Ref. Ares(2021)7044077 - 23/11/2021

945345 – H2O

Health Outcomes Observatory

WP2 – Technical Infrastructure and Interoperability

D2.2 Architecture Blueprint

Lead contributor	Fabian Prasser (2-Charité – Universitätsmedizin Berlin)
	fabian.prasser@charite.de
Other contributors	Kees van Bochove (9-The Hyve)
	Joris Borgdorff (9-The Hyve)
	Alexey Khlestov (9-The Hyve)
	Julia Kurps (9-The Hyve)
	Felix Wirth (2-Charité)
	Diogo Telmo Neves (2-Charité)
	Matthias Rose (2-Charité)
	Dimitar Toshev (20-Roche)
	Karel De Witte (20-Roche)
	Thomas Metcalfe (20-Roche)
	Georg Duftschmid (1-Medizinische Universität Wien)
Reviewers	Antonella Chiucchiuini (14-Takeda)
	Tanja Stamm (1-Medizinische Universität Wien)
	Peter Singleton (8-European Institute for Innovation through Health Data)
	Antie Hottgenroth (17-Lilly)
	Roseann Chu (14-Takeda)

Document History

Version	Date	Description
V0.3	31/05/2021	First Draft
V0.6	15/08/2021	Comments
V0.9	31/08/2021	Draft
V1.0	30/09/2021	Final Version

Table of Contents

Definitions

- Participants of the H2O Consortium are referred to herein according to the following codes:
  1. MUW — Medizinische Universität Wien
  2. Charité — Charité - Universitätsmedizin Berlin
  3. EMC — Erasmus University Medical Center
  4. ICS-HUVH — Institut Català de la Salut - Hospital Universitari Vall d'Hebron
  5. KCL — King's College Hospital
  6. KUL — Katholieke Universiteit Leuven
  7. EPF — European Patients' Forum
  8. I-HD — The European Institute for Innovation through health data
  9. The Hyve — The Hyve
  10. TEAMIT — Teamit Research
  11. KUH — Karolinska University Hospital
  12. UniSR — Vita-Salute San Raffaele University
  13. IKNL — Netherlands Comprehensive Cancer Organisation
  14. TAKEDA — Takeda Pharmaceuticals International AG
  15. NVS — Novartis Pharma AG
  16. ABBVIE — Abbvie INC
  17. Lilly — Eli Lilly and Company Limited
  18. MDT — Medtronic International Trading SARL
  19. Pfizer — Pfizer Limited
  20. ROCHE — F. Hoffmann-La Roche Ltd
  21. SARD — Sanofi-Aventis Recherche & Developpement
  22. JDRF — JDRF International (AP)
  23. Trial Nation — Trial Nation (AP)

- Grant Agreement. (Including its annexes and any amendments) The agreement signed between the beneficiaries of the action and the IMI2 JU for the undertaking of the H2O project (Grant Agreement No. 945345).
- Project. The sum of all activities carried out in the framework of the Grant Agreement.
- Consortium. The H2O Consortium, comprising the above-mentioned legal entities.
- Consortium Agreement. Agreement concluded amongst H2O participants for the implementation of the Grant Agreement. The agreement shall not affect the parties' obligations to the Community and/or to one another arising from the Grant Agreement.

Abbreviations

Acronym / Abbreviation	Meaning
API	Application Programming Interface
app	Application
CDM	Common Data Model
EHR	Electronic Health Record
ePRO	Electronic Patient Reported Outcome
ETL	Extract, Transform, and Load
FHIR	Fast Healthcare Interoperability Resources
GB	Gigabyte
GDPR	General Data Protection Regulation
GP	General Practitioner
H2O	Health Outcomes Observatory
HL7	Health Level Seven International
ID	Identity
IMI	Innovative Medicines Initiative
LDAP	Lightweight Directory Access Protocol
LOINC	Logical Observation Identifiers Names and Codes
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
PRO	Patient Reported Outcome
QR	Quick Response (code)
RADAR-base	Remote Assessment of Disease And Relapses
SNOMED CT	Systemised Nomenclature of Medicine – Clinical Terms
TB	Terabyte
vCPUs	Virtual Central(ized) Processing Units
WP	Work Package

Preamble

The Health Outcomes Observatory (H2O) project (https://health-outcomes-observatory.eu/) project is designed to enable a value-based approach in healthcare systems, improving their sustainability by helping them optimize care delivery and the use of their resources around outcomes that matter to patients. Patients will be provided with digital tools which will allow them to measure their symptoms in a standardized manner, integrate their patient generated data with clinically generated outcomes data and communicate with their health professionals and other healthcare providers and align on the right course of action. National observatories will be set up in four European countries (Austria, Germany, the Netherlands and Spain), initially focused on diabetes, inflammatory bowel disease and oncology. These will provide integrated information on patient outcomes for patients and their healthcare providers for use in clinical care. H2O aims to eventually launch observatories throughout Europe, covering many more disease categories and countries.

