Cancer Specialist and CVD treating physician: medical staff responsible for the study
- HelpDesk Level 1 - local contact for support (in local language)
- HelpDesk Level 2 - contact within the development team for support, in case the Level 1 contact person cannot resolve the issue locally

People participating per site (the form is filled out by research associate)

Generation of the prototype
- ○ Generation 1
- ○ Generation 2

reset

Site name:
- ○ NEMC - Sihtasutus Põhja-Eesti Regionaalhaigla
- ○ UM - Maastricht University
- ○ MUG - Medizinische Universität Graz

reset

Data Protection Impact Assessment (DPIA) available:
- ○ Yes
- ○ No (if not, provide data at which it is expected to be performed)

reset

Today Y-M-D

Research Associates

Research Associate 1:

Name: (example: John Smith)

Email-contact: (example: john.smith@gmail.com)

ID / pseudonym: (example: NEMCBCC1 , NEMCCVC1, ...)

Expert Curator role?
- ○ Yes
- ○ No

reset

Research Associate 2:

Name: (example: John Smith)

Email-contact: (example: john.smith@gmail.com)

ID / pseudonym: (example: NEMCBCC1 , NEMCCVC1, ...)

Expert Curator role?
- ○ Yes
- ○ No

reset

Expert data curators	
Expert data curator 1:	
Name:	(example: John Smith)
Email-contact:	(example: john.smith@gmail.com)
ID / pseudonym:	(example: NEMCBCC1 , NEMCCVC1, ...)
Expert data curator 2:	
Name:	(example: John Smith)
Email-contact:	(example: john.smith@gmail.com)
ID / pseudonym:	(example: NEMCBCC1 , NEMCCVC1, ...)
Expert data curator 3:	
Name:	(example: John Smith)
Email-contact:	(example: john.smith@gmail.com)
ID / pseudonym:	(example: NEMCBCC1 , NEMCCVC1, ...)

Breast cancer specialists	
Breast cancer specialist 1:	
Name:	(example: John Smith)
Email-contact:	(example: john.smith@gmail.com)
ID / pseudonym:	(example: NEMCBCC1 , NEMCCVC1, ...)
Breast cancer specialist 2:	
Name:	(example: John Smith)
Email-contact:	(example: john.smith@gmail.com)
ID / pseudonym:	(example: NEMCBCC1 , NEMCCVC1, ...)
CVD specialists	
CVD specialist 1:	
Name:	(example: John Smith)
Email-contact:	(example: john.smith@gmail.com)
ID / pseudonym:	(example: NEMCBCC1 , NEMCCVC1, ...)
CVD specialist 2:	
Name:	(example: John Smith)
Email-contact:	(example: john.smith@gmail.com)
ID / pseudonym:	(example: NEMCBCC1 , NEMCCVC1, ...)

Help Desk	
Help Desk level 1: (local contact in local language)	
Name:	(example: John Smith)
Email-contact:	(example: john.smith@gmail.com)
WhatsApp Account:	(example: +43 664 1234567)
Help Desk level 2: (local contact in local language)	
Name:	(example: John Smith)
Email-contact:	(example: john.smith@gmail.com)
WhatsApp Account:	(example: +43 664 1234567)

Patients Screening (including eligibility criteria and follow-up)

Form to be used to screen the approximately up to 50 patients per site/ therapeutic areas; out of these 50, only 15 will actually be enrolled.

VISIT 0 (on site): Patient Screening Form (the form is filled out by research associate)	
Generation of the prototype	○ Generation 1 ○ Generation 2 reset
Data of the patient and the research associate (the form is filled out by research associate)	
Site name:	○ NEMC - Sihtasutus Põhja-Eesti Regionaalhaigla ○ UM - Maastricht University ○ MUG - Medizinische Universität Graz reset
Therapeutic Area:	○ Breast Cancer ○ Cardiovascular reset
Patient pseudonym: *Each patient will be assigned a pseudonym that will be composed by the concatenation of* **the acronym for the site (NEMC, MUG, UMUC)* **the acronym for the therapeutic area (BC, CV)* **a sequential number based on the sequence of enrollment* *The full identification information of the patient will be kept separately by the Research associate; the Research associate is also responsible for reentering the same pseudonym during G2 - for patients that participated in Generation 1.*	(example: NEMCBCP1 , NEMCCVP1, ...)

Screening	
Date of Screening:	Today Y-M-D
Inclusion criteria (BC and CVD patients)	
Data available in hospital systems and/or in patient registry within related medical center?	○ Yes ○ No reset
Is the participant an owner and user of a smartphone?	○ Yes ○ No reset
Consent to give access to personal data (based on scope provided in the ICF) to the hospital data curator?	○ Yes ○ No reset
Agree to sign a collaboration agreement with the relevant HDI - for the duration of the study?	○ Yes ○ No reset
Agree to test the AIDAVA prototype (G1 and G2) and to delegate to a representative data curator in case of issues?	○ Yes ○ No reset
Is the patient comfortable with using digital devices and computers?	○ Yes ○ No reset
Exclusion criteria (BC and CVD patients)	
Vulnerable person and/or children/minor?	○ Yes ○ No reset
Inability/unwillingness to use smartphone applications?	○ Yes ○ No reset

Conclusion (BC and CVD patients)	
All inclusion and exclusion criteria checked?	○ Yes ○ No
Project introduction done?	○ Yes ○ No
Informed Consent Form signed?	○ Yes ○ No
All mandatory checkboxes in Informed Consent Form filled?	○ Yes ○ No
Comments from patient (if any):	
Comments from study nurse (if any):	
Conclusion:	○ Screening failure ○ Patient enrolled

Signature of ICF and agreement (for successfully screened patient - to be enrolled)

Informed Consent Form signed?	○ Yes ○ No reset
HDI agreement signed?	○ Yes ○ No reset

Data extraction (enrolled patients only - after V0 and before V1)

Data Extraction from Hospital and transfer to AIDAVA: initiated on (date of the first transfer)	Today Y-M-D
Data Extraction from Hospital and transfer to AIDAVA: any issue?	(free text)
Data Extraction from HDI and transfer to AIDAVA: initiated on (date of the first actual transfer of data)	Today Y-M-D
Data Extraction from Hospital and transfer to AIDAVA: any issue?	(free text)

Patients monitoring information

Form to be used to monitor each patient once then have been enrolled, across the different visits in G1 and G2

GENERATION 1 and GENERATION 2

VISIT 1 (on site): Patient monitoring form (the form is filled out by research associate)

Generation of the prototype	○ Generation 1 ○ Generation 2 reset
Site name:	○ NEMC - Sihtasutus Põhja-Eesti Regionaalhaigla ○ UM - Maastricht University ○ MUG - Medizinische Universität Graz reset
Therapeutic Area:	○ Breast Cancer ○ Cardiovascular reset

Data of the patient and the research associate (the form is filled out by research associate)

Patient pseudonym:	_____
Patient email adress: *Needed to contact the patient for an on-line meeting invite. Note that the patient identification can be known to the study participants - but we still prefer to keep a pseudonym whenever possible, except for the contact email.*	(example: john.smith@gmail.com)
Name of research associate: *(need to be included in the list of participants)*	
Name of expert data curator: *(need to be included in the list of participants)*	

Visit 1 - on site (Week 1/Day 1) (the form is filled out by research associate)

Date of Visit:	Today Y-M-D
Health and digital literacy levels (see questionnaire Health and digital literacy):	
Health Literacy Score:	5
Digital Literacy Score:	_____
ePRO-QLY (EQ-5D) filled in through person apps?	Yes / No reset
Training session for the patient on the prototype done?	Yes / No reset
Comments:	(free text)

VISIT 1 (Week 1/Day 5) - Data extraction / calculation (the form is filled by research associate)

Date:	Today Y-M-D
Data curation (Week 1) by patient done?	Yes / No reset
Data curation (Week 1) by expert curator done?	Yes / No reset
Did the patient check their International Patient Summary extract (transferred to HDI automatically)?	Yes / No reset
Automatic extraction BC registry, query done (BC patients) and viewed by BC specialists?	Yes / No reset
Comments:	(free text)

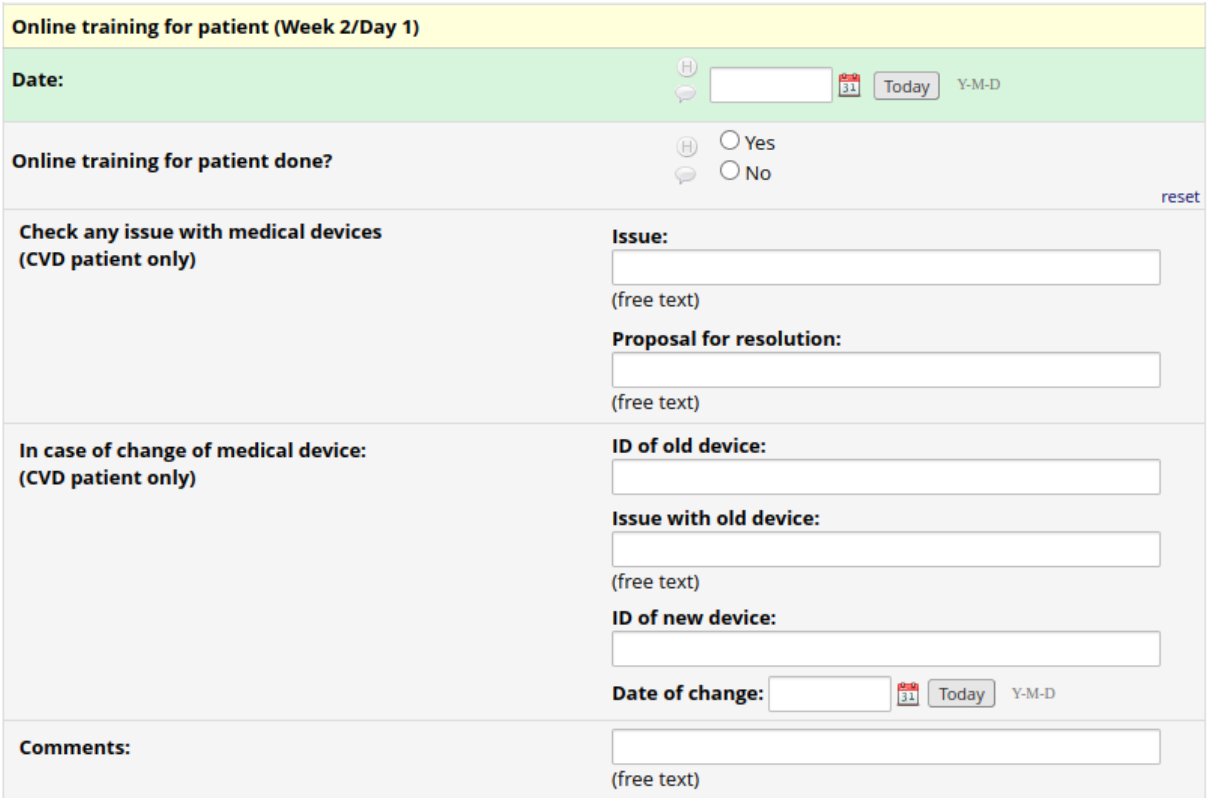
For BC patients

VISIT 1 (Week 1/Day 5) - Data extraction / calculation (the form is filled by research associate)	
Date:	Today Y-M-D
Data curation (Week 1) by patient done?	Yes / No reset
Data curation (Week 1) by expert curator done?	Yes / No reset
How many times did you check your blood pressure with the BPM during the last week (expected to be a minimum of 3 times during the week, 3 times a day)?	(e.g. 2 or 3 how many times during the week) (e.g. frequency of each measurement)
Did the patient check their International Patient Summary extract (transferred to HDI automatically)?	Yes / No reset
SMART score (CVD patients) available and **- viewed by CVD specialists** **- calculated by system**	Yes / No reset
Comments:	(free text)

For CVD patients

Online training for patient (Week 2/Day 1)	
Date:	Today Y-M-D
Online training for patient done?	Yes / No reset
Comments:	(free text)

For BC patients



For CVD patients

Visit 3 - Data extraction / calculation (Week 2/Day 5) (the form is filled by research associate)	
Date:	Today Y-M-D
Data curation (Week 2) by patient done?	Yes / No reset
Data curation (Week 2) by expert curator done?	Yes / No reset
Automatic extraction BC registry, query done (BC patients only) and viewed by BC specialists?	Yes / No reset
Patient Curator form	**Incomplete**
Expert Data Curator Form	**Complete**
Data user/consumer (Breast Cancer Specialist) Form	**Complete**
System Usability Scale	**Complete**
Comments:	(free text)

For BC patients

Visit 3 - Data extraction / calculation (Week 2/Day 5) (the form is filled by research associate)	
Date:	Today Y-M-D
Data curation (Week 2) by patient done?	Yes / No reset
Data curation (Week 2) by expert curator done?	Yes / No reset
SMART score (CVD patients) available and - viewed by CVD specialists - calculated by system	Yes / No reset
Patient Curator form	**Complete**
Expert Data Curator Form	**Complete**
Data user/consumer (CVD Specialist) Form	**Complete**
System Usability Scale	**Complete**
Comments:	(free text)

For CVD patients

Visit 4 - Follow up online (Week 4) (the form is filled out by research associate)	
Date:	Today Y-M-D
Follow-up online meeting done?	Yes / No reset
Comments:	(free text)
Confirm data have been deleted from hospital:	Yes / No reset
Dropping out of the study / Death (the form is filled out by research associate)	
The patient reported withdrawing from the study:	Today Y-M-D
It was not possible to contact the patient. Last contact attempt date:	Today Y-M-D
Date of death:	Today Y-M-D

For BC patients

Visit 4 - Follow up online (Week 4) (the form is filled out by research associate)	
Date:	Today Y-M-D
Follow-up online meeting done?	Yes / No reset
Comments:	(free text)
Confirm data have been deleted from hospital:	Yes / No reset
Dropping out of the study / Death (the form is filled out by research associate)	
The patient reported withdrawing from the study:	Today Y-M-D
It was not possible to contact the patient. Last contact attempt date:	Today Y-M-D
Date of death:	Today Y-M-D
Medical device returned: (CVD patient only)	Yes / No reset

For CVD patients

10.3.3 Patient and Expert Curator Data Collection forms (and related reporting tables)

System usability Scale (Standard version)

Form to be used by the patient and expert curator to assess usability of the AIDAVA prototype in G1 and G2.

System Usability Scale - SUS

Generation of the AIDAVA prototype:
(not on the other tools like QLY or HDI agreement)
○ Generation 1
○ Generation 2
reset

Site name:
○ NEMC - Sihtasutus Põhja-Eesti Regionaalhaigla
○ UM - Maastricht University
○ MUG - Medizinische Universität Graz
reset

Date: Today Y-M-D

User pseudonym:
(example: NEMCBCP1 , NEMCCVP2, ...)

User role:
○ Patient
○ Expert Data Curator
reset

1. I am ready to use this system frequently.
* must provide value
Strongly disagree — Strongly agree: 4
reset

2. I found the system unnecessarily complex.
* must provide value
Strongly disagree — Strongly agree: 2
reset

3. I thought the system was easy to use.
* must provide value
Strongly disagree — Strongly agree: 4
reset

4. I would need the support of a technical person to be able to use this system.
* must provide value
Strongly disagree — Strongly agree: 2
reset

5. I found the various functions in this system were well integrated.
* must provide value
Strongly disagree — Strongly agree: 5
reset

6. There was too much inconsistency in this system.
* must provide value

Strongly disagree — Strongly agree: 1 reset

7. Most people would learn to use this system very quickly.
* must provide value

Strongly disagree — Strongly agree: 4 reset

8. I found the system very awkward to use.
* must provide value

Strongly disagree — Strongly agree: 2 reset

9. I felt very confident using the system.
* must provide value

Strongly disagree — Strongly agree: 4 reset

10. I needed to learn a lot of things before I could get going with the system.
* must provide value

Strongly disagree — Strongly agree: 2 reset

SUS Score:
(0-100)

80 View equation

Explainability/Causability evaluation

1. For me it is important to know, where the different parts of the curated health data in AIDAVA are coming from.
* must provide value

Strongly disagree — Strongly agree
Change the slider above to set a response
reset

In my opinion, information regarding this aspect is sufficient in AIDAVA.
* must provide value

Strongly disagree — Strongly agree
Change the slider above to set a response
reset

2. For me it is important to know, who has curated my health data and which tools have been used during the curation process.
* must provide value

Strongly disagree — Strongly agree
Change the slider above to set a response
reset

In my opinion, information regarding this aspect is sufficient in AIDAVA.
* must provide value

Strongly disagree — Strongly agree
Change the slider above to set a response
reset

3. For me it is important to know, whether a health data curation method has been applied by a human or an algorithm.
* must provide value

Strongly disagree — Strongly agree
Change the slider above to set a response
reset

In my opinion, information regarding this aspect is sufficient in AIDAVA.
* must provide value

Strongly disagree — Strongly agree
Change the slider above to set a response
reset

AIDAVA specific questions	
1. I would recommend AIDAVA to my friend, colleague, or family member.	Strongly disagree — Strongly agree Change the slider above to set a response reset
2. I am ready to work with AIDAVA when available on the market.	Strongly disagree — Strongly agree Change the slider above to set a response reset
3. I understand the purpose of data curation.	Strongly disagree — Strongly agree Change the slider above to set a response reset
4. I am ready to spend the needed time to ensure proper data curation of my data.	Strongly disagree — Strongly agree Change the slider above to set a response reset
5. This system is unique and different from anything else available.	Strongly disagree — Strongly agree Change the slider above to set a response reset
6. This system will allow me to manage my health records better.	Strongly disagree — Strongly agree Change the slider above to set a response reset
Please write down any suggestions you might have on how we could improve/ make AIDAVA easier to use.	Expand

Expert Data curators Form

Form to be used by the expert curator to enter performance metrics needed to assess the AIDAVA prototype in G1 and G2.

Expert Data Curator Form - G1 & G2

Generation of the prototype	○ Generation 1 ○ Generation 2 reset
Site name:	○ NEMC - Sihtasutus Põhja-Eesti Regionaalhaigla ○ UM - Maastricht University ○ MUG - Medizinische Universität Graz reset
Expert Curator ID / pseudonym:	
Number of patients supported:	(example: 10 or 12 ..)
Period of time during which patient records were supported:	
Start:	Today Y-M-D
End:	Today Y-M-D
Number of days:	View equation

M2

Estimated time spent working with the system to perform ingestion and curation:

Week:	Date:	Hours:
Week 1	2024-07-01 - 2024-07-07	
Week 2	2024-07-08 - 2024-07-14	
Week 3	2024-07-15 - 2024-07-21	
Week 4	2024-07-22 - 2024-07-28	
Week 5	2024-07-29 - 2024-08-04	
Week 6	2024-08-05 - 2024-08-11	
Week 7	2024-08-12 - 2024-08-18	
Week 8	2024-08-19 - 2024-08-25	
Week 9	2024-08-26 - 2024-09-01	
Week 10	2024-09-02 - 2024-09-08	
Week 11	2024-09-09 - 2024-09-15	
Week 12	2024-09-16 - 2024-09-22	
Week 13	2024-09-23 - 2024-09-29	
Week 14	2024-09-30 - 2024-10-06	

G1 - Total Time (ingestion and curation):	View equation

M2

Estimated time spent working with the system for other activities (training, visit, visualization):

Week:	Date:	Hours:
Week 1	2024-07-01 - 2024-07-07	
Week 2	2024-07-08 - 2024-07-14	
Week 3	2024-07-15 - 2024-07-21	
Week 4	2024-07-22 - 2024-07-28	
Week 5	2024-07-29 - 2024-08-04	
Week 6	2024-08-05 - 2024-08-11	
Week 7	2024-08-12 - 2024-08-18	
Week 8	2024-08-19 - 2024-08-25	
Week 9	2024-08-26 - 2024-09-01	
Week 10	2024-09-02 - 2024-09-08	
Week 11	2024-09-09 - 2024-09-15	
Week 12	2024-09-16 - 2024-09-22	
Week 13	2024-09-23 - 2024-09-29	
Week 14	2024-09-30 - 2024-10-06	

G1 - Total Time (training, visit, visualization): View equation

For Generation 1

M2

Estimated time spent working with the system to perform ingestion and curation:

Week:	Date:	Hours:
Week 1	2026-02-01 - 2026-02-07	
Week 2	2026-02-08 - 2026-02-14	
Week 3	2026-02-15 - 2026-02-21	
Week 4	2026-02-22 - 2026-02-28	
Week 5	2026-03-01 - 2026-03-07	
Week 6	2026-03-08 - 2026-03-14	
Week 7	2026-03-15 - 2026-03-21	
Week 8	2026-03-22 - 2026-03-28	
Week 9	2026-03-29 - 2026-04-04	
Week 10	2026-04-05 - 2026-04-11	
Week 11	2026-04-12 - 2026-04-18	
Week 12	2026-04-19 - 2026-04-25	
Week 13	2026-04-26 - 2026-05-02	
Week 14	2026-05-03 - 2026-05-09	

G2 - Total Time (ingestion and curation): View equation

M2

Estimated time spent working with the system for other activities (training, visit, visualization):

Week:	Date:	Hours:
Week 1	2026-02-01 - 2026-02-07	
Week 2	2026-02-08 - 2026-02-14	
Week 3	2026-02-15 - 2026-02-21	
Week 4	2026-02-22 - 2026-02-28	
Week 5	2026-03-01 - 2026-03-07	
Week 6	2026-03-08 - 2026-03-14	
Week 7	2026-03-15 - 2026-03-21	
Week 8	2026-03-22 - 2026-03-28	
Week 9	2026-03-29 - 2026-04-04	
Week 10	2026-04-05 - 2026-04-11	
Week 11	2026-04-12 - 2026-04-18	
Week 12	2026-04-19 - 2026-04-25	
Week 13	2026-04-26 - 2026-05-02	
Week 14	2026-05-03 - 2026-05-09	

G2 - Total Time (training, visit, visualization): View equation

For Generation 2

This form is linked with a table that contains all the metrics manually entered in REDCap (as .csv-export) that are needed to evaluate the system from an expert curator perspective.