The main aim of Work Package 2 (WP2) of H2O is the design and implementation of the technical architecture of the H2O project, to support data management according to the project vision. This document serves to capture the current planning status, to communicate it within the consortium and with the public and to coordinate it with the parties involved.

1. Introduction

When designing the architecture specification, there was a need to decide which data (i.e., patient-reported data, including actively captured data e.g. from questionnaires as well as passively collected data e.g. from sensors, clinical outcomes data, further clinical data) will be stored at which level (i.e. healthcare providers, national observatories) and how data flows between them. Notably, choices along those axes influence important aspects, such as (a) complexity and costs of building and maintaining the infrastructure locally and on the national levels, and (b) supported functionalities. WP2 believes that the architecture described in this document provides optimal trade-offs, e.g., between maximum functionality at minimum complexity and costs as well as adequate privacy protection. The solution described has been derived by evaluating the complete spectrum of options regarding data types and which level they are managed on, as well as relevant functional components (i.e., tools for patients, analytical processing, and individual-level access by healthcare professionals).

It is planned to deploy an initial version of the architecture in a relatively short time span. Therefore, following the approach laid out in the proposal, we suggest an iterative development process and describe the first two phases of the architecture in this document. In the first phase, we will utilize mostly existing technologies and solutions to create a minimum viable product, which will be extended with more state-of-the-art functionalities and dashboards in the second phase.

2. Prerequisites

In the remainder of this document, we will refer to different important terms and consider specific scopes, which are explained below.

2.1. Scope

- This document focuses on the description of a high-level architecture with important processes, systems, and data flows only. Certain specifics will be worked out during implementation and the further course of the project.
- Only some information security and data protection aspects are covered in this document. Detailed information security and data protection concepts, including details on confidentiality, integrity, and availability, such as backups, as well as a detailed data governance framework, will be developed during the further course of the project together with WP1.
- The data governance framework to be developed in the future will also describe who is in charge of data quality assurance on different levels. Automated quality checks are supported by the architecture described in this document.

2.2. Stakeholders

- Stakeholders with need to access aggregate data: patient organisations, healthcare payers and providers, regulatory authorities, scientists, and industry.
- Stakeholders with additional need to access individual-level data: patients and healthcare professionals.

2.3. Assumptions

- Conceptually, patient-reported data will always be stored centrally on the national level (physically it may be distributed, e.g., across different data centers as a measure to ensure availability).
- A central healthcare professional-facing interface is also required to allow outpatient clinics and private practices to join the observatories (particularly important for Diabetes).
- Patient-reported data can be entered through patients' own devices as well as devices temporarily made available in healthcare settings, e.g., in a GP's waiting room.

2.4. Types of Data

- **User Master Data (patients, healthcare professionals, etc.):** This data, which directly identifies individuals, is particularly relevant for identity management, record linkage and onboarding processes.
- **H2O-internal System Metadata:** This data will be required at various levels. This architecture specification will not cover any details of these implementation aspects.
- **Patient-generated Data:** This data will be provided by patients through PRO apps and potentially by further means (e.g., sensor data collected through wearables).

- **Clinical Data:** This data includes healthcare professional-generated outcomes and further clinical data relevant to H2O, such as information on therapy or socio-demographic properties.

We note that this document describes the first two versions of the H2O architecture, which will be continuously and iteratively extended to capture more aspects of the project. This may also lead to the addition of new types of data, e.g., for data exchanged with registries.

2.5. Levels of Identifiability of Data

- **Identified Data:** This includes individual-level personal data managed in H2O including directly identifying attributes, such as names. Important examples include data displayed in the healthcare providers dashboards or patient apps.

- **Pseudonymized Data:** This includes individual-level personal data managed in H2O in a form where directly identifying attributes, such as insurance numbers, names, and addresses, have been removed and replaced by a surrogate identifier (a pseudonym).

- **Anonymous or Aggregated Data:** This level includes data that has been aggregated across a group of individuals or modified in a way that it is not considered personal data under the GDPR anymore. An alternative approach may be to use data synthetization as an anonymization process, by generating artificial data derived from personal data that captures important statistical properties. These types of data are important on the observatory level, to generate scientific insights, stakeholder reports or aggregates for patients to compare with.

Obviously, these different levels of data are associated with different degrees of identifiability and require different degrees of protection. As mentioned in the Section “Prerequisites”, these aspects will be covered by detailed information security and data protection concepts, which will be developed together with WP1 in the future.

3. High-level Overview

The infrastructure of H2O will consist of three layers. First, a European Observatory for conducting research studies throughout Europe. Second, National Observatories for each of the participating countries in the H2O project. Third, local infrastructure deployed by each of the participating healthcare providers.

The European Observatory is used to coordinate aggregated research queries to federated data sets hosted at the care providers and may potentially store aggregated, anonymous, or synthesized data and metadata for further dissemination.

Each National Observatory has a user/patient registry[1] and a healthcare professional/healthcare provider registry in order to implement identity management and record linkage on a national level and to link individuals with their care provider. The National Observatory will store patient-

[1] The user/patient registry is a database for person identities and should not be understood as a disease registry.

reported data on individual level, selected clinical data entered through the healthcare professional portal and at a later stage also individual-level outcomes data and further clinical data pushed by care providers. Requests for data analyses, which might also be based on additional data provided by the participating sites, e.g., on controls, are coordinated at the National Observatory and executed in a federated manner, where the analysis comes to the data and not the data to the analysis. As a result, only highly aggregated and hence anonymous results of analyses will be stored at the national level. The National Observatory will also be the point of entry for patients and healthcare professionals, where they can log in and view their personal data and obtain access to aggregated data.

Finally, each of the local healthcare providers will integrate some clinical data from their EHR with H2O, initially for analytical purposes on an aggregated level. ePRO data from the National Observatory will be merged into this data set for analytical purposes. Linkage will be made possible by enabling healthcare providers to store pseudonyms for their patients together with patient master data at the national level in the user registry. At a later stage, a healthcare provider may also upload clinical data on individual-level to a National Observatory, e.g., for display in a dashboard together with PRO data. The healthcare provider will need to maintain a mapping between the patient identifier and their own patient records.

Not all healthcare providers participating in H2O will have the resources or the capacity to run the local infrastructure themselves. Therefore, the National Observatory may host a multi-tenant infrastructure replicating the setup expected from the care provider level on their behalf. This should be a temporary provisioning until the care providers have organized in a way that they can host that infrastructure themselves.

4. Building Blocks

The following sections describe the basic building blocks of the first two versions of the H2O architecture. We note that some information, in particular in sections 4.1 (Identity Management), 4.2 (User Onboarding) and 4.3 (Consent Management, Pseudonymization and Data Linkage) is high-level and that details will need to be worked out iteratively also in pilot installations, e.g., regarding how to exactly manage multiple identities from across providers, patient-held devices and how consent choices may affect this.

4.1. Identity Management

Most importantly, H2O will need to manage the identities of patients and healthcare professionals. In addition, authentication and authorization will also be needed to control access to resources for stakeholders using aggregated data, such as reports, provided by the observatories. These user-facing parts of the observatories will be implemented in later stages of H2O and are hence not covered here (see Section “Future Work” for further details).

To enable patients and healthcare professionals to connect to the observatories, we suggest implementing a two-stage identity management solution, which basically consists of two types of services that can be separated from each other to implement pseudonymization, if needed in a certain observatory due to national policies. Both services are shown in Figure 1. They will manage the following types of entities:

- **Patient User:** Typically, a patient, identified by typical identifying data and equipped with a H2O-specific pseudonym. Users can have different identities when they are patients at different healthcare providers. This is modelled via a patient profile (see below).
- **Healthcare Professional User:** Typically, a healthcare professional, also identified by typical parameters.
- **Patient Profile:** Links a user (i.e., a patient) to a certain service provider. Might reference additional information on consent given for certain types of processing. A robust and practical process for reliably linking the patient's ePRO identity to the provider's identity needs to be developed as part of the project.
- **Service Provider:** Represents a healthcare provider and links providers with institutional users and patients.

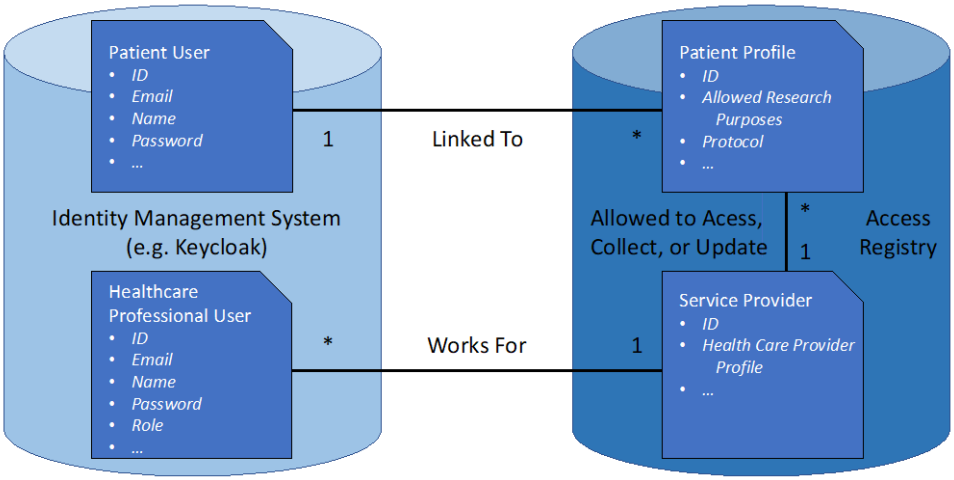

Figure 1: Basic identity management.