Record ID record_id	Generation of the prototype generation_edcf_2	Site name: site_name_edcf_2	Expert curator ID / pseudonym: pseudonym_edcf_2
MUG_BC_Test	Generation 1 (1)	MUG - Medizinische Universität Graz (3)	MUGBCC4 (2)
MUG_CVD_Test	Generation 2 (2)	MUG - Medizinische Universität Graz (3)	MUGCVC3 (1)

G1 - Total Time (ingestion and curation) in minutes: edcf_2_g1_total_min	G1 - Total Time (training, visit, visualization) in minutes: edcf_2_g1_total_min_2	G2 - Total Time (ingestion and curation) in minutes: edcf_2_g2_total_min	G2 - Total Time (training, visit, visualization) in minutes: edcf_2_g2_total_min_2
2160	4200		
		2580	3780

Patient Data curators Forms

Form to be used by the patient - as a curator - to enter performance metrics needed to assess the AIDAVA prototype in G1 and G2.

Patient Data Curator Form - G1 & G2

Generation of the prototype	○ Generation 1 ○ Generation 2 reset
Site name:	○ NEMC - Sihtasutus Põhja-Eesti Regionaalhaigla ○ UM - Maastricht University ○ MUG - Medizinische Universität Graz reset
Therapeutic Area:	○ Breast Cancer ○ Cardiovascular reset
Patient ID / pseudonym:	MUGCVP1
Period of curation:	
Start:	2023-09-01 Today Y-M-D
End:	2023-09-14 Today Y-M-D

M2

Estimated time spent working with the system to perform ingestion and curation:

Day:	Date:		Hours	:	Minutes
Day 1	2023-09-01	View equation	2	:	15
Day 2	2023-09-02	View equation	3	:	00
Day 3	2023-09-03	View equation	5	:	45
Day 4	2023-09-04	View equation	5	:	30
Day 5	2023-09-05	View equation	1	:	10
Day 6	2023-09-06	View equation	2	:	05
Day 7	2023-09-07	View equation	1	:	00
Day 8	2023-09-08	View equation	3	:	30
Day 9	2023-09-09	View equation	4	:	00
Day 10	2023-09-10	View equation	1	:	00
Day 11	2023-09-11	View equation	3	:	15
Day 12	2023-09-12	View equation	2	:	20
Day 13	2023-09-13	View equation	3	:	10
Day 14	2023-09-14	View equation	1	:	10

Total Time in Minutes (ingestion and curation): 2350 View equation

M2

Estimated time spent working with the system for other activities (training, visit, visualization):

Day:	Date:		Hours	:	Minutes
Day 1	2023-09-01	View equation	2	:	10
Day 2	2023-09-02	View equation	1	:	00
Day 3	2023-09-03	View equation	1	:	20
Day 4	2023-09-04	View equation	1	:	20
Day 5	2023-09-05	View equation	0	:	10
Day 6	2023-09-06	View equation	0	:	15
Day 7	2023-09-07	View equation	1	:	15
Day 8	2023-09-08	View equation	1	:	00
Day 9	2023-09-09	View equation	0	:	45
Day 10	2023-09-10	View equation	2	:	00
Day 11	2023-09-11	View equation	1	:	15
Day 12	2023-09-12	View equation	0	:	30
Day 13	2023-09-13	View equation	0	:	45
Day 14	2023-09-14	View equation	0	:	15

Total Time in Minutes (training, visit, visualization): 840 View equation

This form is linked with a table that contains all the metrics manually entered in REDCap (as .csv-export) that are needed to evaluate the system from a patient - as a curator - perspective.

Record ID record_id	Generation of the prototype generation_pdcf	Site name: site_name_pdcf	Therapeutic Area: ta_pdcf
MUG_BC_Test	Generation 1 (1)	MUG - Medizinische Universität Graz (3)	Breast Cancer (1)
MUG_CVD_Test	Generation 1 (1)	MUG - Medizinische Universität Graz (3)	Cardiovascular (2)

Total Time in Minutes (ingestion and curation): pdcf_time_total	Total Time in Minutes (training, visit, visualization): pdcf_time_total_2
1975	910
2350	840

10.3.4 Data user/consumers spreadsheet: Breast Cancer Specialist and CVD clinicians

Breast Cancer

Form to be used by the Breast Cancer Specialist to enter time related to management of the BC registry

Breast Cancer Specialist - Data User Form - G0

Site name:
- ○ NEMC - Sihtasutus Põhja-Eesti Regionaalhaigla
- ○ UM - Maastricht University
- ○ MUG - Medizinische Universität Graz

reset

Data User ID / pseudonym:

To be computed for 45 patients across sites
(15 patients in your own site + time to get info on 30 patients in the other 2 sites)

TBC1. Estimated time to compute the percentage of patients - across sites - treated with breast conserving therapy and whole breast radiotherapy that receive a boost to the tumour bed without AIDAVA (in min):
(example: 2 or 6 or ..)

TBC2. Estimated time to compute the number of patients - across sites - undergoing surgery and radiotherapy, without adjuvant chemotherapy: % of patients with time between surgery and radiotherapy > 8 weeks without AIDAVA (in min):
(example: 2 or 6 or ..)

TBC3. Estimated time to compute the number of patients - across sites - receiving radiotherapy: Percentage of patients receiving any nodal radiotherapy without AIDAVA (in min):
(example: 2 or 6 or ..)

This form is linked with a table that contains all the metrics manually entered in REDCap (as .csv-export) that are needed to evaluate the system from a BC specialist perspective.

Breast Cancer	G1					G2			
	BC1	BC2	BC3	Comments		BC1	BC2	BC3	Comments
NEMC									
MUG									
Maastro									
MEAN									
MEDIAN									

CVD specialist

Form to be used by the CVD specialist to enter the manually computed SMArt score and the time needed to compute this form manually

CVD Specialist - Data User Form - G1 & G2	
Generation of the prototype	○ Generation 1 ○ Generation 2 reset
Site name:	○ NEMC - Sihtasutus Põhja-Eesti Regionaalhaigla ○ UM - Maastricht University ○ MUG - Medizinische Universität Graz reset
Data User ID / pseudonym:	
Patient ID / pseudonym:	_____
Date:	Today Y-M-D
CVD1. SMART risk score computed by the physician(s) for the patient (from this site): **U-prevent SMART risk score calculator**	SMART risk score computed by CVD Specialist SMART risk score computed by another person Name of this other person
TCVD1. Time required for the treating physician to compute the SMART risk score (collect + analyze + enter data into a standalone web page calculator) in min:	(example: 10 or 20 or ..)
How useful was the patient status graphical display (with balloons) - to assess the status of the patient: - to communicate with the patient on their status:	(scale 1 to 5) (scale 1 to 5)
Please enter any comments related to the graphical display:	(free text)

This form is linked with a table that contains all the metrics manually entered in REDCap (as .csv-export) that are needed to evaluate the system from a CVD specialist perspective.

CVD	G1					G2			
	CVD1	TCVD1		Comments		CVD2	TCVD2		Comments
NEMC									
MUG									
UMUC									
MEAN									
MEDIAN									

10.4 Feedback to the patient

- Communication every 2 months

- **Objective**: provide regular information on the project to maintain interest the patient to decrease dropout during the G2 development/ improvement phase
- **Medium**: Regular newsletters (send through emails; additional online meetings after 9 months will be explored).
- **Expected planning and content**
    - Across all newsletters: in each communication include a message highlighting the advantages of using AIDAVA, key milestones and achievements in the project, the value of patients' participation in the evaluation, and the important steps the patients can take to improve the quality of care by using it.
    - Proposed focus per newsletters

Number	Time	Content
1	Oct 2024	Thank you - highlights from evaluation
2	Jan 2025	Comments from Ethical Advisory Board and data privacy aspects Comments Sustainability Advisory Board on potential development and future use of the application
3	Apr 2025	Perspective from Health Data Intermediaries on further extension and use of AIDAVA like tools. How does this fit in the wider perspective of EHDS
4	Jul 2025	Progress on development of G2 - how patients feedback was key to improving features for G2. What is coming in G2 and what we expect from patients and
5	Oct 2025	G2 is coming - confirmation of new features - we count on you to help evaluate this

10.5. List of Data Elements composing the Breast Cancer Registry

The table below represents the proposed list of DE for the Breast Cancer registry. It is based on the expertise of the Breast Cancer specialist from the different sites; it is however not expected to be considered as a final list ready for scientific validation. This could be done - if relevant - after we demonstrate how it could be automatically generated through the AIDAVA publishing module.

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
Patient ID	422549004 |Patient-related Identification code (observable entity)|	398093005 |Social security number (observable entity)|	*Value*	423901009 |Identification code (observable entity)| 398093005 |Social security number (observable entity)|	
Age at diagnosis	423493009 |Age at diagnosis (observable entity)|	Clinical diagnosis (contextual qualifier) (qualifier value)	*Number*	410680006 |Number (attribute)|	
Sex	184100006 |Patient sex (observable entity)|		Female	248152002 |Female (finding)|	
			Male	248153007 |Male (finding)|	
Prior ipsilateral BC	429087003 |History of malignant neoplasm of breast (situation)|	255208005 |Ipsilateral (qualifier value)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
Lesion ID	309049000 |Lesion specimen (specimen)|		*Value*		
Radiological tumor size	263605001 |Length dimension of neoplasm (observable entity)|	363679005 |Imaging (procedure)|	*Millimeter*	246115007 |Size (attribute)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
Laterality	76752008 |Breast structure (body structure)|	272741003 |Laterality (attribute)|	Right	24028007 |Right (qualifier value)|	73056007 |Right breast structure (body structure)|
			Left	7771000 |Left (qualifier value)|	80248007 |Left breast structure (body structure)|
			Unknown	373068000 |Undetermined (qualifier value)|	312863000 |Patient data not recorded (finding)|
Clinical T-stage		1222585009 |American Joint Committee on Cancer clinical T category allowable value|	T0 - No evidence of primary tumour	1228882005 |American Joint Committee on Cancer cT0 (qualifier value)|	
			T1a - Tumour > 1 mm but <= 5 mm in greatest dimension	1228892002 |cT1a|1228892002 |American Joint Committee on Cancer cT1a (qualifier value)|	
			T1b - Tumour > 5 mm but <= 10 mm in greatest dimension	1228895000 |American Joint Committee on Cancer cT1b (qualifier value)|	
			T1c - Tumour > 10 mm but <= 20 mm in greatest dimension	1228899006 |American Joint Committee on Cancer cT1c (qualifier value)|	
			T1mi - Tumour <= 1 mm in greatest dimension	1228891009 |American Joint Committee on Cancer cT1mi (qualifier value)|	
			T1NOS - Tumour <= 20 mm in greatest dimension	1228889001 |American Joint Committee on Cancer cT1 (qualifier value)|	
			T2 - Tumour > 20 mm but <= 50 mm in greatest dimension	1228929004 |American Joint Committee on Cancer cT2 (qualifier value)|	
			T3 - Tumour > 50 mm in greatest dimension	1228938002 |American Joint Committee on Cancer cT3 (qualifier value)|	
			T4a - Extension to the chest wall, not including only pectoralis muscle adherence/invasion	1228945002 |American Joint Committee on Cancer cT4a (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
			T4b - Ulceration and/or ipsilateral satellite nodules and/or oedema (including peau d'orange) of the skin which do not meet the criteria for inflammatory carcinoma	1228946001 |American Joint Committee on Cancer cT4b (qualifier value)|	
			T4c - Both T4a and T4b	1228947005 |American Joint Committee on Cancer cT4c (qualifier value)|	
			T4d - Inflammatory carcinoma	1228948000 |American Joint Committee on Cancer cT4d (qualifier value)|	
			T4NOS - Tumour of any size with direct extension to the chest wall and/or to the skin (ulceration or skin nodules)	1228944003 |American Joint Committee on Cancer cT4 (qualifier value)|	
			Tis - Carcinoma in situ	1228884006 |American Joint Committee on Cancer cTis (qualifier value)|	
			Tis (DCIS) - Ductal carcinoma in situ	1228885007 |American Joint Committee on Cancer cTis(DCIS) (qualifier value)|	
			Tis (LCIS) - Lobular carcinoma in situ	1228884006 |American Joint Committee on Cancer cTis (qualifier value)|	109888004 |Lobular carcinoma in situ of breast (disorder)|
			Tis (Paget's) - Paget's disease of the nipple is NOT associated with invasive carcinoma and/or carcinoma in situ (DCIS and/or LCIS) in the underlying breast parenchyma. Carcinomas in the breast parenchyma associated with Paget disease are categorized based on the size and characteristics of the parenchymal disease, although the presence of Paget disease should still be noted.	1228888009 |American Joint Committee on Cancer cTis(Paget) (qualifier value)|	
			TX - Primary tumour cannot be assessed	1222604002 |American Joint Committee on Cancer cTX (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
Clinical N-stage	399534004 |cN category (observable entity)|	1222588006 |American Joint Committee on Cancer clinical N category allowable value (qualifier value)|	N0 - No regional lymph node metastases	1229967007 |American Joint Committee on Cancer cN0 (qualifier value)|	
			N1mi - Micrometastases (approximately 200 cells, larger than 0.2 mm, but none larger than 2.0 mm)	1229974002 |American Joint Committee on Cancer cN1mi (qualifier value)|	
			N1NOS - Metastases to movable ipsilateral level I, II axillary lymph node(s)	1229973008 |American Joint Committee on Cancer cN1 (qualifier value)|	
			N2a - Metastases in ipsilateral axillary lymph nodes fixed to one another (matted) or to other structures	1229981009 |American Joint Committee on Cancer cN2a (qualifier value)|	
			N2b - Metastases only in clinically detected ipsilateral internal mammary nodes and in the absence of clinically evident axillary lymph node metastases	1229982002 |American Joint Committee on Cancer cN2b (qualifier value)|	
			N2NOS - Metastases in ipsilateral level I, II axillary lymph nodes that are clinically fixed or matted; or in clinically detected ipsilateral internal mammary nodes in the absence of clinically evident axillary lymph node metastasis	1229978004 |American Joint Committee on Cancer cN2 (qualifier value)|	
			N3a - Metastases in ipsilateral infraclavicular lymph node(s)	1229985000 |American Joint Committee on Cancer cN3a (qualifier value)|	
			N3b - Metastases in ipsilateral internal mammary lymph node(s) and axillary lymph node(s)	1229986004 |American Joint Committee on Cancer cN3b (qualifier value)|	
			N3c - Metastases in ipsilateral supraclavicular lymph node(s)	1229987008 |American Joint Committee on Cancer cN3c (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
			N3NOS - Metastases in ipsilateral infraclavicular (level III axillary) lymph node(s) with or without level I, II axillary lymph node involvement; or in clinically detected ipsilateral internal mammary lymph node(s) with clinically evident level I, II axillary lymph node metastasis; or metastasis in ipsilateral supraclavicular lymph node(s) with or without axillary or internal mammary lymph node involvement	1229984001 |American Joint Committee on Cancer cN3 (qualifier value)|	
			NX - Regional lymph nodes cannot be assessed (e.g., previously removed)	1229966003 |American Joint Committee on Cancer cNX (qualifier value)|	
Clinical M-stage	399387003 |cM category (observable entity)|	1222587001 |American Joint Committee on Cancer clinical M category allowable value (qualifier value)|	M0 - No distant metastasis	1229901006 |American Joint Committee on Cancer cM0 (qualifier value)|	
			M1 - Distant detectable metastases as determined by classic clinical and radiographic means and/or histologically proven larger than 0.2 mm	1229903009 |American Joint Committee on Cancer cM1 (qualifier value)|	
Date of diagnosis	432213005 |Date of diagnosis (observable entity)|	373793006 |Basis of cancer diagnosis (finding)| 258415003 |Biopsy specimen (specimen)|	Date	410671006 |Date (attribute)|	
Number of nodes resected	444025001 |Number of lymph nodes examined by microscopy in excised specimen (observable entity)|		Number	410681005 |Count of entities (property) (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
Number of involved nodes	444384007 |Number of regional lymph nodes containing metastatic neoplasm in excised specimen (observable entity)|		Number	410681005 |Count of entities (property) (qualifier value)|	
Conclusion lymph node mapping	443497002 |Excision of sentinel lymph node (procedure)|		Negative	261985008 |No metastases|	1163303005 |Structure of sentinel lymph node (body structure)|
			Positive for malignant cells ≤ 0.2 mm (itc)	444510006 |Number of lymph nodes containing isolated metastatic neoplastic cells in excised specimen (observable entity)|	
			Positive for malignant cells > 0.2 and < 2.0 mm	444511005 |Number of lymph nodes containing micrometastases in excised specimen (observable entity)|	
			Positive for malignant cells ≥ 2.0 mm	444422007 |Number of lymph nodes containing macrometastases in excised specimen (observable entity)|	
			Not performed	262008008 |Not performed (qualifier value)|	
Date of diagnosis invasive / insitu component	399737008 |Date of pathology report (observable entity)|	126926005 |Neoplasm of breast (disorder)| 397199005 |Specimen from breast obtained by excision (specimen)|	Date	399651003 |Date of report (observable entity)|	410671006 |Date (attribute)|
Histology invasive breast cancer	250537006 |Histopathology finding (finding)|		Invasive breast carcinoma of no special type (invasive ductal carcinoma, not otherwise specified)	82711006 |Infiltrating duct carcinoma (morphologic abnormality)|	8500/3

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
	713609000 |Invasive carcinoma of breast (disorder)|		Invasive lobular carcinoma	89740008 |Lobular carcinoma (morphologic abnormality)|	8520/3
			Tubular carcinoma	4631006 |Tubular adenocarcinoma (morphologic abnormality)|	8211
			Cribriform carcinoma	30156004 |Cribriform carcinoma (morphologic abnormality)|	8201
			Mucinous carcinoma	72495009 |Mucinous adenocarcinoma (morphologic abnormality)|	8480/3
			Invasive micropapillary carcinoma	703578005 |Invasive micropapillary carcinoma of breast (morphologic abnormality)|	8507
			Carcinoma with apocrine differentiation	57141000 |Apocrine adenocarcinoma (morphologic abnormality)|	8401
			Metaplastic carcinoma	128705006 |Metaplastic carcinoma (morphologic abnormality)|	8575

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
			Other	713609000 |Invasive carcinoma of breast (disorder)|	8504/3 8290/3 8200/3 8502/3 8525/3 8315/3 8315/3 8314/3 8509/3 8430/3 8503/3 8522/3 8523/3)mixed? 8524/3 8550/3 8562/3 8983/3 8470/3 8240/3 8249/3 8246/3 8041/3 8013/3
Size invasive component *on pathology*	200001000004104 |Greatest length dimension of excised primary malignant neoplasm (observable entity)|		*Millimeter*	246115007 |Size (attribute)|	
Topography	372064008 |Malignant neoplasm	*363698007 |Finding site (attribute)|*	C50.0	188147009 |Malignant neoplasm of nipple and areola of female breast (disorder)|	188163001 |Malignant neoplasm of nipple and areola of male breast (disorder)|

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
	of female breast (disorder)|		C50.1	353521000119108 |Primary malignant neoplasm of central portion of female right breast (disorder)|	93884005 |Primary malignant neoplasm of male breast (disorder)| 49058007 |Structure of central portion of breast (body structure)|
			C50.2	373089009 |Primary malignant neoplasm of breast upper inner quadrant (disorder)|	
			C50.3	373090000 |Primary malignant neoplasm of breast lower inner quadrant (disorder)|	
			C50.4	373088001 |Primary malignant neoplasm of breast upper outer quadrant (disorder)|	
			C50.5	373091001 |Primary malignant neoplasm of breast lower outer quadrant (disorder)|	
			C50.6	372092003 |Primary malignant neoplasm of axillary tail of breast (disorder)|	
			C50.8	188157005 |Malignant neoplasm of overlapping sites of breast (disorder)|	
			C50.9	254837009 |Malignant neoplasm of breast (disorder)|	363698007 |Finding site (attribute)| 10003008 |Non-specific (qualifier value)|
Grade of differentiation invasive cancer	373372005 |Histological grade finding (finding)|		Well differentiated (I)	369790002 |Nottingham Combined Grade I: 3-5 points (finding)|	373375007 |Well differentiated histological grade finding (finding)|

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
			Moderately differentiated (II)	369791003 |Nottingham Combined Grade II: 6-7 points (finding)|	373377004 |Moderately differentiated histological grade finding (finding)|
			Poorly differentiated (III)	369792005 |Nottingham Combined Grade III: 8-9 points (finding)|	373373000 |Poorly differentiated histological grade finding (finding)|
			Cell type not determined	384668003 |Nottingham Combined Grade cannot be determined (finding)|	60815008 |Grade not determined (finding)|
Minimal resection margins invasive breast cancer	384960007 |Surgical margin involved by malignant neoplasm in excised tissue specimen (observable entity)|		Free	373067005 |No (qualifier value)|	
			Not free	373066001 |Yes (qualifier value)|	
Progesterone receptor	13892007 |Progesterone receptor assay measurement (procedure)|		Not tested	416237000 |Procedure not done (situation)|	
			Positive	441773004 |Positive measurement finding (finding)|	416561008 |Progesterone receptor positive tumor (disorder)|
			Negative	442225006 |Negative measurement finding (finding)|	441118006 |Progesterone receptor negative neoplasm (disorder)|
			Unknown	261665006 |Unknown (qualifier value)|	
Oestrogen receptor	83302001 |Estrogen receptor assay (procedure)|		Not tested	416237000 |Procedure not done (situation)|	
			Positive	441773004 |Positive measurement finding (finding)|	416053008 |Estrogen receptor positive tumor (disorder)|
			Negative	442225006 |Negative measurement finding (finding)|	441117001 |Estrogen receptor negative neoplasm (disorder)|
			Unknown	261665006 |Unknown (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
HER2 receptor	43934000 |Non-endocrine receptor assay (procedure)|	434363004 |Human epidermal growth factor receptor 2 gene detection by fluorescence in situ hybridization (procedure)|	Not tested	416237000 |Procedure not done (situation)|	
			Positive	441773004 |Positive measurement finding (finding)|	1162602001 |Human epidermal growth factor receptor 2 gene amplification detected (finding)| 427685000 |Human epidermal growth factor 2 positive carcinoma of breast (disorder)|
			Negative	442225006 |Negative measurement finding (finding)|	705105000 |Human epidermal growth factor 2 gene amplification negative (finding)| 431396003 |Human epidermal growth factor 2 negative carcinoma of breast (disorder)|
			Unknown	261665006 |Unknown (qualifier value)|	
Histology in situ component	189336000 |Carcinoma in situ of breast (disorder)|	250537006 |Histopathology finding (finding)|	Ductal carcinoma in situ	1162814007 |Non-infiltrating intraductal carcinoma (morphologic abnormality)|	
			Lobular carcinoma in situ	77284006 |Lobular carcinoma in situ (morphologic abnormality)|	
			Lobular carcinoma in situ pleomorphic	444591006 |Pleomorphic lobular carcinoma in situ (morphologic abnormality)|	
			Paget disease of the nipple	2985005 |Paget's disease, mammary (morphologic abnormality)|	
			Ductal carcinoma in situ, papillary	30566004 |Noninfiltrating intraductal papillary adenocarcinoma (morphologic abnormality)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
			Other	703546002 |Solid papillary carcinoma in situ (morphologic abnormality)| 703545003 |Encapsulated papillary carcinoma (morphologic abnormality)|	Solid papillary 8509/2 Encapsulated papillary 8504/2
Grade of differentiation in situ component	189336000 |Carcinoma in situ of breast (disorder)|	373372005 |Histological grade finding (finding)|	Well differentiated (I)	373375007 |Well differentiated histological grade finding (finding)|	*369781009 |Ductal carcinoma in situ nuclear pleomorphism, grade 1: monotonous nuclei, 1.5 - 2.0 red blood cells diameters, with finely dispersed chromatin and only occasional nucleoli (finding)|*
			Moderately differentiated (II)	373377004 |Moderately differentiated histological grade finding (finding)|	*369782002 |Ductal carcinoma in situ nuclear pleomorphism, grade 2: neither nuclear grade 1 nor nuclear grade 3 (finding)|*
			Poorly differentiated (III)	373373000 |Poorly differentiated histological grade finding (finding)|	*369783007 |Ductal carcinoma in situ nuclear pleomorphism, grade 3: Markedly pleomorphic nuclei, usually >2.5 red blood cells diameters, with coarse chromatin and multiple nucleoli (finding)|*
			Cell type not determined	60815008 |Grade not determined (finding)|	
Size of DCIS at pathology	2710001000004109 |Greatest length dimension of ductal carcinoma in situ in excised neoplasm of		*Millimeter*	246115007 |Size (attribute)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
	breast (observable entity)|				
Minimal resection margins non invasive	396691004 |Surgical margin involved by carcinoma in situ in excised tissue specimen (observable entity)|		Free	373067005 |No (qualifier value)|	
			Not free	373066001 |Yes (qualifier value)|	
pT-stage	384625004 |pT category (observable entity)|	1222589003 |American Joint Committee on Cancer pathological T category allowable value (qualifier value)|	T0 - No evidence of primary tumour	1228951007 |American Joint Committee on Cancer pT0 (qualifier value)|	
			T1a - Tumour > 1 mm but <= 5 mm in greatest dimension	1228959009 |American Joint Committee on Cancer pT1a (qualifier value)|	
			T1b - Tumour > 5 mm but <= 10 mm in greatest dimension	1228962007 |American Joint Committee on Cancer pT1b (qualifier value)|	
			T1c - Tumour > 10 mm but <= 20 mm in greatest dimension	1229846008 |American Joint Committee on Cancer pT1c (qualifier value)|	
			T1mi - Tumour <= 1 mm in greatest dimension	1228958001 |American Joint Committee on Cancer pT1mi (qualifier value)|	
			T1NOS - tumour <= 20 mm in greatest dimension	1228957006 |American Joint Committee on Cancer pT1 (qualifier value)|	
			T2 - Tumour > 20 mm but <= 50 mm in greatest dimension	1229852009 |American Joint Committee on Cancer pT2 (qualifier value)|	
			T3 - Tumour > 50 mm in greatest dimension	1229859000 |American Joint Committee on Cancer pT3 (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
			T4a - Extension to the chest wall, not including only pectoralis muscle adherence/invasion	1229865000 |American Joint Committee on Cancer pT4a (qualifier value)|	
			T4b - Ulceration and/or ipsilateral satellite nodules and/or oedema (including peau d'orange) of the skin which do not meet the criteria for inflammatory carcinoma	1229866004 |American Joint Committee on Cancer pT4b (qualifier value)|	
			T4c - Both T4a and T4b	1229867008 |American Joint Committee on Cancer pT4c (qualifier value)|	
			T4d - Inflammatory carcinoma	1229866004 |American Joint Committee on Cancer pT4b (qualifier value)|	
			T4NOS - Tumour of any size with direct extension to the chest wall and/or to the skin (ulceration or skin nodules)	1229865000 |American Joint Committee on Cancer pT4a (qualifier value)|	
			Tis - Carcinoma in situ	1228953005 |American Joint Committee on Cancer pTis (qualifier value)|	
			Tis (DCIS) - Ductal carcinoma in situ	1228954004 |American Joint Committee on Cancer pTis(DCIS) (qualifier value)|	
			Tis (LCIS) - Lobular carcinoma in situ	1228953005 |American Joint Committee on Cancer pTis (qualifier value)|	109888004 |Lobular carcinoma in situ of breast (disorder)|
			Tis (Paget's) - Paget's disease of the nipple is NOT associated with invasive carcinoma and/or carcinoma in situ (DCIS and/or LCIS) in the underlying breast parenchyma. Carcinomas in the breast parenchyma associated with Paget disease are categorized based on the size and characteristics of the parenchymal disease, although the presence of Paget disease should still be noted.	1228956002 |American Joint Committee on Cancer pTis(Paget) (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
			TX - Primary tumour cannot be assessed	1228950008 |American Joint Committee on Cancer pTX (qualifier value)|	
			Unknown / Not done	261665006 |Unknown (qualifier value)|	
ypT-stage	384625004 |pT category (observable entity)| 395073001 |Cancer treatment started (situation)| 373847000 |Neo-adjuvant - intent (qualifier value)|	1222595002 |American Joint Committee on Cancer ypT category allowable value (qualifier value)|	T0 - No evidence of primary tumour	1228863001 |American Joint Committee on Cancer ypT0 (qualifier value)|	
			T1a - Tumour > 1 mm but <= 5 mm in greatest dimension	1228872009 |American Joint Committee on Cancer ypT1a (qualifier value)|	
			T1b - Tumour > 5 mm but <= 10 mm in greatest dimension	1228897008 |American Joint Committee on Cancer ypT1b (qualifier value)|	
			T1c - Tumour > 10 mm but <= 20 mm in greatest dimension	1228905006 |American Joint Committee on Cancer ypT1c (qualifier value)|	
			T1mi - Tumour <= 1 mm in greatest dimension	1228870001 |American Joint Committee on Cancer ypT1mi (qualifier value)|	
			T1NOS - tumour <= 20 mm in greatest dimension	1228869002 |American Joint Committee on Cancer ypT1 (qualifier value)|	
			T2 - Tumour > 20 mm but <= 50 mm in greatest dimension	1228910005 |American Joint Committee on Cancer ypT2 (qualifier value)|	
			T3 - Tumour > 50 mm in greatest dimension	1228917008 |American Joint Committee on Cancer ypT3 (qualifier value)|	
			T4a - Extension to the chest wall, not including only pectoralis muscle adherence/invasion	1228923003 |American Joint Committee on Cancer ypT4a (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
			T4b - Ulceration and/or ipsilateral satellite nodules and/or oedema (including peau d'orange) of the skin which do not meet the criteria for inflammatory carcinoma	1228924009 |American Joint Committee on Cancer ypT4b (qualifier value)|	
			T4c - Both T4a and T4b	1228925005 |American Joint Committee on Cancer ypT4c (qualifier value)|	
			T4d - Inflammatory carcinoma	1228926006 |American Joint Committee on Cancer ypT4d (qualifier value)|	
			T4NOS - Tumour of any size with direct extension to the chest wall and/or to the skin (ulceration or skin nodules)	1228922008 |American Joint Committee on Cancer ypT4 (qualifier value)|	
			Tis - Carcinoma in situ	1228865008 |American Joint Committee on Cancer ypTis (qualifier value)|	
			Tis (DCIS) - Ductal carcinoma in situ	1228866009 |American Joint Committee on Cancer ypTis(DCIS) (qualifier value)|	
			Tis (LCIS) - Lobular carcinoma in situ	1228866009 |American Joint Committee on Cancer ypTis(DCIS) (qualifier value)|	109888004 |Lobular carcinoma in situ of breast (disorder)|
			Tis (Paget's) - Paget's disease of the nipple is NOT associated with invasive carcinoma and/or carcinoma in situ (DCIS and/or LCIS) in the underlying breast parenchyma. Carcinomas in the breast parenchyma associated with Paget disease are categorized based on the size and characteristics of the parenchymal disease, although the presence of Paget disease should still be noted.	1228868005 |American Joint Committee on Cancer ypTis(Paget) (qualifier value)|	
			TX - Primary tumour cannot be assessed	1228862006 |American Joint Committee on Cancer ypTX (qualifier value)|	
			Unknown / Not done	261665006 |Unknown (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
pN-stage	371494008 |pN category (observable entity)|	1222590007 |American Joint Committee on Cancer pathological N category allowable value (qualifier value)|	N0 - No regional lymph node metastasis identified histologically	1229947003 |American Joint Committee on Cancer pN0 (qualifier value)|	
			N0(i+) - Malignant cells in regional lymph node(s) no greater than 0.2 mm (detected by H&E or IHC including ITC)	1229949000 |American Joint Committee on Cancer pN0(i+) (qualifier value)|	
			N0(mol+) - Positive molecular findings (RT-PCR), but no regional lymph node metastases detected by histology or IHC	1229950000 |American Joint Committee on Cancer pN0(mol+) (qualifier value)|	
			N1 - Micrometastases; or metastases in 1 to 3 axillary lymph nodes; and/or in internal mammary nodes with metastases detected by sentinel lymph node biopsy but not clinically detected	1229951001 |American Joint Committee on Cancer pN1 (qualifier value)|	
			N1a - Metastases in 1 to 3 axillary lymph nodes, at least one metastasis greater than 2.0 mm	1229954009 |American Joint Committee on Cancer pN1a (qualifier value)|	
			N1b - Metastases in internal mammary nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected	1229955005 |American Joint Committee on Cancer pN1b (qualifier value)|	
			N1c - Metastases in 1 to 3 axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometstases detected by sentinel lymph node biopsy but not clinically detected	1229956006 |American Joint Committee on Cancer pN1c (qualifier value)|	
			N1mi - Micrometastases (approximately 200 cells, larger than 0.2 mm, but none larger than 2.0 mm)	1229952008 |American Joint Committee on Cancer pN1mi (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
			N2a - Metastases in 4-9 axillary lymph nodes, including at least one that is larger than 2 mm	1229959004 |American Joint Committee on Cancer pN2a (qualifier value)|	
			N2b - Metastasis in clinically detected internal mammary lymph node(s) in the absence of axillary lymph node metastasis	1229960009 |American Joint Committee on Cancer pN2b (qualifier value)|	
			N2NOS - Metastasis in 4-9 ipsilateral axillary lymph nodes or in clinically detected ipsilateral internal mammary lymph node(s) in the absence of axillary lymph node metastasis	1229957002 |American Joint Committee on Cancer pN2 (qualifier value)|	
			N3a - Metastases in 10 or more axillary lymph nodes (at least one larger than 2 mm) or metastasis in infraclavicular lymph nodes	1229963006 |American Joint Committee on Cancer pN3a (qualifier value)|	
			N3b - Metastases in clinically detected ipsilateral internal mammary lymph nodes in the presence of 1 or more positive axillary lymph nodes; or in more than 3 axillary lymph nodes and in internal mammary lymph nodes with microscopic or macroscopic metastasis detected by sentinel lymph node biopsy but not clinically detected	1229964000 |American Joint Committee on Cancer pN3b (qualifier value)|	
			N3c - Metastases in ipsilateral supraclavicular lymph nodes	1229965004 |American Joint Committee on Cancer pN3c (qualifier value)|	
			N3NOS - Metastases in 10 or more axillary lymph nodes; or in infraclavicular (level III axillary) lymph nodes; or in clinically detected ipsilateral internal mammary lymph nodes	1229962001 |American Joint Committee on Cancer pN3 (qualifier value)|	
			NX - Regional lymph nodes cannot be assessed (e.g., previously removed, or not removed for pathologic study)	1229945006 |American Joint Committee on Cancer pNX (qualifier value)|	
			Unknown / Not done	261665006 |Unknown (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
ypN-stage	371494008 |pN category (observable entity)| 395073001 |Cancer treatment started (situation)|	1222596001 |American Joint Committee on Cancer ypN category allowable value (qualifier value)|	N0 - No regional lymph node metastasis identified histologically	1229878000 |American Joint Committee on Cancer ypN0 (qualifier value)|	
	373847000 |Neo-adjuvant - intent (qualifier value)|		N0(i+) - Malignant cells in regional lymph node(s) no greater than 0.2 mm (detected by H&E or IHC including ITC)	1229881005 |American Joint Committee on Cancer ypN0(i+) (qualifier value)|	
			N0(mol+) - Positive molecular findings (RT-PCR), but no regional lymph node metastases detected by histology or IHC	1229880006 |American Joint Committee on Cancer ypN0(mol+) (qualifier value)|	
			N1 - Micrometastases; or metastases in 1 to 3 axillary lymph nodes; and/or in internal mammary nodes with metastases detected by sentinel lymph node biopsy but not clinically detected	1229884002 |American Joint Committee on Cancer ypN1 (qualifier value)|	
			N1a - Metastases in 1 to 3 axillary lymph nodes, at least one metastasis greater than 2.0 mm	1229887009 |American Joint Committee on Cancer ypN1a (qualifier value)|	
			N1b - Metastases in internal mammary nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected	1229889007 |American Joint Committee on Cancer ypN1b (qualifier value)|	
			N1c - Metastases in 1 to 3 axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometstases detected by sentinel lymph node biopsy but not clinically detected	1229890003 |American Joint Committee on Cancer ypN1c (qualifier value)|	
			N2a - Metastases in 4-9 axillary lymph nodes, including at least one that is larger than 2 mm	1229893001 |American Joint Committee on Cancer ypN2a (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
			N2b - Metastasis in clinically detected internal mammary lymph node(s) in the absence of axillary lymph node metastasis	1229896009 |American Joint Committee on Cancer ypN2b (qualifier value)|	
			N2NOS - Metastasis in 4-9 ipsilateral axillary lymph nodes or in clinically detected ipsilateral internal mammary lymph node(s) in the absence of axillary lymph node metastasis	1229892006 |American Joint Committee on Cancer ypN2 (qualifier value)|	
			N3a - Metastases in 10 or more axillary lymph nodes (at least one larger than 2 mm) or metastasis in infraclavicular lymph nodes	1229898005 |American Joint Committee on Cancer ypN3a (qualifier value)|	
			N3b - Metastases in clinically detected ipsilateral internal mammary lymph nodes in the presence of 1 or more positive axillary lymph nodes; or in more than 3 axillary lymph nodes and in internal mammary lymph nodes with microscopic or macroscopic metastasis detected by sentinel lymph node biopsy but not clinically detected	1229899002 |American Joint Committee on Cancer ypN3b (qualifier value)|	
			N3c - Metastases in ipsilateral supraclavicular lymph nodes	1229900007 |American Joint Committee on Cancer ypN3c (qualifier value)|	
			N3NOS - Metastases in 10 or more axillary lymph nodes; or in infraclavicular (level III axillary) lymph nodes; or in clinically detected ipsilateral internal mammary lymph nodes	1229897000 |American Joint Committee on Cancer ypN3 (qualifier value)|	
			NX - Regional lymph nodes cannot be assessed (e.g., previously removed, or not removed for pathologic study)	1229877005 |American Joint Committee on Cancer ypNX (qualifier value)|	
			Unknown / Not done	261665006 |Unknown (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
pM-stage	371497001 |pM category (observable entity)|	1222591006 |American Joint Committee on Cancer pathological M category allowable value (qualifier value)|	Not applicable - pM cannot be determined from the submitted specimen(s)	260415000 |Not detected (qualifier value)|	
			pM1: Histologically proven metastases larger than 0.2 mm	1229916009 |American Joint Committee on Cancer pM1 (qualifier value)|	
Date of surgery primary tumour	392090004 |Operation on breast (procedure)|	63348002 |Excision of benign tumor of breast (procedure)| 46116005 |Excision of malignant tumor of breast (procedure)|	Date	439272007 |Date of procedure (observable entity)|	
Type of surgery	392090004 |Operation on breast (procedure)|	46116005 |Excision of malignant tumor of breast (procedure)| 63348002 |Excision of benign tumor of breast (procedure)|	Not performed	416237000 |Procedure not done (situation)|	
			Breast conserving surgery	64368001 |Excision of part of breast (procedure)|	
			Mastectomy	384723003 |Radical mastectomy (procedure)|	172043006 |Simple mastectomy (procedure)|
			Performed, unknown type	46116005 |Excision of malignant tumor of breast (procedure)|	69466000 |Unknown procedure (finding)|
			Unknown	261665006 |Unknown (qualifier value)|	
Re-resection	395165008 |Re-excision of breast for clearance of tumor margins (procedure)|		Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
If yes, Date of re-resection	395165008 |Re-excision of breast for clearance of tumor margins (procedure)|		Date	439272007 |Date of procedure (observable entity)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
Breast reconstruction	33496007 |Reconstruction of breast (procedure)|		Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
If yes, Date of reconstructions	33496007 |Reconstruction of breast (procedure)|		Date	439272007 |Date of procedure (observable entity)|	
Axillary lymph node dissection performed	79544006 |Complete axillary lymphadenectomy (procedure)|	234262008 |Excision of axillary lymph node (procedure)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
If yes, Date of axillary lymph node dissection	79544006 |Complete axillary lymphadenectomy (procedure)|	234262008 |Excision of axillary lymph node (procedure)|	Date	439272007 |Date of procedure (observable entity)|	
Neoadjuvant chemotherapy	1259200004 |Neoadjuvant antineoplastic chemotherapy (procedure)|		Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
Date start neoadjuvant chemotherapy	1259200004 |Neoadjuvant antineoplastic chemotherapy (procedure)|		Date	439272007 |Date of procedure (observable entity)| 413946009 |Date treatment started (observable entity)|	
Adjuvant chemotherapy	*367336001 |Chemotherapy (procedure)|*	1259201000 |Adjuvant drug therapy (procedure)|	Yes	373066001 |Yes (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
	373846009 |Adjuvant - intent (qualifier value)|		No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
Date start adjuvant chemotherapy	*367336001 |Chemotherapy (procedure)|*		Date	439272007 |Date of procedure (observable entity)| 413946009 |Date treatment started (observable entity)|	
Neoadjuvant endocrine therapy	169413002 |Hormone therapy (procedure)|	372688009 |Antineoplastic agent (substance)|	Yes	373066001 |Yes (qualifier value)|	
	373847000 |Neo-adjuvant - intent (qualifier value)|		No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
Date start neoadjuvant endocrine therapy	169413002 |Hormone therapy (procedure)|	372688009 |Antineoplastic agent (substance)| 373847000 |Neo-adjuvant - intent (qualifier value)|	Date	439272007 |Date of procedure (observable entity)| 413946009 |Date treatment started (observable entity)|	
Adjuvant endocrine therapy	169413002 |Hormone therapy (procedure)|	372688009 |Antineoplastic agent (substance)| 1259201000 |Adjuvant drug therapy (procedure)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
Date start adjuvant endocrine therapy	169413002 |Hormone therapy (procedure)|	372688009 |Antineoplastic agent (substance)| 1259201000 |Adjuvant drug therapy (procedure)|	Date	439272007 |Date of procedure (observable entity)| 413946009 |Date treatment started (observable entity)|	
Neoadjuvant anti HER2 therapy	1237262009 |Receptor tyrosine-protein kinase erbB-2 inhibitor therapy (procedure)|	784176007 |Substance with human epidermal growth factor receptor 2 inhibitor mechanism of action (substance)| 373847000 |Neo-adjuvant - intent (qualifier value)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
Date start neoadjuvant anti HER2 therapy	1237262009 |Receptor tyrosine-protein kinase erbB-2 inhibitor therapy (procedure)|	784176007 |Substance with human epidermal growth factor receptor 2 inhibitor mechanism of action (substance)| 373847000 |Neo-adjuvant - intent (qualifier value)|	Date	439272007 |Date of procedure (observable entity)| 413946009 |Date treatment started (observable entity)|	
Adjuvant anti HER2 therapy	1237262009 |Receptor tyrosine-protein kinase erbB-2 inhibitor therapy (procedure)|	1259201000 |Adjuvant drug therapy (procedure)| 784176007 |Substance with human epidermal growth factor receptor 2 inhibitor mechanism of action (substance)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
Date start adjuvant anti HER2 therapy	1237262009 |Receptor tyrosine-protein kinase erbB-2 inhibitor therapy (procedure)|	1259201000 |Adjuvant drug therapy (procedure)| 784176007 |Substance with human epidermal growth factor receptor 2 inhibitor mechanism of action (substance)|	Date	439272007 |Date of procedure (observable entity)| 413946009 |Date treatment started (observable entity)|	
Neoadjuvant radiotherapy	108290001 |Radiation oncology AND/OR radiotherapy (procedure)|	373847000 |Neo-adjuvant - intent (qualifier value)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
Start neoadjuvant radiotherapy	108290001 |Radiation oncology AND/OR radiotherapy (procedure)|	373847000 |Neo-adjuvant - intent (qualifier value)|	Date	439272007 |Date of procedure (observable entity)| 413946009 |Date treatment started (observable entity)|	
Adjuvant radiotherapy	108290001 |Radiation oncology AND/OR radiotherapy (procedure)|	373846009 |Adjuvant - intent (qualifier value)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
Target volume: Chest wall	428624002 |Radiotherapy to chest wall (procedure)|		Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
Target volume: Interpectoral nodes	168522007 |Radiotherapy for lymphatic irradiation (procedure)|		Yes	373066001 |Yes (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
	245274001 |Interpectoral lymph node group (body structure)|		No	373067005 |No (qualifier value)|	
Target volume: Internal mammary nodes	168522007 |Radiotherapy for lymphatic irradiation (procedure)|		Yes	373066001 |Yes (qualifier value)|	
	245282001 |Internal mammary lymph node group (body structure)|		No	373067005 |No (qualifier value)|	
Total dose prescribed	399077005 |Prescribed external beam radiation therapy dose (observable entity)|		*Gy*	229029004 |Gray (qualifier value)|	
Number of fractions to elective regions	445232009 |Boost radiation therapy (procedure)| 410673009 |Region (attribute)| 103390000 |Elective (qualifier value)|	228862004 |Number of fractions (observable entity)|	*Number*	260299005 |Number (qualifier value)|	
If boost, Fraction dose boost	445565002 |Total boost radiation dose delivered (observable entity)|		*Gy*	229029004 |Gray (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
If boost, Number of fractions to boost regions	445232009 |Boost radiation therapy (procedure)|	228862004 |Number of fractions (observable entity)|	*Number*	260299005 |Number (qualifier value)|	
Dermatitis	49084001 |Dermatitis caused by radiation (disorder)|	108290001 |Radiation oncology AND/OR radiotherapy (procedure)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
Lymphedema	439128001 |Lymphedema caused by radiation (disorder)|	108290001 |Radiation oncology AND/OR radiotherapy (procedure)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
Rib fracture	704169000 |Pathological fracture of rib (disorder)|	109301001 |Radiation injury of bone (disorder)|	Yes	373066001 |Yes (qualifier value)|	
		108290001 |Radiation oncology AND/OR radiotherapy (procedure)|	No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
Heart toxicity	430401005 |Heart disease caused by ionizing radiation (disorder)|	108290001 |Radiation oncology AND/OR radiotherapy (procedure)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
Radiation Pneumonitis	84004001 |Radiation pneumonitis (disorder)|	108290001 |Radiation oncology AND/OR radiotherapy (procedure)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
Hypertension	38341003 |Hypertensive disorder, systemic arterial (disorder)|		Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
Smoking at the time of diagnosis	77176002 |Smoker (finding)|		Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
Diabetes	73211009 |Diabetes mellitus (disorder)|		Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
			Unknown	261665006 |Unknown (qualifier value)|	
BMI	60621009 |Body mass index (observable entity)|	301331008 |Finding of body mass index (finding)|	*Number*	410680006 |Number (attribute)|	
Date of last follow up	185353001 |Appointment date (finding)|	308273005 |Follow-up status (finding)|	Date	410671006 |Date (attribute)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
Status at last follow up	308273005 |Follow-up status (finding)|		Alive with no evidence of disease	438949009 |Alive (finding)| 51231003 |Disease condition determination, cured (finding)|	
			Alive with disease, index tumour successfully salvaged	438949009 |Alive (finding)|	
			Alive with disease, index tumour not salvaged	438949009 |Alive (finding)|	
			Died from intercurrent disease	419099009 |Dead (finding)|	184305005 |Cause of death (observable entity)| (other disorder than 372064008 |Malignant neoplasm of female breast (disorder)|)
			Died from index tumour	419099009 |Dead (finding)|	184305005 |Cause of death (observable entity)| 372064008 |Malignant neoplasm of female breast (disorder)|
			Died from toxicity	419099009 |Dead (finding)|	184305005 |Cause of death (observable entity)| 373866000 |Death due to chemotherapy toxicity (event)|
			Died from unknown cause	419099009 |Dead (finding)|	184305005 |Cause of death (observable entity)| 261665006 |Unknown (qualifier value)|
Local recurrence	314955001 |Local recurrence of malignant tumor of breast (disorder)|		Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
If YES, date of local recurrence	314955001 |Local recurrence of malignant tumor of breast (disorder)|		Date	410671006 |Date (attribute)|	
Regional recurrence	25173007 |Recurrent neoplasm (disorder)|	25173007 |Recurrent tumor (finding)| 312500006 |Regional lymph node structure (body structure)| 94392001 |Secondary malignant neoplasm of lymph node (disorder)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
If YES, date of regional recurrence	25173007 |Recurrent neoplasm (disorder)|	25173007 |Recurrent tumor (finding)| 312500006 |Regional lymph node structure (body structure)| 94392001 |Secondary malignant neoplasm of lymph node (disorder)|	Date	410671006 |Date (attribute)|	
Distant metastases	128462008 |Metastatic malignant neoplasm (disorder)|	399409002 |Distant metastasis present (finding)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
If YES, date of distant metastases	128462008 |Metastatic malignant neoplasm (disorder)|	399409002 |Distant metastasis present (finding)|	Date	410671006 |Date (attribute)|	