Information on patients (general users) and institutional users will be managed by another system than patient profiles and data on service providers. These two systems can also be hosted by different entities to have a clear separation of concerns and pseudonymization by-design and default:

- **An Identity Management System:** To create a user account and log into the registry. This should be a well-tested public solution, preferably with the option to link to other identity directories and authentication providers, such as Google Identity, self-sovereign identity providers (e.g., Sovrin, BrightID, IDX), a national login (e.g., DigiD in the Netherlands) or institutional LDAP. Following our iterative approach, we will start with a simple separate identity management system and aim for integration with national systems later, as the hurdles are quite high today.
- **An Access Registry:** To determine what resources a user has access to. This is administered using role-based access controls. For example, a service provider admin account will be able to add healthcare professionals to a service provider but not necessarily be able to view any data themselves. Users themselves will be in charge of

linking patient profiles and approving or denying certain processing activities by certain healthcare providers or for certain research purposes via the patient profile.

4.2. User Onboarding

Users can be onboarded to a national observatory's services in two different ways. The first way consists of self-registration by users who want to use the apps provided by H2O for outcomes documentation, before having been in contact with a participating institution. The second option will enable healthcare professionals involved in one of H2O's use cases to make patients aware of the services provided. In the first case, after self-registration, a token is generated (i.e., a unique, error-correcting identification number, which is easy to read or spell) that can be shown to healthcare professionals, for example, in the form of a short alphanumeric code or a barcode. By entering this code through the healthcare professional portal, linkage to a participating institution can be performed and further features as well as options to use and share the data collected will be enabled. In the second case, a shadow patient profile and user account can be generated by healthcare professionals through the dashboard, again generating a token, but in this case for the patient. In the patient app, this token can then be used to complete the account and further capture and share patient-generated data. It will also be possible to use H2O's survey app on devices, such as tablets onsite in the healthcare provider or a GP's waiting room. For this purpose, it will be possible to log in such local devices momentarily, linking it to a patient's account in H2O and log it out when they are done.

4.3. Consent Management, Pseudonymization and Data Linkage

For managing patient consent, we will implement a two-stage process for (1) enabling patients to accept the terms and conditions of using the H2O ePRO app, also for healthcare purposes and communication with healthcare professionals, and (2) enabling patients to accept or decline data use for research processes. We expect the second process to be more heavyweight and complex to design, due to national and local research consenting policies. The details will be worked out in close cooperation between WP1 (Data Management Plan, Consent, Information Governance Framework) and WP 2 especially regarding the scope for data management and data sharing (which could potentially slightly vary between observatories and between studies).

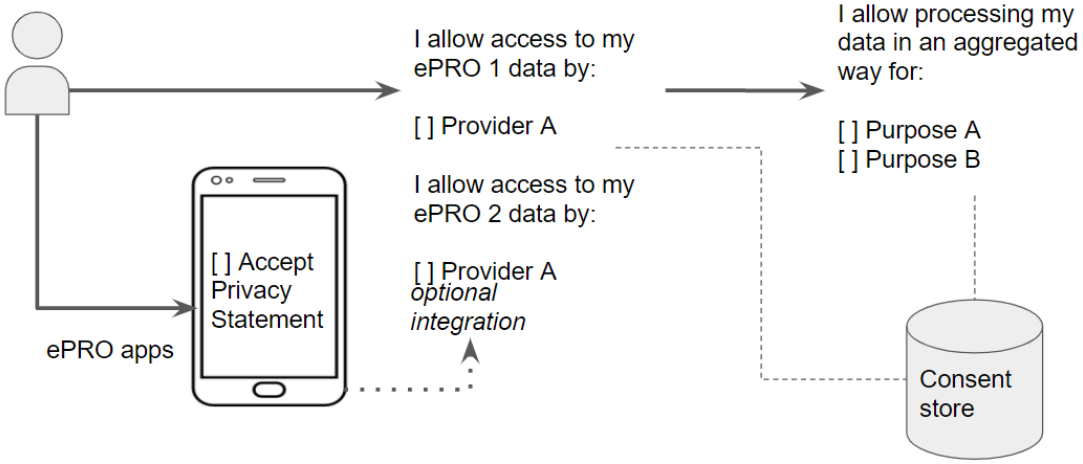

Figure 2: Basic consenting process.

An example, reflecting both types of consenting processes, is illustrated in Figure 2. As can be seen in the figure, we strive for a dynamic consenting process in which patients can control the usage of their data in a fine-grained manner. To connect an ePRO app with a patient record, a consent store is needed as part of the user registry component. The following flows will need to be enabled:

1. Patient User Self-registration

Users may register themselves either on the H2O website or in the H2O ePRO app. Before any data can be collected and before the registration can be completed, the user must agree with the privacy policy, the terms, and the conditions for collecting ePRO data in the platform. They may optionally allow processing of their data for creating ePRO statistics. Data collection may start immediately. Of course, users will also be able to unregister, which is mandatory in their GDPR subject rights, and the respective processes will be further elaborated in the according documents (data protection, information security).