Data Element	Variable SNOMED code + FSN	Variable alternative SNOMED codes	Permissible values	Values SNOMED code + FSN	Values alternative (additional) SNOMED code + FSN
Postoperative multi-disciplinary team meeting	60631000210103 |Breast cancer multidisciplinary meeting (procedure)|	262061000 |Postoperative period (qualifier value)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
Preoperative multi-disciplinary team meeting	60631000210103 |Breast cancer multidisciplinary meeting (procedure)|	262068006 |Preoperative (qualifier value)|	Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
Was cancer found in screening?	268547008 |Screening for malignant neoplasm of breast (procedure)|	243878006 |Breast neoplasm screening status (finding)|	Yes	373066001 |Yes (qualifier value)|	185710003 |Breast screening abnormal - told patient (finding)|
			No	373067005 |No (qualifier value)|	185713001 |Breast screening non-attender (finding)|
Is patient receiving palliative treatment?	362964009 |Palliative procedure (procedure)|		Yes	373066001 |Yes (qualifier value)|	
			No	373067005 |No (qualifier value)|	
Is patient enrolled in registred trial?	709491003 |Enrollment in clinical trial (procedure)|		Yes	373066001 |Yes (qualifier value)|	428024001 |Clinical trial participant (person)|
			No	373067005 |No (qualifier value)|	

10.6. SMART risk score

```
############################
#SMART2 risk score function
############################

#Last update: 27 - 01 - 2022
smart2 <- function(age, sex, isSmoking, sbp, diabetesDiagnosis, coronaryArteryDisease,
cerebrovascularDisease, peripheralArteryDisease, aorticAneurysm, yearsSinceFirstDiagnosis,
hdlCholesterol, totalCholesterol, creatinin, crp, hrFutureEvent = 1, usingAnticoag, region){

LP <- NULL
tenYearRisk <- NULL

 sex <- ifelse(sex==1, "M", "F")

  #Bepaal recalibratiefactor voor geografische regio
  ratioRegion <- ifelse(region[18]=="low", 0.81590,
          ifelse(region=="mod", 0.6973285,
          ifelse(region=="high", 0.5085825,
          ifelse(region=="vhigh", 0.2285371,
          ifelse(region=="Asia", 0.4043255,
          ifelse(region=="Australia", 1.0040808,
          ifelse(region=="Japan", 0.8825678,
          ifelse(region=="North America", 0.4714961,
          ifelse(region=="Latin America", 0.5075670, NA)))))))))

  #Berekenen eGFR obv CKDepi
  gfr <- ifelse(sex=="M", 32788*creatinin^(-1.154)*age^(-0.203),
32788*creatinin^(-1.154)*age^(-0.203)*0.742)

  #Berekenen non-HDL obv totaal cholesterol en HDL
  nonHdl <- totalCholesterol - hdlCholesterol

  LP= -0.0349602236 * age +
          0.0005510715 * age^2 +
          ifelse(sex=="M", 0.2876587433, 0) +
          0.3455832714 * isSmoking +
          0.0018913154 * sbp +
          0.3181706587 * diabetesDiagnosis +
          0.2947019539 * coronaryArteryDisease +
          0.3483178604 * cerebrovascularDisease +
          0.3303566308 * aorticAneurysm +
```

[18] As defined in European Risk Regions (heartscore.org) - Netherland: Low; Austria: moderate; Estonia: High

```
        0.2244665798 * peripheralArteryDisease +
        0.0476995851 * yearsSinceFirstDiagnosis -
        0.0016497342 * yearsSinceFirstDiagnosis^2 +
        0.5403642493 * log(nonHdl) -
        0.0396752081 * gfr +
        0.0002186126 * gfr^2 +
        0.1517601731 * log(crp) -
        0.2107210313 * usingAnticoag

 tenYearRisk <- 1 - (1-0.1658228)^exp(LP + 0.0463729 - log(ratioRegion))
 tenYearRisk <- ifelse(tenYearRisk==1,0.9999999,tenYearRisk)

 tenYearRiskWithTreatment <- 1 - (1-tenYearRisk)^(hrFutureEvent) #hrFutureEvent komt uit
behandeleffectenscript

 return(
        list(
        currentTenYearRisk = tenYearRisk,
        tenYearRiskWithTreatment = tenYearRiskWithTreatment
        )
 )
}
```

11. Regulatory, Ethical, and Study Oversight Considerations

11.1 Regulatory and Ethical Considerations

From an ethical point of view, the AIDAVA-curated data will NOT be presented to clinicians and patients as being reliable for final use, but as a research prototype output to be validated before use (if they choose to) in the care process or clinical research activities.
The partners intend, provided that this project demonstrates success, to obtain Medical Device Regulation certification so that the outputs can be accepted and used. This will only occur after the project.

- This study will be conducted in accordance with the protocol and with the following:
  - Consensus ethical principles derived from international guidelines including the Declaration of Helsinki and Council for International Organizations of Medical Sciences (CIOMS) international ethical guidelines
  - Applicable ICH Good Clinical Practice (GCP) guidelines
  - Applicable laws and regulations
- The protocol, protocol amendments, ICF, SIP will be submitted to an IRB/IEC by the investigator and reviewed and approved by the IRB/EC active in each of the study centres, before the study is initiated.
- Any amendments to the protocol will require IRB/EC approval before implementation of changes made to the study design, except for changes necessary to eliminate an immediate hazard to study participants.

11.2 Financial Disclosure: not applicable

11.3 Informed Consent Process

- The investigator or the research associate will explain the nature of the study, including the risks and benefits, to the potential participant and answer all questions regarding the study.
- Potential participants must be informed that their participation is voluntary. They will be required to sign a statement of informed consent that meets the requirements of local regulations, ICH guidelines, privacy and data protection requirements, where applicable, and the IRB/IEC or study centre.
- The written informed consent must be obtained before the participant is enrolled in the study and the date the written consent was obtained. The authorised person obtaining the informed consent must also sign the ICF.
- Participants must consent to the most current version of the ICF(s) during their participation in the study.
- A copy of the ICF(s) must be provided to the participant and kept on record at the recruiting site.

11.4 Data Protection

The participant must be informed that their personal study-related data will be used by the project in accordance with local data protection law. The level of disclosure must also be explained to the participant who will be required to give consent for their data to be used as described in the informed consent.

Information technology systems used to collect, process, and store study-related data are secured by technical and organisational security measures designed to protect such data against accidental or unlawful loss, alteration, or unauthorised disclosure or access.

Additional information related to data processing and security aspects can be found in Annex 4 on Data Sharing Agreement.

References

[1] "[No title]." https://www.fda.gov/media/108378/download#:~:text=The%20process%20of%20registering%20or,of%20answering%20the%20trial%20questions. (accessed May 18, 2023).

[2] "- glossary -." https://glossary.ramit.be/public/term.cfm?termID=1330 (accessed May 17, 2023).

[3] "enrollment or enrolment in clinical trials - Google Search." https://www.google.com/search?q=enrollment+or+enrolment+in+clinical+trials&sxsrf=APwXEdd3FGTz6TT4Cb16H0IBP7rVb4YaZA%3A1684312235445&ei=q5BkZNrtGt6lkdUPwZGV6Ag&ved=0ahUKEwia7--D-Pv-AhXeUqQEHcFIBY0Q4dUDCA8&uact=5&oq=enrollment+or+enrolment+in+clinical+trials&gs_lcp=Cgxnd3Mtd2l6LXNlcnAQAzIICCEQFhAeEB0yCAghEBYQHhAdOgoIABBHENYEELADOgoIABCKBRCwAxBDOgUIABCABDoICAAQigUQhgM6BggAEBYQHjoFCCEQoAE6BwghEKABEAo6BQghEJIDOgQIIRAVSgQIQRgAUJ0KWMMcYL4eaAJwAXgAgAF9iAGDDpIBBDE0LjWYAQCgAQHIAQrAAQE&sclient=gws-wiz-serp (accessed May 17, 2023).

[4] "ESC Prevention of CVD Programme: SMART Risk Score." https://www.escardio.org/Education/ESC-Prevention-of-CVD-Programme/Risk-assessment/SMART-Risk-Score (accessed May 17, 2023).

[5] "U-prevent." https://u-prevent.com/ (accessed Jul. 31, 2023).

[6] J. A. N. Dorresteijn *et al.*, "Development and validation of a prediction rule for recurrent vascular events based on a cohort study of patients with arterial disease: the SMART risk score," *Heart*, vol. 99, no. 12, pp. 866–872, Jun. 2013.

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

Deliverable No. D1.4
Annex 2 - Templates for Study Information Package (SIP) and Informed Consent Form (ICF)

Contractual Submission Date: 30/11/2023

Actual Submission Date: 30/11/2023

Responsible partner: P8- NMEC

Funded by
the European Union

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

Deliverable number	**Annex 2 to Deliverable D1.4**
Deliverable title	**D1.4. Description of Assessment Study (and supporting material)**
Type[1]	R
Dissemination level[2]	SEN
Work package number	WP1
Work package leader	P8 - NEMC and P2 - b!loba
Author(s)	Isabelle de Zegher (P2-b!loba) Nathan Lea, Maria Christofidou (P5-IHD)
Reviewers	Mall Maasik, Kerli Norak, Katrin Lepik, Eno-Martin Lotman (P6-NEMC) Isabella Cina (P12-EHN), Laura Suacedo-Cuevas (P11- ECPC) Dominik Steiger (MID) Hanne Muller (DME)
Keywords	

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).

The work of MID received funding by the Swiss State Secretariat for Education, Research and Innovation (SBFI), subvention contract 22.00093, REF-1131-52104.

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

Document History

Version	Date	Description
V1	March 2023	First version by IHD based on a a generic template
V2	April 2023	Adaptation to AIDAVA
V3	August 2023	Updates following comments from patient consultants
V4	November 2023	Final version ready

[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 Abbreviations and definitions

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

Foreword

This document includes a template of the Study Information Package (SIP), together with the Informed Consent Form (ICF) to be used for the Assessment Study Protocol of AIDAVA for the 2 identified use cases. They are based on a broad template that incorporates the particulars for an informed consent form that Partner i~HD have encountered for multi-site projects across Europe; the first version of this SIP has been reviewed by the patient consultants from the 2 patients associations, EHN and ECPC, supporting the project, based on a structured questionnaire.

The proposed template includes details of AIDAVA and should be provided to potential participants so that they can make an informed decision about whether they wish to participate.

These forms are expected to be adapted, translated into the local languages for participants and be submitted for approval by the local Ethics Committee, together with the Study Protocol. It is advised that each site checks if this template is sufficient to meet local requirements around the specification and consent arrangements in each jurisdiction, and completes it if needed. Also, if a site decides to have a different version - based on this template - for BC and CVD patients, this is possible as long as there is consistency with the template.

Information for patients on the assessment of the AIDAVA prototype

Thank you for considering participation in the assessment study of the AIDAVA prototype, an intelligent virtual assistant that aims to support you as a patient to manage, clean and increase the quality and reusability of your personal health data. Your participation in this study is important: it will help us understand how well the solution works and whether it can be used in daily life for patients.

You have been screened as a potential participant to the study as your treating physician assessed that you meet the eligibility criteria described below.

- Your health record is available within the medical centre.
- You confirmed interest in being in more control of your personal health data.
- You confirmed interest in increasing the quality of these data by working with tools like the AIDAVA..
- You own a smartphone and/or a tablet and you are comfortable to work with digital apps.
- You have a confirmed diagnosis of Breast Cancer
  OR you have a confirmed diagnosis of symptomatic type 1 acute ST-Elevation Myocardial Infarction.

Before you decide to participate, it is important for you to understand why the research is being done and what it will involve. Key points, further detailed in this document, are summarised below.

- The prototype is fully funded by a grant from the European Commission (Horizon Europe).
- You will be working with the medical centre who selected you, and with their partner Health Data Intermediary managing non-hospital data (GP, quality of life data, medical device if applicable).
- When participating in the study, you will be introduced and trained to the system and supported by a local research associate. After signing the Informed Consent Form, you will be able to import your data and start the cleaning process; during this cleaning process, the system might ask you a few clarification questions. If you are not able to answer a question raised by the system you have the opportunity to forward the question to an expert curator appointed to you in the project. When the system is ready, you will be requested to fill in a questionnaire, helping to assess the performance of the system. All interactions will happen in your own language.
- The main benefit for you to participate in the study is to help create and have access to your - clean - health record, and to your International Patient Summary (IPS); you can show your IPS to any physician you are visiting, to assess your health condition. You will also contribute to a project that might be of major value for patients, clinical practitioners and researchers. There is no risk associated with the study. The system has no effect on your treatment; one issue for you might be the time needed for the assessment (about 16 hours over 2 weeks, twice over a period of 18 months).
- Your data will be managed securely in full compliance with data protection regulations.
    - You will have the possibility to withdraw from the study at any point in time and your data will be deleted from the prototype unless you express your interest to leave the data for evaluation stated in the informed consent form (not from the source system).
    - Equally, your data will be deleted at the end of the study.
    - No original data in the source databases will be changed due to any activity within the AIDAVA system.

Please read the following information carefully and ask the study team any questions you may have.
We are excited to have you on board for this groundbreaking initiative; your role, as patient in control of your health data, is extremely important to help assess acceptability and effectiveness of the prototype.

Introduction to AIDAVA

What is AIDAVA and what are its aims?

Integrated, high-quality personal health data represents a potential wealth of knowledge for health care systems in general, and concretely to support your treating physician. Today there is no reliable conduit for this data to become easily reusable.

AIDAVA will fill this gap by prototyping and testing an intelligent virtual assistant that aims to support you, as a patient, in the cleaning and reuse of all the personal health data that you have asked to import into the system. **Cleaning** - also referred technically as "curation" - means that your data will be integrated across different data sources, transformed into a format that supports easy reuse and checked for consistency and completeness; **reuse** means that your data can be provided in a format required for a specific use such as reporting on your case in a format for discussion across clinical teams, contribution to clinical research studies, and generation of your International Personal Summary (IPS)[3] for consultation by you and any physician treating you in your medical centre and anywhere in Europe.

More precisely the AIDAVA prototype is designed to support you at different levels:

1. Import your data, following your confirmation; this includes hospital data as well as data managed through your Health Data Intermediary including data from your Blood Pressure Monitoring device for Cardiovascular patients, apps, paper letter from your general practitioner - or other relevant document - that you scanned.
2. Automatically clean this data into the reusable format; in case the system requires clarification it may ask you questions, or decide to ask the more technical questions to a data expert from the clinical centre.
3. Extract from your cleaned data a summary of your health status, in an international standard format enforced across all European countries, called International Personal Summary (IPS)
4. Transfer your personal health summary to the organisation - called **Health Data Intermediary** - that helps you add, control and visualise your personal data.
5. For cardiovascular patients: Transform your data to compute a risk score that will be examined by your treating physician.
   For breast cancer patients: Transform your data and integrate them in a data set that could be used for clinical research on Breast Cancer across the medical centres participating in the project across Europe.

Who is funding AIDAVA?

AIDAVA is a Research and Innovation Actions (RIA), funded by the European Commission from the Horizon Europe program under Grant Agreement no 101057062. A introduction to the project in English can be found at the project homepage. See website (www.AIDAVA.eu).

[3] The International Personal Summary (IPS) is a standard introduced by the health authorities across Europe (and across US and several Asiatic countries) to support exchange of the most critical health information on a patient.

Who is responsible for AIDAVA and who else is involved?

The project is coordinated by the University of Maastricht, in The Netherlands and involves 14 participants from 9 European countries. It started in September 2022 and will end in August 2026.

The consortium includes 3 medical centres (University of Maastricht, Graz University, Northern Estonian Medical Center); you have been selected by one of them. Each medical centre will manage your hospital health record and will be responsible for the evaluation of the AIDAVA prototype with you; it will work with a Health Data Intermediary (MIDATA for Graz university and Northern Estonian Medical Center, and DIGI.me for the University of Maastricht). **Health Data Intermediaries** are emerging organisations proposing to help individuals to be in control of their personal health data; they need to comply with regulations from the European Commission (Data Governance Act).

The consortium also includes 2 patient associations (ECPC for Breast Cancer, EHN for Cardiovascular diseases); together with selected representative patients they are advising the consortium on the patient needs and requirements. They have reviewed the material discussed with you pre-test the prototype before you.

In addition, one member of the consortium (The European Institute for Innovation through Health Data) is a non-profit organisation specialised in Data Privacy and Data Quality. There are also four technology companies and one university collaboration in the development of the prototype. The whole consortium is supported by one project management organisation (Eurice).

Participation

What will participation involve?

To assess the performance of the AIDAVA prototype we need to perform a formal evaluation study with patients. This is a prospective study, across the 3 medical centres identified above, with 15 breast cancer patients and 15 cardiovascular patients, in each site. All patients will be requested to test the prototype during the same period of time, in two periods: Summer 2024 for the first release of the prototype and Spring 2026 for the second release.

If you agree to participate in this study, you will be asked to do the following.:

1. Read thoughtfully and sign the Informed Consent form.
2. Read thoughtfully and sign a collaboration agreement with the Health Data Intermediary (HDI) that will support integration of your personal data (GP data, quality of life data and potentially medical devices data) under your control. It will also help you navigate through your International Patient Summary through their simple and intuitive application. This is further explained in the collaboration agreement.
3. Use a personnel app (BrightFish[4]) to collect quality of life data.
4. For cardiovascular patients only, use a medical device (Withings BPM[5]) to collect your blood pressure.
5. Use the 2 versions of the AIDAVA prototype as directed by the study team. The first version of the prototype will be used to identify needed areas of improvements that will be delivered in the second version of the prototype.

[4] https://brightfish.com/solutions/ehealth/
[5] https://www.withings.com/it/en/bpm-connect

6. Attend 2 on-site and 2 online meetings for the first version of the prototype.
    - V0 on site: screening, introduction to the study, signature of ICF and HDI agreement, possibility to ask questions and get answers from the project team.
    - V1 on site: install the different applications - including the blood pressure monitoring device for cardiovascular patients - on your smartphone or tablet and explain how to use these applications, fill in a questionnaire that will help the system to define your health and digital literacy level that will be used to direct questions to you or to your delegated curator.
    - V2 online: assess your progress and answer any questions you may have.
    - V3 online: answer a questionnaire to collect your experience with the first version of prototype

    You may make some amendments to the visit format (online or on site) or combine visits (E.g 1 and 2) according to your preference by discussing your personal preferences with the project team.

7. Receive information on a regular basis during the 17 month period during which we improve the prototype to deliver the second version; optional participation to information sessions.
8. Attend 1 on-site and 2 online meetings for the second version of the prototype.
    - V1 on site: install the new version of the different applications - including the blood pressure monitoring device for cardiovascular patients - on your smartphone or tablet and refresh the training on how to use them.
    - V2 online: assess your progress and answer any questions you may have on the prototype.
    - V3 online: answer the same questionnaire to collect your experience with this second version of the prototype.
9. Participate in an optional online meeting at the end of the study to provide feedback on the results for the interested participants.

Your participation in this study is expected to last for 2 weeks in 2 periods

- For the first version of the prototype: July - September 2024
- For the second - improved - version of the prototype: February - March 2026

We estimated that your participation in the complete study will take an average of 16 hours (minimum 11 hours - maximum 22 hours) over 2 weeks for the first version of the prototype, and about 12 hours for the second version. And we deeply appreciate the time you will spend on assessing this prototype; it is of critical important that we can assess how useful and how usable the prototype is for patients

Will this affect my care or participation in any other research?

No, participation in this study will not affect your care, nor will it affect participation in any other research study. The medical centre may still invite you for other research studies outside of AIDAVA.

Benefits and issues

Will I directly benefit from participating in AIDAVA and if so, how?

The main benefit for you to participate in the study is that if the prototype is successful you will have access to your International Patient Summary (IPS)[6] which is an important basis for any physician to assess your health condition. The IPS is particularly valuable in emergency health care situations and also useful when requesting a second opinion with another health care provider anywhere in Europe. You will be able to retain this IPS after the study, in an electronic format of your choice (potentially within the Health Data Intermediary if you decide to keep working with the one you used during the study).
In addition you will be offered the possibility to keep your entire[7] clean and reusable health record into an electronic form of your choice.

For cardiovascular (CVD) patients only. It will also help your treating physician to smoothly compute a CVD risk score that will support your health care provider in making decisions about your health care management and monitoring. Though this is currently possible in the existing health care system, it is very time consuming for your health care providers.

In addition you will contribute to an existing, innovative European project that might be of major value for the overall health system´s data management: as personal health data are of higher quality and reusable - with the agreement of the patient - treating physicians will be able to make decisions more effectively and clinical researchers will have more potential for discovery and testing of new treatments.

You will not be paid for participating in this study. However, you will be reimbursed for any study-related expenses.

Potential issues during the study?

There is no discomfort expected from using the AIDAVA prototype and related apps. The only discomfort we can envisage is the time we are expecting you to spend while testing the prototype, for each of the two versions of the prototype, as detailed above.

To minimise any risk related to data privacy when handling your personal data, the prototype includes critical functions such as access control, secure storage and traceability; and the AIDAVA team has been putting in place a clear process to respect ethical and data privacy constraints during the evaluation study. In addition, the whole assessment study has been approved by the local Ethics Committee (date and approval number= TBD in each site) .

[6] As described above, the International Personal Summary (IPS) is a standard introduced by the health authorities across Europe (and across US and several Asiatic countries) to support exchange of the most critical health information on a patient.
[7] Based on all the data available in the medical centre that selected you, complemented by data provided by the Health Data Intermediary under your directions

Information in scope and information management

What kind of information will be used in AIDAVA and how will it be handled?

AIDAVA will handle information contained in your medical record, available within the medical centre as well as quality of life information - and for the cardiovascular (CVD) patients, parameters extracted from the blood pressure medical device - that are managed through your Health Data Intermediary. In addition, and whenever possible, we will include as well data available from the record of your general practitioner. That information will be imported in AIDAVA, only with your express agreement at the beginning of the cleaning process.

If you decide to withdraw from the study, and at the end of the study, your data will be completely deleted from the AIDAVA prototype. And you will be asked if you do want to keep your data in the cleaned, reusable format (called the Personal Health Knowledge Graph), as well as your International Personal Summary (IPS referred above).

Your data will not leave Europe.

We take data privacy very seriously. All information collected about you during this study will be kept confidential with strict control on who has access to these data, as defined in the study protocol of AIDAVA. When your data is used for analysis, your personal identification information will be removed before it is published; pseudonymisation and anonymization techniques will be applied whenever needed

Will data about me be used by other parties outside of AIDAVA?

Your data will be used only by the AIDAVA research team and staff explicitly mentioned in the study protocol approved by the local Ethics Committee, including the Health Data Intermediary mentioned earlier.

Withdrawal and complaints

What if I do not want to participate anymore?

Your participation in this study is voluntary. You may choose not to participate or you may withdraw from the study at any time without giving any reasons. Withdrawal will not affect your care or treatment you receive in any way. The personal health data collected by you/for you will no longer be used for the purposes of AIDAVA's research, but will need to be retained for legal compliance reasons. In addition - if you explicitly consent to this - we will keep the data provided through the evaluation questionnaires.

What if I have questions or complaints?

If you have any questions about this study, please contact first the local study team and if needed your clinical team

- Local study team: *[name, email, phone]*
- Clinic contact details/ treating team: remain the same [*name, email, phone]*
- Local Study Sponsor: *[name, email, phone]*

If you have any complaints about data protection matters, please contact the local Data Protection Officer

- Local Data Protection Officer : *[name, email, phone]*

Note: when this document has been adapted/translated to the local needs, please also refer to the privacy policy of their recruitment centre.

In addition, if you have any reason for dissatisfaction, you can always refer to the Supervisory Authority of your country

- National Supervisory Authority : *[name, email, phone]*

We hope that this information has been helpful and that it has answered any questions you may have had. Thank you again for considering participation in our software prototype assessment study.

Informed Consent for participation in the evaluation study of the AIDAVA prototype

I, <<PARTICIPANT Name>> of <<PARTICIPANT ADDRESS>> declare that:

The check boxes below are mandatory; **you have to agree on all of them to participate in the study**.

- ☐ I have been provided, read and fully understood the Study Information Pack for participating in AIDAVA.
- ☐ I have had an opportunity to ask questions and they have been answered fully to my satisfaction.
- ☐ I have had an opportunity to discuss and have discussed participation with my medical professional.
- ☐ I understand that my participation is voluntary, that I can withdraw at any time without providing any reason, and that this withdrawal will not affect my care in any way both now and in the future.
- ☐ I have read the above information and have had the opportunity to ask questions. I understand the purpose, procedures, duration, risks and discomforts, benefits, and confidentiality of this study.
- ☐ I freely give my consent to participate in this study.
- ☐ I understand that participation will require me to enter into an agreement with the Health Data Intermediary (specify name Digi.me or MIDATA) and register for their services.
- ☐ I understand that my data might be accessed by the software developer (GND) for system correction

The check boxes below are optional; **even if you do not agree you will be able to participate in the study.**

- ☐ In case of withdrawal, I consent that my data related to evaluation are used - with the same data security safeguards as for my health data - up to the final evaluation at the end of the project.
- ☐ I consent to being contacted about my involvement in the AIDAVA project, for communication purposes.

Participant	**Name**	**Pseudonym**	**Signature**	**Date**
	----------	----------	----------	---/----/2024
Research Officer	**Name**		**Signature**	**Date**
	----------		----------	---/----/2024

Time sheet (to be given to patient for each generation)

Site name	○ NEMC	○ UM	○ MUG
Patient ID / pseudonym	_____ (example NEMCBCC1 , NEMCCVC2, …)		
Period of time of evaluation	START: _____	END: _____	

	Date	Activity	Time **curation** (download, curation)		Time **other** (visit, training, visualisation ..)		Comments from the patient
			Hours	Min	Hours	Min	
WEEK 1							
Monday							
Tuesday							
Wednesday							
Thursday							
Friday							
Saturday							
Sunday							
WEEK 2							
Monday							
Tuesday							
Wednesday							
Thursday							
Friday							
Saturday							
Sunday							
END OF EVALUATION							

ANNEXES

Annex 1. MIDATA Collaboration Agreement

Introduction to MIDATA

MIDATA Cooperative, a non-profit institution situated in Zurich, Switzerland, acts as a Health Data Intermediary. It offers personal, encrypted health data accounts. Account holders of MIDATA accounts can give explicit consent to data sharing with third parties and secondary use of their data by third parties. MIDATA acts on their behalf, guaranteeing the informational self determination of its account holders and upholding their sovereignty (see Articles 4 and 5 of the Articles of Association of the cooperative, https://www.midata.coop/wp-content/uploads/2019/08/MIDATA_Statuten_20190626_EN.pdf).

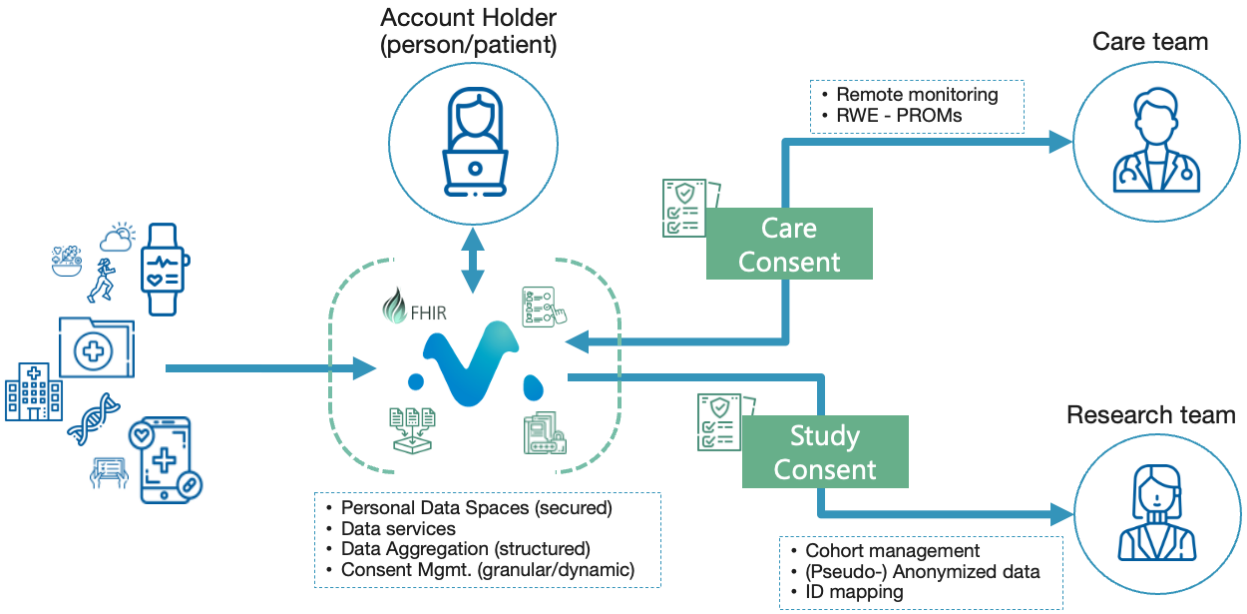


Figure 1: MIDATA platform

For MIDATA accounts, the following General Terms and Conditions and Privacy Policy apply:
General Terms and Conditions: https://ch.midata.coop/#/public/terms?which=midata-terms-of-use
Privacy Policy: https://ch.midata.coop/#/public/terms?which=midata-privacy-policy

Terms of Use: Participation in AIDAVA

These Terms of Use concern your use of the MIDATA platform, as part of your participation in the AIDAVA study [cite study designation/number ethics approval] [link to SIP/ICF].
At account creation, you - as a patient - will be displayed and prompted to accept these Terms of Use specific for the AIDAVA study. By accepting these terms, you set the electronic consents for the interaction of your MIDATA health data account with the AIDAVA prototype.