2. Linking a Patient to a Healthcare Provider

A healthcare provider may log into the user registry to register a new patient. The registration should contain at least a pseudonym for the ID of the patient record in their own systems, but it may also contain full name and birth date to validate the patient record. Once the patient record is created, the user can create a token for the patient. That token should be provided to the patient in a secure manner as determined by healthcare provider policy (e.g., face-to-face, a postal letter, or a secure messaging facility). The patient can then enter that token either on the H2O website or via the H2O ePRO app, for example by scanning a QR code. When the patient enters this token, consent is requested for sharing data with the healthcare provider. This consent request will include at least that the healthcare provider may read the patient's personal ePRO data. In the second version (see below) this will also include permission for the healthcare provider to upload the patient's clinical outcomes data to the H2O national observatory for use in a dashboard.

User registration may occur before or after linking the app to a healthcare provider or it may not occur altogether. As long as the patient profile is not linked to the user, it will not be possible to link data from different healthcare providers to the same person. No clinical outcomes data will be uploaded to H2O for patient profiles that are not linked to a user.

3. Verifying the Identity of a Patient

The healthcare provider may want to verify that the user that is linked to the healthcare provider is indeed the patient in question. For this purpose, the healthcare provider should have a face-to-face or at least a direct contact with the patient. The patient navigates to the H2O website or the H2O ePRO app. The patient then starts an action to finish linking to a provider, which will then show a short-lived token. The healthcare provider logs in to the user registry, to the patient record, and starts the verification process. The healthcare professional verifies the identity of the user and then enter the short-lived token provided by the patient.

4. Linking an External ePRO Application

A user can go to the H2O website and register. The user can create a token there to link their account to an external ePRO application. The external ePRO application must in some way accept this token to be able to send ePRO data regarding that user. Two methods could be envisioned here:

1. The external ePRO app has a dedicated action to link to H2O.
2. The external ePRO app has an API that allows the data of a patient to be linked to an external system. This API is then automatically called.

4.4. Data Harmonization, Standardization and Analytics

A national observatory, which servers the global purpose of data analysis and knowledge generation as well as the local purpose of sharing data between patients and healthcare professionals, must meet different data presentation objectives. For this reason, we plan for the observatories to include two parallel data presentations, one for clinical purposes (Section 4.4.2) and one for analytical purposes (Section 4.4.1). It is planned to start without a central repository for clinical purposes in the first version (see Section 5).

4.4.1. OMOP/OHDSI

The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) aims to standardize observational healthcare data for analytical purposes. The OMOP CDM is supported by the OHDSI (Observational Health Data Sciences and Informatics) community, which seeks to collectively generate real-world evidence that promotes better health decisions and better care. The community offers (1) an extensive standardized vocabulary for health data, (2) tools and methods to transform data into the OMOP CDM and analyse it subsequently and (3) a global research network enabling cross-country and cross-institutional data analyses. In H2O, OMOP/OHDSI will be used for aggregated analytics across healthcare providers and observatories due to its high degree of standardization, which significantly accelerates studies[2].

4.4.2. FHIR

Fast Healthcare Interoperability Resources (FHIR) is a framework for the exchange of healthcare data developed by the Health Level Seven International (HL7) organization. FHIR defines approximately 120 health data objects, called resources (patient, practitioner, encounter, …), as well as specifications, how the resources can be exchanged. A well-defined maturity model describes the state of development of resources. The resources are described as a common platform or foundation, which means that each real-world implementation of FHIR requires a profiling process, in which rules about the actual usage of resources and terminologies are defined. In H2O, FHIR will be used (together with further terminology standards) to manage transactional data and to represent individual-level clinical and outcomes data in an interoperable manner. We note that HL7 FHIR is traditionally a framework for designing interoperability standards (or profiles) that can be used for communicating clinical information

[2] European Medicines Agency (EMA/614680/2018) - A Common Data Model for Europe? – Why? Which? How? Workshop report, 2018.

between repositories, and potentially between apps and repositories. While it was not originally designed to serve as a clinical data repository, increasingly more solutions that implement this have become available. We expect to be able to identify FHIR-compatible clinical data repositories offering all required features (such as audit trails, logs, rights, and roles management) and, as most hospitals are currently establishing or plan to establish FHIR infrastructures, this will likely reduce efforts for ingesting data into H2O significantly in the future.

4.5. Data Collection and Retrieval

4.5.1. RADAR-base

RADAR-base (Remote Assessment of Disease And Relapses) is an open source platform to collect patient reported data, which was also developed as part of an IMI project. RADAR-base supports two means of data collection: (1) Passive data collection using sensors measuring patient data automatically (e.g., movement, location); (2) Active data collection by means of questionnaires. RADAR-base can be administered by a management portal and data can be visualized in a dashboard. In addition, RADAR-base allows the export of structured data to researchers. In H2O, RADAR-base will be used as a project-specific app for collecting electronic patient-reported outcomes (ePROs). In the further course of the project, we aim to also create interfaces to relevant existing apps.

4.5.2. Patient and Healthcare Professional Dashboards

The data collected and managed at a national observatory will be provided to patients and healthcare professionals through a dedicated web-based dashboard (e.g., providing patient info in relation to local, national, and international averages/norms). From a technology perspective, the same application will be used for both types of users, but different types of data will be displayed depending on whether a user is associated with a patient or an institutional role. The data that is made accessible through the dashboards will be increased iteratively in the different versions of the overall architecture and specified together with the disease work packages. Moreover, the healthcare professional dashboard will also feature data entry forms that can be used by healthcare professionals using EHR systems that cannot be technically connected to the national observatories, e.g., GPs using local medical practice management software. It is expected that the data entry forms will be simple, capturing a relevant core set of outcomes-related clinical data. A risk with these dashboards is that it will cross the line of becoming a clinical decision support tool falling under the Medical Device Regulation. In that case the platform will need to comply with the relevant regulations, which we expect to be able to work around by paying careful attention to the features implemented.

5. Architecture Versions

To be able to provide initial versions of the H2O architecture quickly, the development process will be conducted in two iterations each offering different levels of services. The specifics of these versions of the architecture are described below.

5.1. Iteration 1: Collection of patient-generated data on the national level combined with federated analytics of outcomes-related clinical data stored locally

In this initial iteration, ePRO will be stored on individual-level at the national observatories, while outcomes-related clinical data is stored at the local healthcare provider level only (note that we currently do not expect healthcare providers to perform additional data entry through the dashboards). No central repository for clinical purposes is foreseen in the first version, although we might need to provide limited solutions enabling GPs to document some data through the healthcare professional dashboard. Analytics is supported in a federated manner on a combination of clinical outcomes-related data and patient-generated data. Data entered via the entry forms at a national observatory will also be considered, by pulling it into the national OMOP instance. The major technical components required on the national level include: (1) identity and access management, (2) dashboards for healthcare professionals and patients, (3) RADAR-base for ePROs, (4) OMOP/OHDSI implementation. The major technical components required on the local healthcare provider level include: (1) OMOP/OHDSI implementation, (2) interface component to pull patient-generated data from the national level into the local OMOP/OHDSI instance. An overview of these components is illustrated in Figure 3. Data collected in RADAR-base will in parallel be accessed by the dashboards and integrated into the local OMOP/OHDSI implementations.

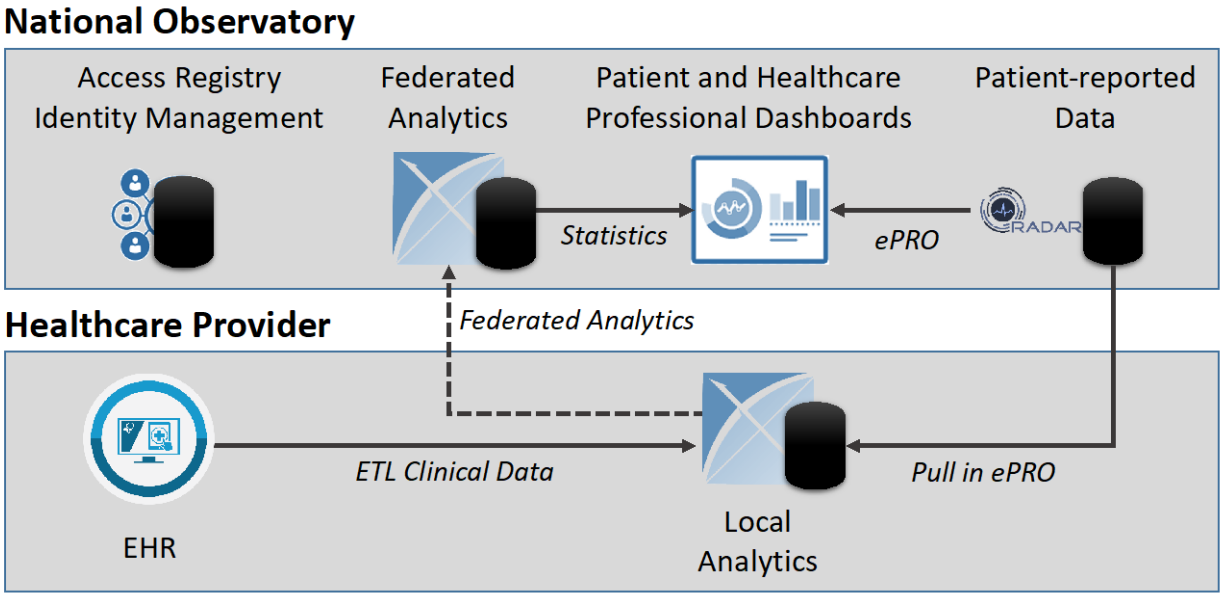

Figure 3: Architecture in the first iteration.