During your participation in the AIDAVA study, you will use the MIDATA platform to collect different types of personal health data:

- GP data

- Quality of life questionnaire via connected app
- Withings Blood Pressure Monitor data and further data from your Withings account

Under your explicit consent, these data will be shared with the AIDAVA study and curated. A copy of the resulting curated data will be transmitted back to your MIDATA account and visualised by you.

Data storage within MIDATA

Personal health data you enter into MIDATA is stored in a personal encrypted data account on the MIDATA platform. Your personal encrypted data account is opened in a registration process.
When registering on the MIDATA platform, the following required information must be provided as part of the registration process:

- Name
- First name
- Email address
- Gender
- Date of birth
- Country of residence

This information is used for registration and identification purposes and, if necessary, for MIDATA to handle requests.

During registration, you accept the General Terms and Conditions (https://ch.midata.coop/#/public/terms?which=midata-terms-of-use) and the Privacy Policy (https://ch.midata.coop/#/public/terms?which=midata-privacy-policy).

Only you have access to and control over the personal health data which are encrypted and stored in your data account.
You can also connect your already existing MIDATA data account.

Consent to data sharing with the AIDAVA study

You consent to the use of the data entered by you as part of your participation in the AIDAVA study. The following data will be shared with [enter hospital name] operating the AIDAVA prototype:

- GP data
- Quality of life questionnaire via connected app
- Withings Blood Pressure Monitor data and further data from your Withings account

Risks and benefits for participants

Please see section on Benefits and Issues in the document “Information for patients on the assessment of the AIDAVA prototype” [link to SIP/ICF]

Withdrawal process and data deletion

You have the option at any time to opt out of the sharing of collected data on the MIDATA portal https://ch.midata.coop. You can export your data at any time and delete your MIDATA account and thus all data stored there.

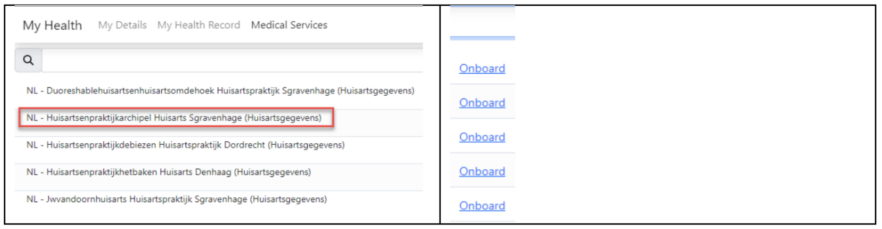
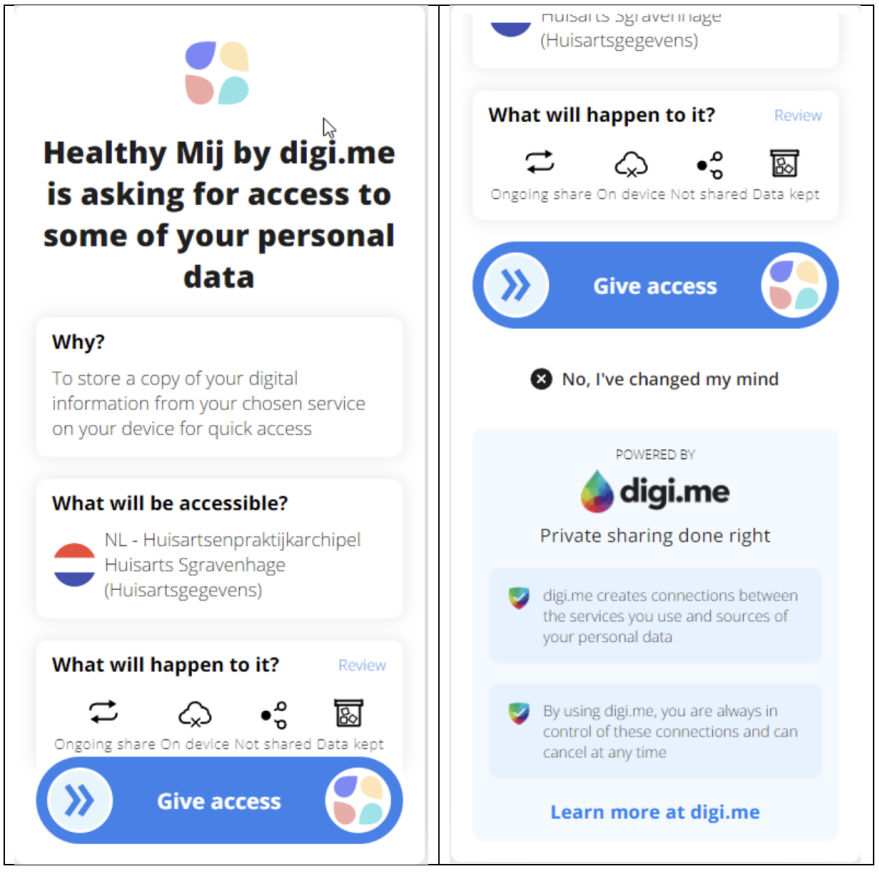
Contact point for questions

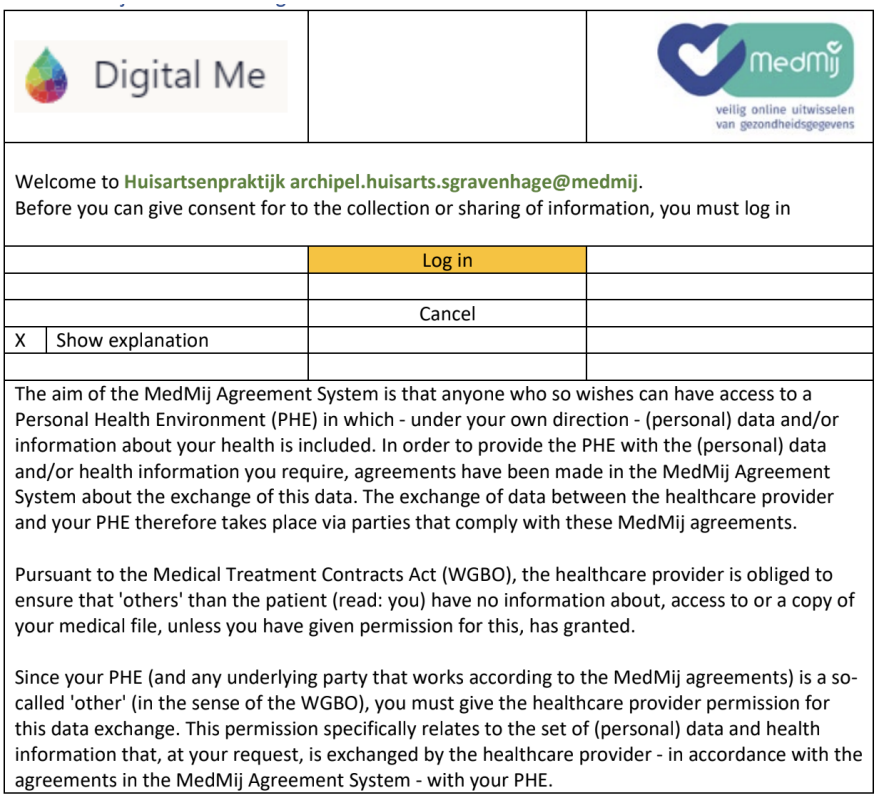
[Address]
[Email]

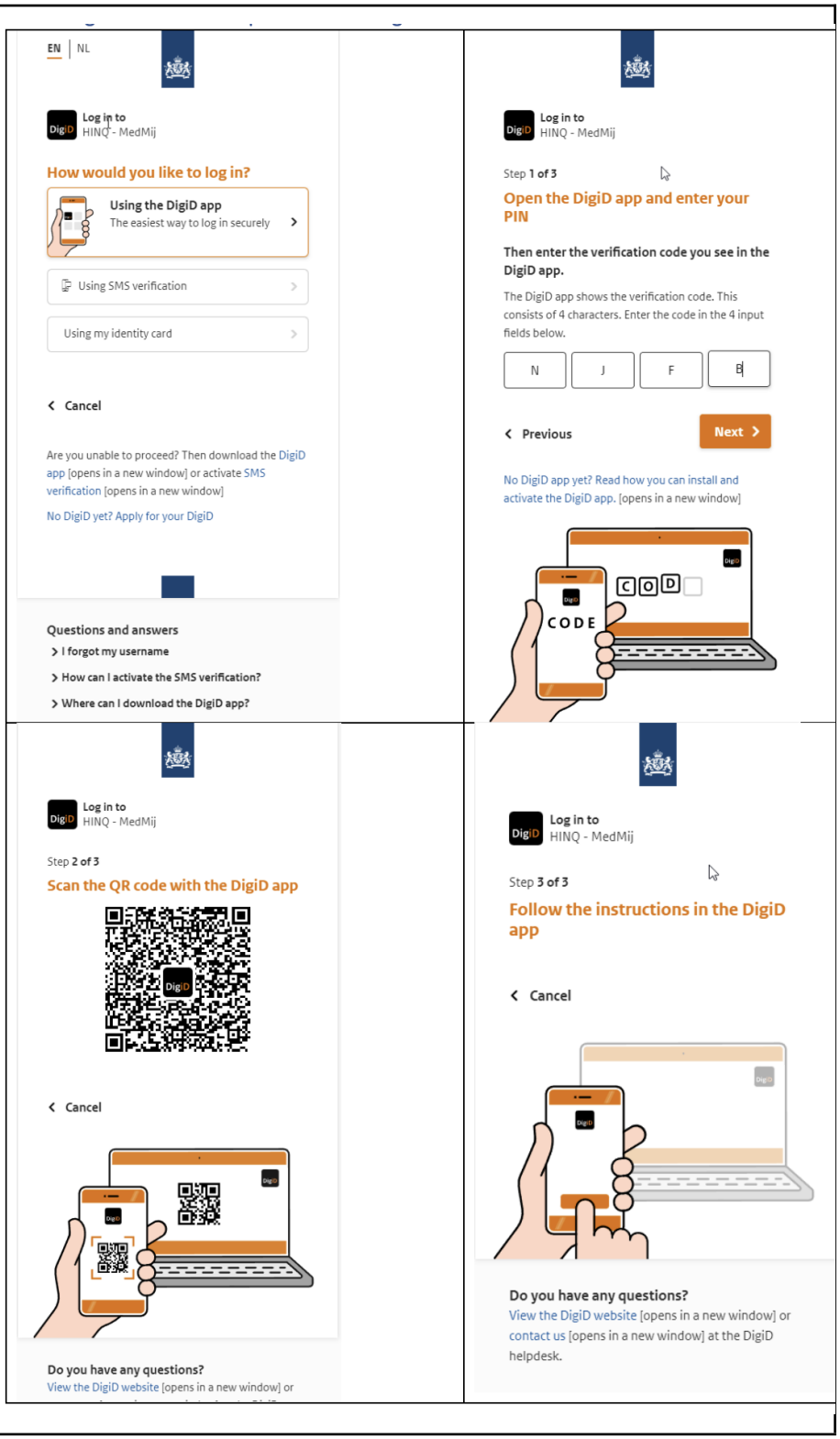
Annex 2. DIGI.me Collaboration Agreement

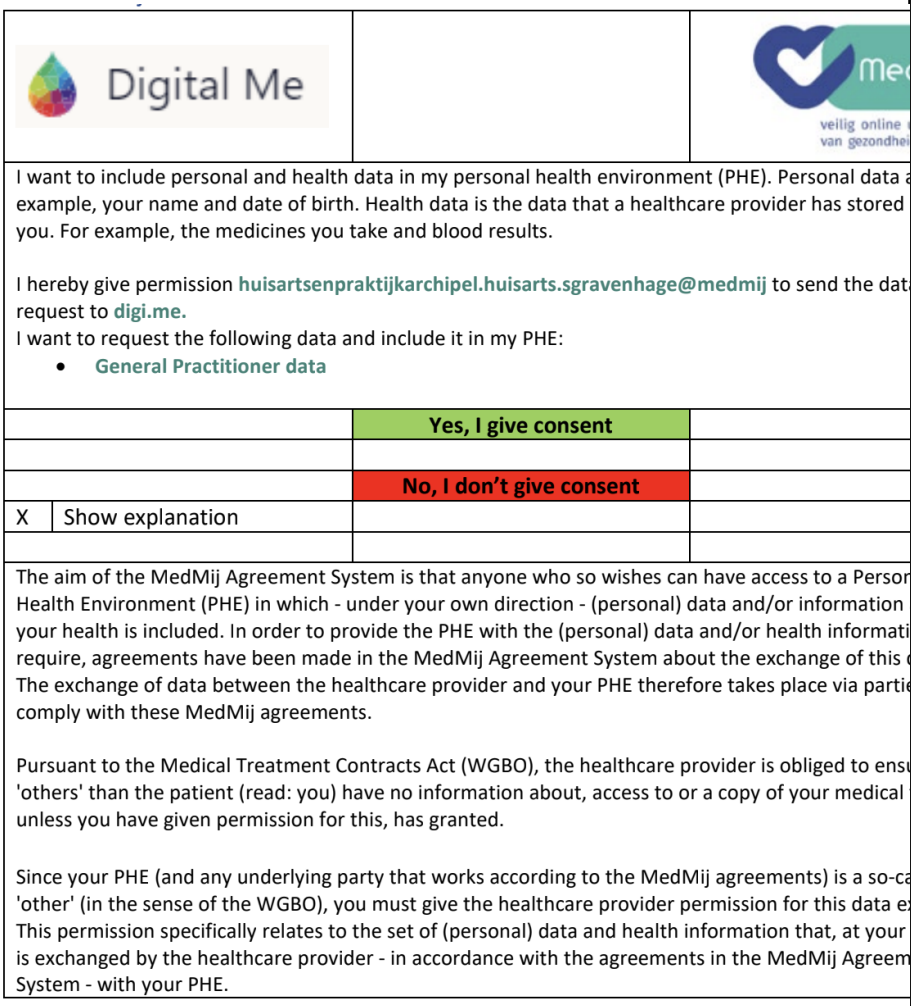
We will follow the national process, defined by MedMij. MedMij's core task is to facilitate the digital exchange of health data between residents of the Netherlands and their care-givers. MedMij is also creating confidence that this is done in a secure, affordable, future-proof and user-friendly way. One of the ways MedMij is doing this is by designing a solid framework. ("MedMij Framework" 2021)

Phase 1. Agree to have data managed by Digi.me

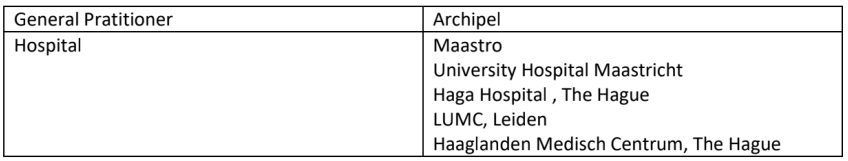
Step 1. Patient selects their health care provider through the MedMij national portal	
Step 2. Patient give consent that their data from the provider are sent to Digi.me	

Step 3. Patient need to log in to identify itself - from DIGI.me	

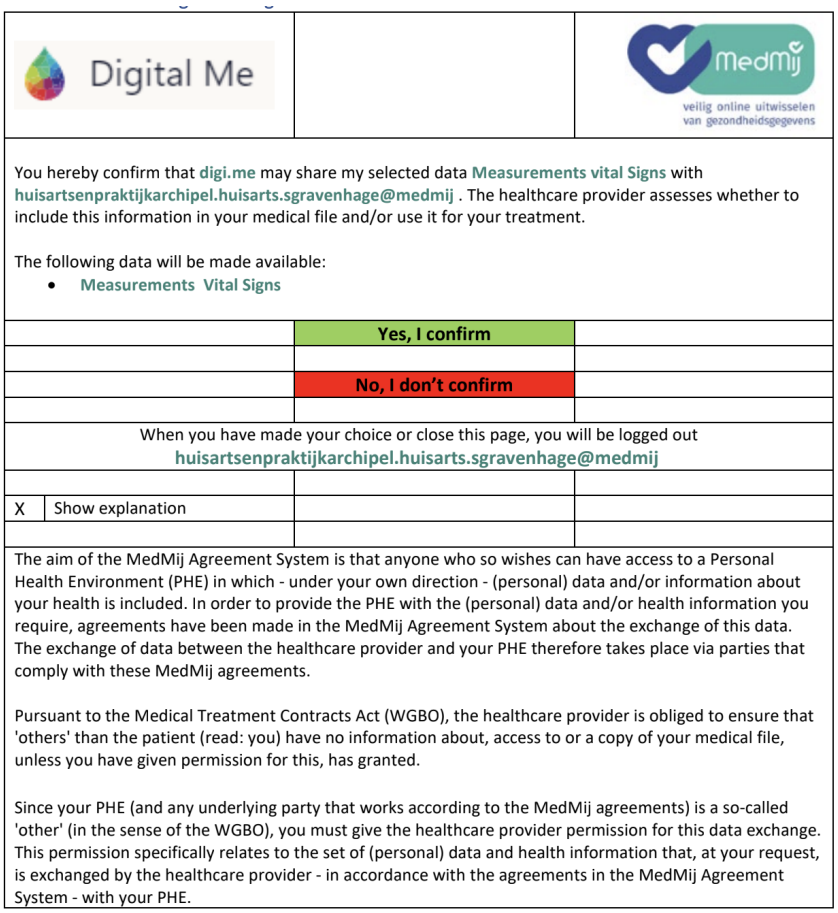
Step 4. Patient provides credential through national digital identification system (used for multiples services in the NL)	

Step 5. Patient confirms consent.	

Phase 2. Agree to share data with the AIDAVA project

Step 1. Select the provider/ hospital with whom to share data in the context of AIDAVA	
Step 2 and 3. Login and confirm identity	Same as step 3 and 4 in the previous phase .

Step 4. Confirm consent to share data



"MedMij Framework." 2021. MedMij. December 22, 2021. https://medmij.nl/en/medmij-framework/.

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

Deliverable No. D1.4
Annex 3 - Training plan

Contractual Submission Date: 30/11/2023

Actual Submission Date: 30/11/2023

Responsible partner: P8- NEMC

Funded by
the European Union

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

Deliverable number	**Annex 3 to Deliverable D1.4**
Deliverable title	**D1.4. Description of Assessment Study (and supporting material)**
Type[1]	R
Dissemination level[2]	SEN
Work package number	WP1
Work package leader	P8 - NEMC and P2 - b!loba
Author(s)	Isabelle de Zegher (P2-b!loba) Mall Maasik, Kerli Norak (P6-NEMC)
Reviewers	Katrin Lepik, Eno-Martin Lotman (P6-NEMC) Markus Kreutzthaler (MUG) Petros Kalendralis (UM) Dominik Steiger (MID) Hanne Muller (DME)
Keywords	Training

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).

The work of MID received funding by the Swiss State Secretariat for Education, Research and Innovation (SBFI), subvention contract 22.00093, REF-1131-52104.

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

Document History

Version	Date	Description
V1	April 2023	First version following identification of needs from user personas
V2	9.11.2023	Draft ready with initial review by partners
V3	30.11.2023	Final version ready

[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 Abbreviations and definitions

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

AIDAVA - AI powered Data Curation & Publishing Virtual Assistant
BC - Breast Cancer
BPM - Blood pressure monitor
CVD - Cardiovascular Disease
D - Day
EHDS - European Health Data Space
EQ-5D QLY - quality of life questionnaire
G0 - Generation 0 (current situation)
G1 - Generation 1 (of prototype)
G2 - Generation 2 (of prototype)
GDPR - General Data Protection Regulation
HDI - Health Data Intermediary
ICF - Informed Consent Form
RA - Research associate
SIP - Study Information Package
UI - User interface
V - Visit
W - week

Foreword

This document contains the overview of the training plans regarding the ADAVA prototype testing phase. The details on media types and actual duration will be agreed upon drafting stages. Any training materials that include information given to patients when recruiting have been drafted, reviewed and completed and sent for ethical committees for approval. Any changes to the materials regarding patients will be subject to a change to the ethical applications of each site.