It is yet to be decided whether the healthcare professional dashboard will use a FHIR server for storing data entered through the dashboard. However, as H2O aims for interoperability on all layers of the architecture, this is the primary choice at the moment. ePRO data will be stored in the RADAR-base backend. The most important data flows are also depicted in Figure 3. Most development efforts at the local level will include mapping of clinical data and ePRO data pulled in from the national level to OMOP/OHDSI. We note that ePRO data can already be mapped to the OMOP data model on the national level but will be pulled into the local implementations. In the latter context, a transformation procedure will be provided from H2O that will need to be

executed locally to reduce efforts. The transformation must adhere to each user's consent settings in the H2O Access Registry, which will be checked automatically.

5.1.1. Use Cases to Enable

Major use cases that will be enabled by the first version of the architecture are listed in Table 1.

Table 1: Main use cases enabled by the first version of the architecture.

Role	Use Case
Healthcare Professional	Read ePRO data on dashboard using central authentication
Healthcare Professional	Read statistics on dashboard using central authentication
Healthcare Professional	Enter clinical data into own EHR, which will become available to H2O
Healthcare Professional	Enter and read selected clinical data into/from the central repository
Patient	Collect ePRO data without having contact to a healthcare provider
Patient	Read ePRO data on dashboard using central authentication
Patient	Read statistics on dashboard using central authentication
Patient	Grant/deny access to ePRO data to healthcare professional
Patient	Grant/deny access to data for cohort analysis
Healthcare provider admin	Grant/deny read/write roles to healthcare professionals
Researcher	Create cohort analysis on example data or anonymous open data
Researcher	Submit cohort analysis for running on H2O OHDSI installations
Observatory admin	Grant/deny cohort analysis
Healthcare provider admin	Grant/deny cohort analysis

As can be seen, the first iterations trades development complexity on the local and national level off against extended functionalities for patients and healthcare professionals. Most importantly, the dashboards will focus on ePRO data and not provide an integrated view of and control over ePRO data together with outcomes-relevant clinical data.

5.2. Iteration 2: Collection of patient-generated data and selected outcomes-related clinical data on the national level combined with federated analytics of additional outcomes-related clinical data stored locally

The second version of the H2O architecture, which is illustrated in Figure 4, will extend the first version with central storage of outcomes-related clinical data. This will enable H2O to display ePRO data as well as clinical data in the healthcare professional and patient dashboards. For this purpose, FHIR repositories will be installed at the national level (if not already installed during the first development cycle) and required on the local level. It is noteworthy that this only applies

to outcomes-related data strictly necessary in the dashboard. Data which is only needed for some analyses can remain at the healthcare provider level. The design decision to push selected clinical data from the healthcare provider to the national level instead of dynamically pulling it in is driven by the experience that dynamic access by external parties to resources stored at the healthcare provider level is typically not compliant with the healthcare providers' information security policies. Analyses will be performed in a federated manner based on OMOP/OHDSI in the second version as well.

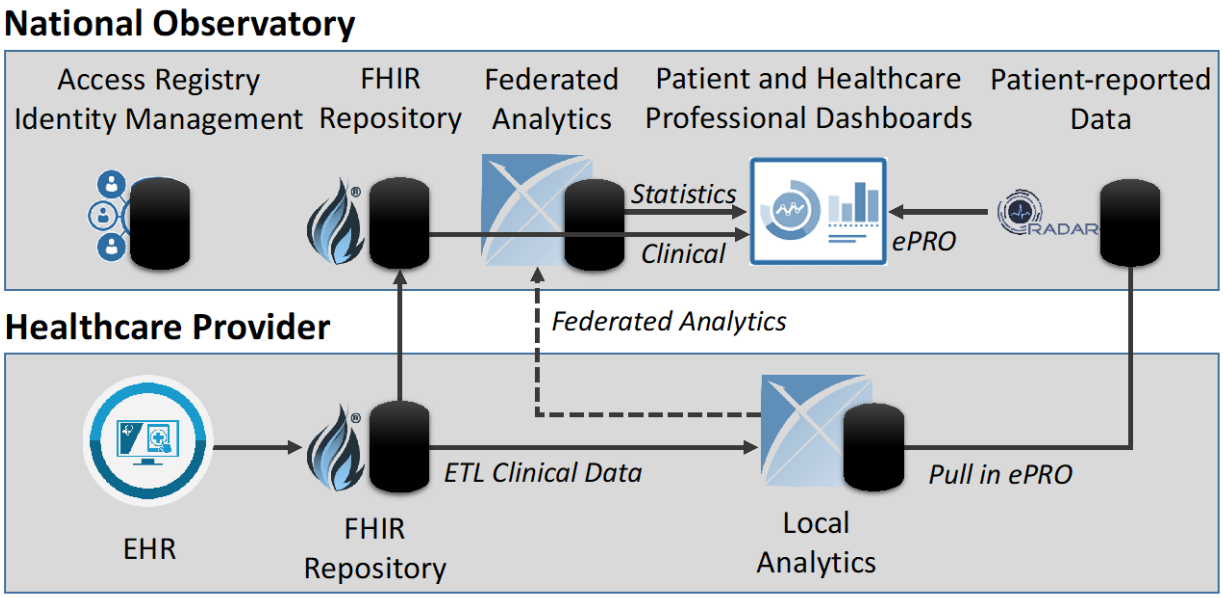

Figure 4: Architecture extensions in the second iteration.

The most important addition to the first iteration is the integration and central storage of individual-level data in an interoperable form, based on a common dataset specified with FHIR profiles and using international terminologies and ontologies like SNOMED CT and LOINC. This will introduce new efforts, in particular at the local healthcare provider level, where either a FHIR server needs to be deployed or H2O-relevant data will need to be integrated into an existing FHIR server. We note that healthcare providers are free to either map data to FHIR and then map the FHIR representation to the OMOP CDM, the other way around or to implement both mappings independently. Local sites should strive for maximum data quality in both repositories, though.

Finally, the storage of individual-level data on a broader scale at the national level may increase information security and data protection requirements to deal with concerns from regulators, patients, and privacy defenders. Also, a more stringent data governance framework will be required to make sure that there is no misuse of data beyond the scope of H2O goals. Also, hurdles for participating might be higher on the local level, as healthcare providers cede control over a subset of the data they generated.

5.2.1. Use Cases to Enable

The major use cases that will be enabled by the second version of the architecture are listed in Table 2.

Table 2: Main use cases enabled by the second version of the architecture (functionalities provided in addition to the first version are highlighted).

Role	Use Case
Healthcare Professional	Read ePRO data on dashboard using central authentication
Healthcare Professional	Read statistics on dashboard using central authentication
Healthcare Professional	Read clinical outcomes on dashboard using central authentication
Healthcare Professional	Enter clinical data into own EHR, which will become available to H2O
Healthcare Professional	Enter and read selected clinical data into/from the central repository
Patient	Collect ePRO data without having contact to a healthcare provider
Patient	Read ePRO data on dashboard using central authentication
Patient	Read statistics on dashboard using central authentication
Patient	Read clinical outcomes on dashboard using central authentication
Patient	Grant/deny access to ePRO data to healthcare professional
Patient	Grant/deny access to EHR data to a healthcare professional or provider not part of the provider that generated that EHR data
Patient	Grant/deny access to data for cohort analysis
Healthcare provider admin	Grant/deny access to EHR data to a patient
Healthcare provider admin	Grant/deny read/write roles to healthcare professionals
Researcher	Create cohort analysis on example data or anonymous open data
Researcher	Submit cohort analysis for running on H2O OHDSI installations
Observatory admin	Grant/deny cohort analysis
Healthcare provider admin	Grant/deny cohort analysis

6. Future Work

The two first iterations described in this document aim to provide the basic functionalities needed to further increase the number of use cases supported by H2O. Important topics which we plan to address in revised versions of the specification and implementation include:

- The design and development of user interfaces for specific types of users / stakeholders that want to perform analyses or access reports in a collaborative process. At the current stage of the architecture, we expect that this will be implemented with manual processes and not supported through the interfaces and portals of the platform.
- Analogously, the current architecture version doesn't support connections to the European observatory. We expect that the exchange of aggregate data and reports will be a manual process until automated connections and according user interfaces will be established in future versions.

- For some use cases and to increase acceptance by patients and healthcare professionals, we plan to also establish interfaces enabling the integration of external ePRO products and platforms.
- Another important use case is to provide the ability of getting in touch with patients for a follow-up study, for patients in a certain cohort or with a certain profile to enable enrolment in clinical trials or observational studies. A communication path between the European and National Observatories via the healthcare providers is needed to establish this. A request procedure would need to be set up. In it, the request will detail the study, its purpose and a full specification of the cohort definition. The procedure will most likely need multiple levels of approval, from the observatories, healthcare providers, and patient consents.
- In many European countries, national eHealth infrastructures and EHR systems have been or are being established. Such infrastructures can provide important functionalities relevant to H2O (e.g., identity management, provision of patient-facing apps, integration of clinical data). We plan to make appropriate integrations, but this will require significant efforts due to the heterogeneous technical and legal frameworks around these infrastructures.

7. Hardware Specification

The H2O platform for National Observatories is planned to run on a Kubernetes cluster. This makes it scalable and deployable on public cloud platforms and on-premise clouds alike. In the pilot phase of the first iteration, the following hardware specifications are needed in a national observatory. Information on availability at the partner sites is provided in D2.1 (Assessment of the Local and National Situation).

7.1. Kubernetes Cluster

The cluster should feature at least four worker nodes with 16 GB of main memory and four vCPUs. In addition to that, a shared storage pool of 1 TB is needed. A minimum of 3 nodes is needed to host RADAR-base in a redundant manner together with a national OHDSI installation and the dashboard software. Increasing the number of nodes increases the resilience and reliability of the overall system.

7.2. Backup Storage

As a backup mechanism, an AWS S3-compatible object storage is preferred with at least 1 TB capacity.

7.3. Networking

The Kubernetes cluster's Ingress node must be available from the internet and have a valid domain name record.