1. Modules: timeline and participants

Guidelines:

- Main module common for all sites - add a few slides per country/ per HDI if needed
- All modules for patient should be translated in local languages
- Possibly look into modular durations or an adaptive content system, allowing users to select the depth of information they require.
- Media types will be agreed upon at draft creation stage

Link	Description & objectives	Time (min)	Lead partner	Comments	Media	Draft	Re-viewed	Ready	Update	patient	curator	Interested clinical staff	Site architects	Site Admin	Data users
M1a	Intro: AIDAVA project.	15 pres*creening*	b!lo	to be ready when recruitment of BC patients	study information package + face to face	Nov 23	Jan 24	Feb 24		all patients	curator	Interested clinical staff	Site architects	Site Admin	Data users
M1b	Intro: assessment study		NEMC/ b!lo	to be ready when recruitment of BC patients	study information package + f2f	Nov 23	Jan 24	Feb 24		all patients	curator	Interested clinical staff	Site architects	Site Admin	Data users
M1c	Intro: Health Data Intermediary (HDI)		MiDATA Digi.me	to be ready when recruitment of BC patients	study information package + f2f	Nov 23	Jan 24	Feb 24		all patients	curator	Interested clinical staff	Site architects	Site Admin	Data users
M2a	Details to AIDAVA assessment study - overview	60	NEMC/ b!lo	to be ready when reqruitment of BC patients	study information package + f2f	Dec 23	Jan 24	Feb 24		patients on request		Interested clinical staff inc. research associate			
M2b	Details to AIDAVA assessment study - HDI agreement	5	MiDATA Digi.me	to be ready when reqruitment of BC patients	study information package + f2f	Dec 23	Jan 24	Feb 24		all patients					

Link	Description & objectives	Time (min)	Lead partner	Comments	Media	Draft	Re-viewed	Ready	Update	patient	curator	Interested clinical staff	Site architects	Site Admin	Data users
M2c.1	Details on AIDAVA assessment study - medical device	10	Digi.me + UM	Instructions from vendor+connections +withings account	**training while using device + paper instructions**	Jan 24	Mar 24	May 24		CVD patient only					
M2c.2	Details on AIDAVA assessment study - medical device	10	MiDATA +NEMC, MUG		**training while using device + paper instructions**	Jan 24	Mar 24	May 24							
M2d.1	Details on AIDAVA Assessment study - EQ-5D QLY and 3rd party app for UM	10	Digi.me + UM	version 5L	**training while using app**	Jan 24	Mar 24	May 24		all patients					
M2d.2	Details on AIDAVA Assessment study - EQ-5D QLY and 3rd party app for MUG and NEMC	10	NEMC / MUG Midata		**training while using app**	Jan 24	Mar 24	May 24							
M3a	Data Curation - Basics	10	b!lo			Jan 24		Feb 24			curator	Interested clinical staff	Site architects		
M3b	Data Curation - Advanced	60	b!lo/ UM - DACS			Jan 24		Feb 24					Site architects		
M3c	Data Curation - Data Source Onboarding & update Data Catalogue	60	UM-DACS/ GND			Nov 24	Jan 24	Jan 24					Site architects		
M4a	Using AIDAVA VA - Basics	15	GND	local language	demo	Feb 24	Apr 24	May 24	Jun 24	all patients	curator	Interested clinical staff	Site architects	Site Admin	Data users
M4b	Using AIDAVA VA - Advanced	30	GND			Feb 24		May	Jun 24		curator			Site Admin	
M4c	*Quality control and validation*	*15*	*IHD*			Jan 24		Feb 24			*curator*			*Site Admin*	*Data users*

Link	Description & objectives	Time (min)	Lead partner	Comments	Media	Draft	Re-viewed	Ready	Update	patient	curator	Interested clinical staff	Site architects	Site Admin	Data users
M5	*Data security and privacy advanced*	*15*	*GND + Site Admin*			Jan 24		Feb 24			*curator*				*Data users*
M6	Troubleshooting and support	10	GND			Feb 24		May 24			curator			Site Admin	
M7a	Practice exercises in test environment with synthetic data- Basics	30	b!lo	local language		Feb 24		May 24	Jun 24	all patients	curator	Interested clinical staff	Site architects	Site Admin	Data users
M7b	Practice exercises in test environment with synthetic data- Advanced	30	b!lo			Feb 24		May 24	Jun 24		curator			Site Admin	

2. Training plan for patients

Training plan for patients	Screen (< 15 d before W1)	Generation 1				Develop G2	Rescreen (< 15 d before W1)	Generation 2				
		W1 D1 V1	W1	W1 D5	W2 D1 V2			W1 D1 V1	W1	W1 D5	W2 D1 V2	W2
INTRO TO STUDY (by RA) **Intro: AIDAVA project** Support understanding of the project aim and support readiness to collaborate **Intro: assessment study** Explain to objective and approach of the study as well as eligibility criteria **Intro: Health Data Intermediary (HDI).** Introduction of what HDI are and how they can help the patient to control their data. Explain HDI agreement so that patient feels OK to sign ICF and HDI agreement.	15'-20' BC	15' CVD				(16 to 18 months)	15'-20' BC	15' CVD				
TRAINING ON SITE **Details to AIDAVA assessment study - overview** Explain to patient what is expected during the study **Details on AIDAVA Assessment study - EQ-5D QYL and 3rd party app** Explain to patients why, how, when to use **Using AIDAVA VA - Basics + Practice exercises** Overview of the software, its purpose, and how it can help patients to manage non hospital data and with hospital data. Hands-on exercises based on synthetic data and quizzes **Details on AIDAVA assessment study - medical device** Explain to patients by research associates how to use the medical device, why it is important / CVD patients only		40' + 10' (CVD)						40' + 10' (CVD)				
TRAINING ON LINE (refresh of training on site)					x						x	

3. Detailed training plan

Module	Description & objectives	Time (min)	Owerview of content
M1a	**Intro: AIDAVA project.** **Objectives:** Support understanding of the project aim and readiness to collaborate during screening	15 during pre*screening*	• Objective of the project • Proposed Approach and Benefits • Milestones
M1b	**Intro: assessment study** **Objectives:** Explain objective and approach of the study as well as eligibility criteria		• Objective of the study • Milestones • What we will need from patients, eligibility, workload
M1c	**Intro: Health Data Intermediary (HDI)** **Objectives:** Introduction of what HDI are and how they can help the patient to control their data		• What is HDI? • What are the benefits? • How we work with HDI in AIDAVA?
M2a	**Details to AIDAVA assessment study - overview** **Objectives:** Explain to patient and research associate what is expected during the study	60	• Details on schedule of activities • What we need from patients • How we fill in different forms / questionnaires
M2b	**Details to AIDAVA assessment study - HDI agreement** **Objectives:** Explain HDI agreement so that patient feels OK to sign ICF and HDI agreement	5	• Further details on HDI in context of EHDS and GDPR (high level for patient) • Details on HDI agreement • Present LOGO and UI of HDI • Data Privacy aspects with 3rd party app (Withings BPM, QLY)
M2c.1	**Details on AIDAVA assessment study - medical device** Digi.me + UM **Objectives:** Explain to patients by research associates how to use the medical device, why it is important.	10	• Why we need medical device? • How to use medical device: registering, collect data, visualize data? • What to do in case of problem?
M2c.2	**Details on AIDAVA assessment study - medical device**. MiDATA +NEMC, MUG **Objectives:** Explain to patients by research associates how to use the medical device, why it is important.	10	• Why we need medical device? • How to use medical device: registering, collect data, visualize data? • What to do in case of problem

M2d.1	**Details on AIDAVA Assessment study - EQ-5D QLY and 3rd party app** for UM **Objectives:** Explain to patients why, how, when to use app	10	• Why we need 3rd party app? • How to use medical device: registering, collect data, visualize data? • What to do in case of problem
M2d.2	**Details on AIDAVA Assessment study - EQ-5D QLY and 3rd party app** for MUG and NEMC **Objectives:** Explain to patients why, how, when to use app	10	• Why we need 3rd party app? • How to use medical device: registering, collect data, visualize data? • What to do in case of problem
M3a	**Data Curation - Basics** **Objectives:** Overview of what data curation is and why it's important for patients to understand how to curate their own health data.	10	• Objective of data curation • Challenges • How to curate data with AIDAVA • Benefits for patients, physicians and HealhCare
M3b	**Data Curation - Advanced** **Objectives:** Description of the different data interoperability issues and the related workflow to maximize automation in curation	60	• Data interoperability issues • Orchestration of automation • Review of each workflow
M3c	**Data Curation - Data Source Onboarding** **Objectives:** Detailed description of the process to support expert in resolution of issues	60	• Intro to data standards • Deep dive on the data interoperability issues • Description of data source onboarding
M4a	**Using AIDAVA VA - Basics** **Objectives:** Overview of the software, its purpose, and how it can help patients to manage non hospital data and hospital data	15	• Overview of what AIDAVA does (and it does NOT do) • Acquisition of GP data • Key functionalities (ingestion, curation, view) • Customer support
M4b	**Using AIDAVA VA - Advanced** **Objectives:** In depth instructions on how to use the software interface to solve more complex issues	30	• Admin module • Data source onboarding in each site • Other functionalities
M4c	**Quality control and validation** **Objectives:** Techniques for ensuring the accuracy and completeness of curated data	*15*	• Approach to data quality management • How it is implemented and meaning of scores
M5	**Data security and privacy advanced**	*15*	• Approach to ensure data security and privacy • How it is implemented in AIDAVA

	Objectives: Explanation of the importance of data security and privacy, including best practices for protecting sensitive data and complying with data protection regulations		
M6	**Troubleshooting and support** **Objectives:** Information on how to troubleshoot common issues and where to find support and resources for using the software	10	● Customer support mechanism
M7a	**Practice exercises in test environment with synthetic data-Basics** **Objectives:** Hands -on exercises and quizzes to help patients practise using the software	30	● Exercise based on synthetic data
M7b	**Practice exercises in test environment with synthetic data-Advanced**	30	● Exercise based on synthetic data

4. Feedback from Patient consultants

	Suggest to have a quick after the training to make sure patient understood
Training material (Module 1)	● Project is complex, so could consider creating animations and visuals to support the M1a training (intro to AIDAVA) of training material ○ Also highly recommend to send the animation before patients join the M1 training meeting so they have time to digest the concept of AIDAVA and consider questions they might have ● Helpful to have verbal and visual content in training material to better understand the project
Training material (Module 2)	● Very important to use lay-language for all training material and technical terms - i.e. Even consider terms like “ingest”, which could be simplified to “upload”, “insert” or “add” data ● A glossary would not necessarily be a good solution or “enough” as it will still be complicated to read through the material ● Be clear about the overarching title/aim of each training module and sub-module
On HDI	● Be clear on ○ why there are 2 HDIs in the project ○ why HDIs are involved in different hospital sites ○ which HDI the agreement is connected to => The site patients will only be informed about ONE HDI (MIDATA OR DIGI.ME), the one that works with their referral hospital. They will be asked to check and sign the agreement with that HDI ● HDI agreement must be clear on data privacy principles between the HDI and 3rd Party App so patients understand the different pathways of their personal health data => Pathway is : 3rd party app -> HDI -> AIDAVA Data privacy rules: data sharing is linked to formal data sharing agreement (checked by AIDAVA Data Protection Officer - partner IHD ● HDI agreement should be in lay-language - suggestion to integrate support system in HDI agreement, i.e., include FAQs to each section of the agreement or visuals

5. For admin

Information (as part of the training)	**General information** ● Goals of the AIDAVA tool ● Purpose of the usage of health data in AIDAVA in a detailed way ● How the tool is supposed to interact with the local hospital information system and subsystems ● Define tasks expected from admins **Ensure system can be trusted** ● Legal foundation on which AIDAVA is based ● How AIDAVA protects the data that have been ingested ● Confirm that AIDAVA is not a medical product ● Confirm that AIDAVA does not make diagnoses ● How data in the AIDAVA app (smartphone of a patient) is accessible **Support** ● Clear definition of responsibilities for AIDAVA ● How often there will be updates and security patches ● Need to know if back-ups are necessary and if yes, how

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

Deliverable No. D1.4
Annex 4 - Data Sharing Agreement for data exchange between Hospitals and HDIs

Contractual Submission Date: 30/11/2023

Actual Submission Date: 30/11/2023

Responsible partner: P8- NEMC

Funded by
the European Union

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

Deliverable number	**Annex 4 to Deliverable D1.4**
Deliverable title	**D1.4. Description of Assessment Study (and supporting material)**
Type[1]	R
Dissemination level[2]	SEN
Work package number	WP1
Work package leader	P8 - NEMC and P2 - b!loba
Author(s)	Isabelle de Zegher (P2-b!loba)
Reviewers	Kerli Norak, Elion Sõber (P8-NEMC) Markus Kreuzthaler (MUG) Petros Kalendralis (UM) Dominik Steiger (MID) Hanne Muller (DME)
Keywords	

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).

The work of MID received funding by the Swiss State Secretariat for Education, Research and Innovation (SBFI), subvention contract 22.00093, REF-1131-52104.

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

Document History

Version	Date	Description
V1	September 2023	First draft for review
V2	November 2023	Final version ready

[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

Foreword: how to use this documents

This document is a template Data Sharing agreement, expected to be completed by each partner as a support for their local Ethical Committee approval when requesting approval for the AIDAVA assessment (for generation 1 and generation 2 of the prototype).
We expected we will need 2 of these documents filled in per medical centre:

- One for flow 4 below
- One for flow 5 and 6 below - in a single document

Based on use cases identified in Deliverable D1.1, six data exchanges are needed as described in the table below. This document is focusing on data exchanges at hospital level, i.e. numbers 4, 5 and 6 in Figure 1. digi.me
Data sharing agreements between other Data Providers (such as GP, medical device apps, PROMS personal apps) and the Health Data Intermediary (HDI), i.e. numbers 1, 2 and 3, are out of scope of this document; it is expected that each HDI will use their own existing agreements and that they will transfer the data received from the patient to AIDAVA (data exchange number 5).

Partner MIDATA is the HDI for Partner NEMC (Estonia) and Partner MUG (Austria).
Partner Digi.me is the HDI for Partner UM (The Netherlands).

#	Data Holder	Data Recipient	Data Shared	Data Processor	Comments
1	Holder of GP data based on national practice	HDI	GP Data	HDI *(P-MID for Estonia and Austria* *P-DME for Netherlands)*	
2	Holder of data 3rd Party app (including EQ5D)		QLY data		
3	Holder of medical device data (Withings BPM)		Blood Pressure		Only for CVD patients
4	Hospital (production env)	AIDAVA@ hospital	Data in scope of use cases (Section 10.1 Study protocol)	P-GND	DSA in use from hospital DTS required
5	HDI *(MIDATA or Digi.me)*				DSA template from hospital DTS required
6	Hospital (AIDAVA prototype)	HDI	Patient IPS in HL7 - and potentially full PHKG	HDI	DSA template from hospital DTS based on HL7 IPS standard

Scope of the data exchange

Based on use cases defined in the AIDAVA project and described in the study protocol (Section 2.2. Background) six data exchanges are needed as displayed in Figure 1.

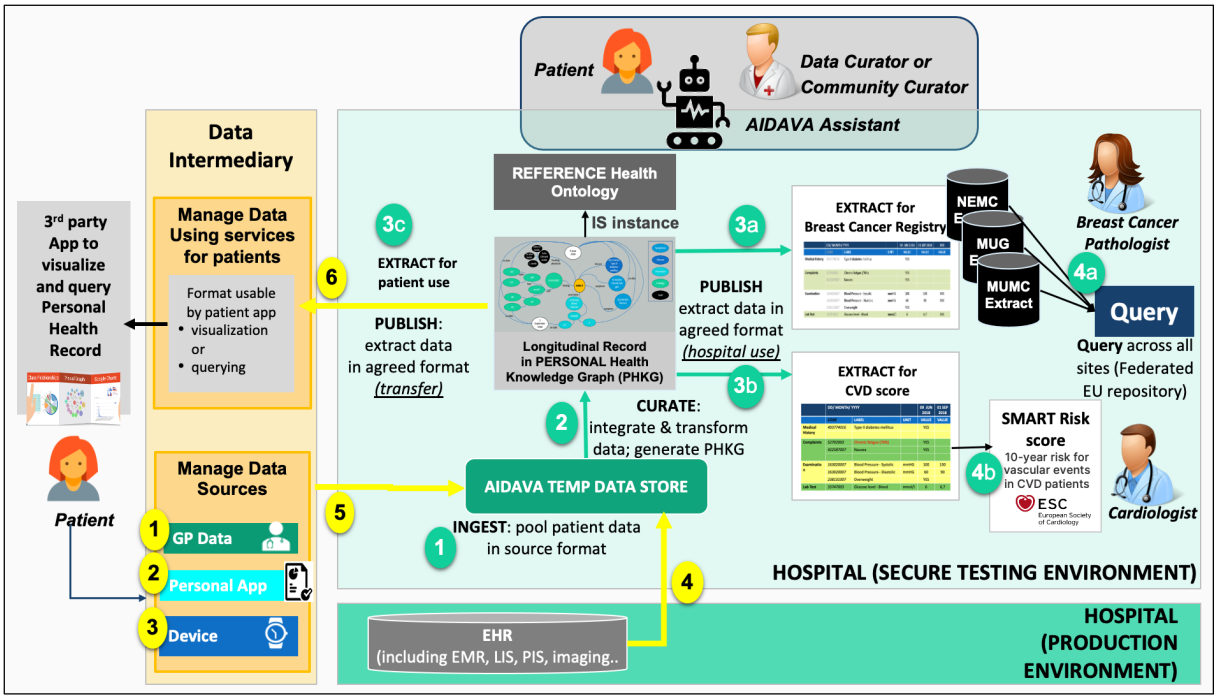


Figure 1. AIDAVA prototype and data flows. Flow and bullet in yellow represent external flows to AIDAVA, subject to Data Sharing/Processing Agreement with Data Transfer Specifications

Data exchange in scope of this document

- Yellow arrow and bullet: 4, 5 and 6
- Data Holder: *to complete (see table above based on flow in scope)*
- Data Recipient: *to complete (see table above based on flow in scope)*
- Data Processor: *to complete (see table above based on flow in scope)*
- Data in scope : *to complete (see table above based on flow in scope)*

All information is considered as sensitive and subject to Article 13 of the Grant Agreement (see Annex), signed by all the consortium partners with the European Commission, and mentioning: *"The parties must keep confidential any data, documents or other material (in any form) that is identified as sensitive in writing ('sensitive information') — during the implementation of the action and for at least until the time-limit set out in the Data Sheet".*

Data privacy and security aspects developed here are compliant with the information governance instruments described in *Deliverable D4.4. Information Governance Framework and Instruments.* Specifically it is expected that each Data Holder has performed a Data Protection Impact Assessment (DPIA); this will be checked at the beginning of the assessment study.

1. Legal and data privacy provisions

In case of data flow 4 - explain which agreement is already in place for such projects (with reference is possible).
In case of data flow 5 and 6, provide link to Data Sharing Agreements already in use within hospital for such projects - with potential adaptation for AIDAVA.

The following agreements need to be completed within the last quarter of 2023.

DSA per Data Flow	Flow	Required	Comments related to DSA
ESTONIA			
NEMC to AIDAVA	4. NEMC to AIDAVA@NEMC	DTS only	Based on current practice and CA (GND as processor)
MIDATA to NEMC	5. HDI to AIDAVA@NEMC	**MIDATA DSA;** DTS	One single contract for both transfers - based on proposal from MIDATA
NEMC to MIDATA	6. AIDAVA@NEMC to HDI	**NEMC DSA**; IPS	
AUSTRIA			
MUG	4. MUG to AIDAVA@MUG	DTS only	Same as NEMC
MIDATA to MUG	5. HDI to AIDAVA@MUG	**MIDATA DSA;** DTS	One single contract for both transfers - based on proposal from MIDATA
MUG to MIDATA	6. AIDAVA@MUG to HDI	**MUG DSA;** IPS	
The Netherlands (Maastro for BC; UMUC for CVD; AIDAVA within UMCU			
MUMC to AIDAVA	4. MUMC to AIDAVA@MUMC	DTS only	(need to check CA)
DME to AIDAVA@MUMC	5. HDI to AIDAVA@MUMC	**DME DSA;** DTS	One single contract for both transfers - based Digi.me / MedMij requirements
AIDAVA@MUMC to DME	6. AIDAVA@MUMC to HDI	**UMCU DSA,** IPS	
Maastro to AIDAVA	4. Maastro to AIDAVA@Maastro	DTS only	(need to check CA)
DME to AIDAVA@Maastro	5. HDI to AIDAVA@Maastro	**DME DSA**; DTS	Same as for MUMC
AIDAVA@Maastro to DME	6. AIDAVAMaastro to HDI	**Maastro DSA**; IPS	

2. Technical provisions

3.1. Data Transfer Specifications

The enclosed table includes a link to the technical specifications for the different documents identified as relevant for the use cases (see as well Assessment Study Protocol - Section 10.1). The same DTS can contain different documents.

link to the file with the data transfer specifications with some explanation in case of several files)
(for clinical sites - files from EHR)

Type of information	Short description on information use	BC	CVD	Reference to DTS	Data Steward
Discharge Summary/Discharge letter	A handover document that explains to other healthcare professionals why the patient was admitted, what has happened to them in hospital, and other information needed to continue care	YES	YES	*To fill in ny hospital*	*Resp for data source*
Medical history	A record of information about a person's health. A personal medical history may include information about allergies, illnesses, surgeries, immunizations, and results of physical exams and tests.	YES	YES	*To fill in ny hospital*	*Resp for data source*
Progress notes	Progress notes are intended to provide an updated analysis of a patient's condition and progress during treatment.	YES	YES	*To fill in ny hospital*	*Resp for data source*
Prescribed medications	A prescription drug is a pharmaceutical that requires a medical prescription to be dispensed and their full list is provided usually at the end of a therapeutic encounter.	YES	YES	*To fill in ny hospital*	*Resp for data source*
Medical imaging reports	Contains the interpretations of images with the main goal of presenting the outcomes of the imaging procedure (e.g. X-ray, MRI) of the patients to physicians	YES	YES	*To fill in ny hospital*	*Resp for data source*
Pathology reports	Contains morphological description, diagnosis, predictive and prognostic factors of tumour and pTNM. Includes reports of cytological specimens, biopsies and surgical specimens	YES	NO	*To fill in ny hospital*	*Resp for data source*
Surgical procedure descriptions	Detailed description of surgical procedure contains type(s) and extent of surgery, specimens removed and intraoperative pathology consultation report	YES	NO	*To fill in ny hospital*	*Resp for data source*
Multidisciplinary meeting reports	Multidisciplinary meeting reports contain information about staging, treatment indications and decisions.	YES	NO	*To fill in ny hospital*	*Resp for data source*
TNM staging	TNM staging describes stage of tumour	YES	NO	*To fill in ny hospital*	*Resp for data source*
Patient referral document	Patient referral documents contain information about current symptoms and disease, medical imaging. In some cases it can be on paper.	YES	NO	*To fill in ny hospital*	*Resp for data source*
Laboratory reports	Describe the results of laboratory tests performed for the patient samples	NO	YES	*To fill in ny hospital*	*Resp for data source*

Type of information	Short description on information use	BC	CVD	Reference to DTS	Data Steward
Echocardiography report	Echocardiography is a diagnostic tool for diagnosis and follow-up of heart disease and its report provides the interpretation of the medical imaging procedure	NO	YES	*To fill in ny hospital*	*Resp for data source*
Coronary angiography report	Coronary angiogram is a procedure that uses X-ray imaging of the heart's blood vessels and it report provides the interpretation of the medical imaging procedure	NO	YES	*To fill in ny hospital*	*Resp for data source*
Ambulance record	A record of care provided during the ambulance stage of treatment.	NO	YES	*To fill in ny hospital*	
Emergency department record	A record of care provided during the emergency department stage of treatment.	NO	YES	*To fill in ny hospital*	*Resp for data source*
ICU/CICU progress notes	Provide an updated analysis of a patient's condition and progress during the ICU stay.	NO	YES	*To fill in ny hospital*	*Resp for data source*

(for HDI only !)

Type of information	Short description on information use	BC	CVD	Reference to DTS	Data Steward
GP record	Contains notes and information from the GP on conditions, lab results, allergies, prescriptions. It is less detailed than any information contained in the EHR. As GP records typically cover the lifetime of a patient, they are offering a truly longitudinal perspective.	YES	YES	*To fill in by HDI*	*Resp for data source*
Personal App	Quality of Life Questionnaire (EQ-5D 9QLY) running on a smartphone -	YES	NO	*To fill in by HDI*	*Resp for data source*
Connected medical device	Certified digital devices that help gather vitals (and potentially many other parameters) directly from the patient. One approved medical device will be identified for this use case.	NO	YES	*To fill in by HDI*	*Resp for data source*

3.2. Processing and Security aspects for patient data during testing

(for detailed description of data transfers- see section 1)

4 = data transfer from Hospital production system to AIDAVA

5 = data transfer from HDI to AIDAVA@hospital

6 = data transfer from AIDAVA@hospital to HDI (IPS and potentially full PHKG)

Question	Answer from Data HOLDER
Transmission method.	• Frequency of the exchange *If 4 or 5 =* Once at the beginning of testing of each generation of the prototype (push from hospital systems into the file share within VPN). Data of each patient will be transferred independently. *If 6 =* Twice during the evaluation period by a patient: after week 1 and after week 2. • Data exchange will be incremental (for 4 and 5): exchange of any data source should be complete, additional data sources can be exchanged as they become available during evaluation. • Security classification and level of risk of the data: SENSITIVE data including personal identifiable data
Transmission encryption (*If data encryption is required, what encryption method will be used?)*	• *If 4 =* (to be specified by each hospital) MAASTRO/MUMC: HTTPS based using SSL • *If 5 =* *For MIDATA (NEMC, MUG)*: Data traffic from and to server via HTTPS, TLS 1.3. Connection and all safety tokens using perfect-forward secrecy. *For Digi.me (MUMC, Maastro):* medmij framework ("MedMij Framework" 2021) • *If 6 =* HTTPS based *Note: AIDAVA UI input by the patient may also require encryption if the UI resides outside of the VPN; if this is inside of the VPN - we will use a https certificate.*
Infrastructure details.	See technical architecture in next section
Error handling *(How errors will be handled during exchange. Include a process flow if required)*	• *If 4 =* (to be specified by each hospital) MAASTRO/MUMC: HTTP status code to indicate the output of the request. • *If 5 =* *For MIDATA (NEMC, MUG)*: Traceability of data processing of sensitive data is ensured: Entries are created in the audit log for each security-critical action. The author is saved for each data record and for each update of a data record. The change history for each data record is available. *For Digi.me (MUMC, Maastro):* The data will be validated by the data manager within the server and the data will be stored within the hospital. Digi.me is not allowed to know the exact location of the data storage according to law. Important point is to not deliver or exchange data directly to the EHR. • *If 6 =* AIDAVA makes detailed logs of everything happening in the system, in case of any error, responsible persons will be notified.

Question	Answer from Data HOLDER
Issues management - contact details.	Person responsible for issue management (Data Holder) ● Name: *TBD* ● Role: *TBD* Person responsible for issue management (Data Recipient) ● Name: *TBD* ● Role: *TBD*
Environment Version Management	During software development and testing, there is typically three separate environments ● **Development environment**: ○ Used by developers, for development, integration and tests. ○ Created and hosted by Partner GND. ● **Test environment**: ○ Used by testers from Partner GND, nominated AIDAVA and site administrators have access to do manual testing and configuration. ○ Created by Partner GND on hardware hosted by the medical centre responsible for evaluation. ● **Production environment**: ○ The live instance, used by all AIDAVA users, changes only according to change management rules. ○ Created by Partner GND on hardware hosted by the medical centre responsible for evaluation.
Data Storage *(How and where will data be stored once it is received? May include server names, systems, hosting location and relevant security standards)*	See technical architecture in next section ● *If 4 or 5 =* AIDAVA local version in the hospital ● *If 6 =* *For MIDATA (NEMC, MUG):* The platform is hosted in Switzerland, using the Swisscom Dynamic Computing Services (DCS) infrastructure, in Swisscom data centres. DCS is certified according to ISO 20000, 27001, 9001, 14001, ISAE 3402, GDPR (data processor). The database is stored in a replica set with 3 replicates. *For Digi.me (MUMC, Maastro):* The data storage is in an Azure cloud environment certified by NEN 7510. It is not allowed to be stored as a HDI without the above certification according to MedMij. That wa salso the reason that we selected Brightfish as it is also certified by NEN 7510 for the EQ5D-5L patient questionnaire.
Data Security (including disaster recovery and access control)	● *If 4 or 5 =* procedures in place in the hospital ● *If 6 =* *For MIDATA (NEMC, MUG):* For complete safety and security features, see annex. The database is backed up using the Swisscom Dynamic Storage service, using geo redundant mirroring in several Swisscom data centres. *For Digi.me (MUMC, Maastro):* According to NEN 7510 and the

Question	Answer from Data HOLDER
	additional Medmij quality certification, there is a data security test yearly done by the NEN 7510 auditor (DNV-GL). The process is evaluated (whether there is a disaster). Penetration test is part of the security test.
Data Disposal.	• *If 4 or 5 =* In line with research governance regulations, and taking into account that AIDAVA is a research prototype with no impact on clinical care or clinical research, all patient data transferred to AIDAVA, the resulting curated data (Personal Health Knowledge Graph (PHKG) and the derived data (IPS of the patient, extract to compute CVD score of the patients and BC registry extracts) will be deleted from the system - by Partner GND - at the end of the project, and the solution will be retired. Patients will have the opportunity to ask and receive a copy of their PHKG in electronic format. • *If 6 =* *For MIDATA (NEMC, MUG):* As a HDI, MIDATA acts in fiduciary capacity on behalf of its data account holders. As data account holders, the patients have sovereign rights over their encrypted data accounts and the data contained therein. They may export their data, or delete their accounts, or further use the accounts. Data sovereignty of the account holders over the data sets in their data accounts is technically ensured and enshrined in the Articles of Association (https://www.midata.coop/wp-content/uploads/2019/08/MIDATA_Statuten_20190626_EN.pdf) and the Privacy Policy (https://ch.midata.coop/#/public/terms?which=midata-privacy-policy); use of the data sets available in the encrypted data accounts by third parties, including MIDATA Cooperative, can only occur with the explicit consent of the account holders. Without such consent, the data cannot be decrypted. *For Digi.me (MUMC, Maastro):* The aim of the MedMij Agreement System is that anyone who wishes can have access to a Personal Health Environment (PHE) in which - under your own direction - (personal) data and/or information about your health is included. In order to provide the PHE with the (personal) data and/or health information you require, agreements have been made in the MedMij Agreement System about the exchange of this data. The exchange of data between the healthcare provider and your PHE therefore takes place via parties that comply with these MedMij agreements. Pursuant to the Medical Treatment Contracts Act (WGBO), the healthcare provider is obliged to ensure that 'others' than the patient (read: you) have no information about, access to or a copy of your medical file, unless you have given permission for this, has granted. Since your PHE (and any underlying party that works according to the MedMij agreements) is a so-called 'other' (in the sense of the WGBO), you must give the healthcare provider permission for this data

Question	Answer from Data HOLDER
	exchange. This permission specifically relates to the set of (personal) data and health information that, at your request, is exchanged by the healthcare provider - in accordance with the agreements in the MedMij Agreement System - with your PHE.

3.3. Technical Architecture

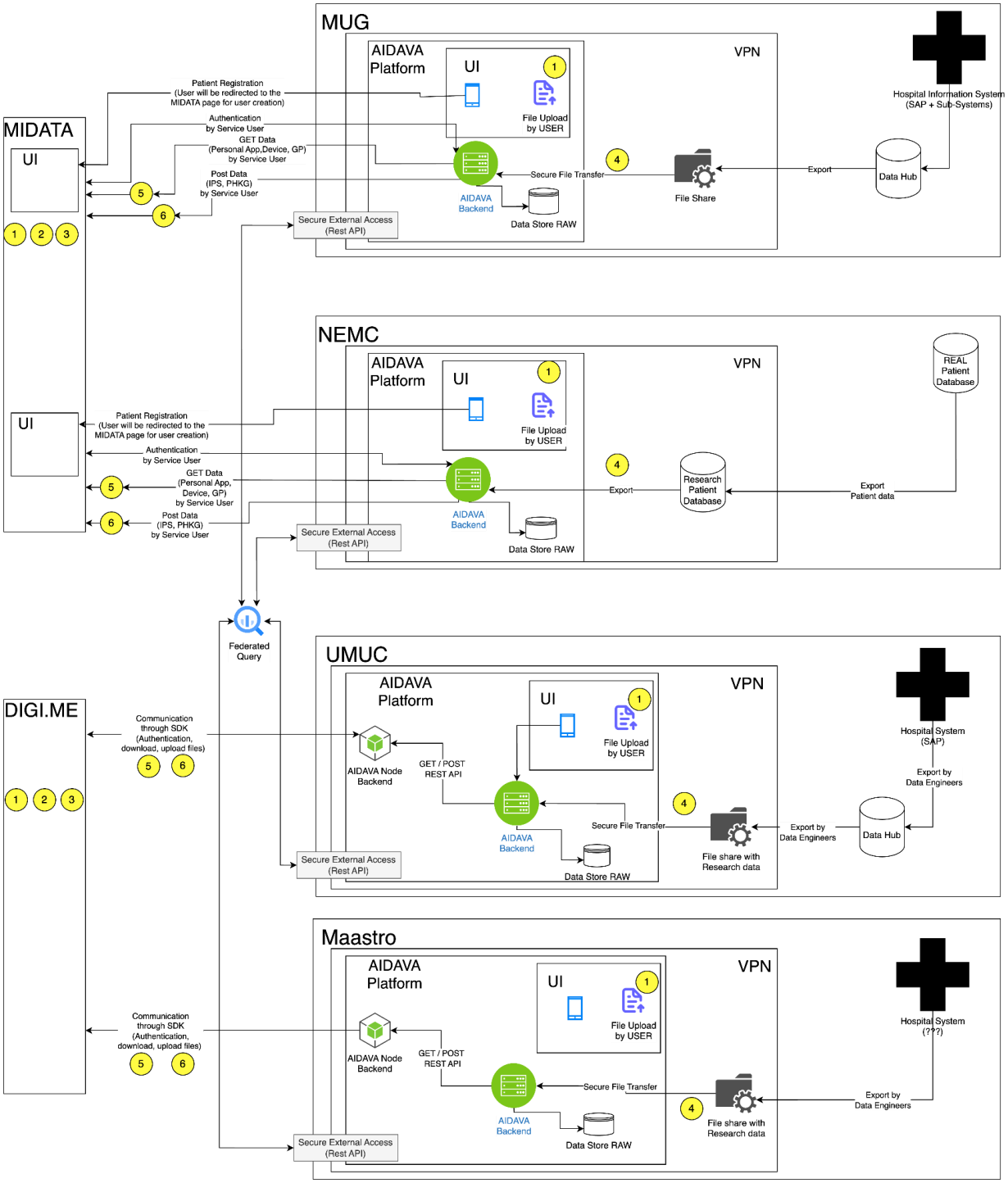
The figure below provides the technical architecture of integration across all sites involved in the project.

(include figure enclosed when submitting - please check for final version)
https://app.diagrams.net/#G1jpolnez6NqWTvtHM45Z1s9XSiQRcK0rI)

(daniel.dallos@gnd.ro)

Question	MUG	NEMC	UM	Maastro
VPN usage by the Patient • Patient will use the their or provided devices from home, that means the AIDAVA system has to be available from outside of the hospital or clinic environment				
Will the patients have VPN connection on these phones to access AIDAVA or not?	No, The WebApp will be deployed on-premise but accessible from outside	We will get an outside service provider to host AIDAVA and provide access to patients.	Will be agreed in next phase of project (task 1.5)	Will be agreed in next phase of project (task 1.5)

First draft of technical architecture:


MUG
AIDAVA Platform
UI
VPN
Hospital Information System (SAP + Sub-Systems)
Patient Registration (User will be redirected to the MIDATA page for user creation)
Authentication by Service User
GET Data (Personal App,Device, GP) by Service User
Post Data (IPS, PHKG) by Service User
MIDATA
UI
File Upload by USER
Secure File Transfer
File Share
Export
Data Hub
AIDAVA Backend
Data Store RAW
Secure External Access (Rest API)
NEMC
REAL Patient Database
Research Patient Database
Export Patient data
Export
Federated Query
UMUC
DIGI.ME
Communication through SDK (Authentication, download, upload files)
AIDAVA Node Backend
GET / POST REST API
Hospital System (SAP)
Export by Data Engineers
File share with Research data
Maastro
Hospital System (???)

ANNEXES

AIDAVA GRANT AGREEMENT ARTICLE 13 — CONFIDENTIALITY AND SECURITY

13.1 Sensitive information

The parties must keep confidential any data, documents or other material (in any form) that is identified as sensitive in writing ('sensitive information') — during the implementation of the action and for at least until the time-limit set out in the Data Sheet (see Point 6).

If a beneficiary requests, the granting authority may agree to keep such information confidential for a longer period.

Unless otherwise agreed between the parties, they may use sensitive information only to implement the Agreement.

The beneficiaries may disclose sensitive information to their personnel or other participants involved in the action only if they:

(a) need to know it in order to implement the Agreement and

(b) are bound by an obligation of confidentiality.

The granting authority may disclose sensitive information to its staff and to other EU institutions and bodies.

It may moreover disclose sensitive information to third parties, if:

(a) this is necessary to implement the Agreement or safeguard the EU financial interests and

(b) the recipients of the information are bound by an obligation of confidentiality.

The confidentiality obligations no longer apply if:

(a) the disclosing party agrees to release the other party

(b) the information becomes publicly available, without breaching any confidentiality obligation

(c) the disclosure of the sensitive information is required by EU, international or national law.

Specific confidentiality rules (if any) are set out in Annex 5.

13.2 Classified information

The parties must handle classified information in accordance with the applicable EU, international or national law on classified information (in particular, Decision 2015/44414 and its implementing rules).

Deliverables which contain classified information must be submitted according to special procedures agreed with the granting authority.

Action tasks involving classified information may be subcontracted only after explicit approval (in writing) from the granting authority.

Classified information may not be disclosed to any third party (including participants involved in the action implementation) without prior explicit written approval from the granting authority.

Specific security rules (if any) are set out in Annex 5.

13.3 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (seeArticle 28).

Such breaches may also lead to other measures described in Chapter 5.

MIDATA platform: Main safety and security features

- The platform is hosted in Switzerland, using the Swisscom Dynamic Computing Services (DCS) infrastructure, in Swisscom data centers. DCS is certified according to ISO 20000, 27001, 9001, 14001, ISAE 3402, GDPR (data processor). The database is stored in a replica set with 3 replicates.
- The database is backed up using the Swisscom Dynamic Storage service, using georedundant mirroring in several Swisscom data centers.
- Yearly external security audits.
- Data are encrypted using multi-level encryption, allowing granular sharing of specific data sets.
- Account holders may encrypt their access key, precluding server-side data breaches.
- Two factor authentication available.
- Data traffic from and to server via HTTPS, TLS 1.3. Connection and all safety tokens using perfect-forward secrecy.
- Traceability of data processing of sensitive data is ensured: Entries are created in the audit log for each security-critical action. The author is saved for each data record and for each update of a data record. The change history for each data record is available.
- Data sovereignty of the account holders over the data sets in their data accounts is technically ensured and enshrined in the Articles of Association (https://www.midata.coop/wp-content/uploads/2019/08/MIDATA_Statuten_20190626_EN.pdf) and the Privacy Policy (https://ch.midata.coop/#/public/terms?which=midata-privacy-policy); use of the data sets available in the encrypted data accounts by third parties, including MIDATA Cooperative, can only occur with the explicit consent of the account holders.

"MedMij Framework." 2021. MedMij. December 22, 2021. https://medmij.nl/en/medmij-framework/.