IMMERSE

Implementing Mobile MEntal health Recording Strategy for Europe

H2020 - 945263

D3.2	**Implementation Guide**
Dissemination level[1]	**PU**
Contractual date of delivery	30.04.2023
Actual date of delivery	30.04.2023
Type[2]	R
Work package	3
Work package lead	Thomas Ganslandt

[1] **Please choose the appropriate reference:**
PU = Public, fully open, e.g. web;
CO = Confidential, restricted under conditions set out in Model Grant Agreement;
CI = Classified, information as referred to in Commission Decision 2001/844/EC.

[2] **Use one of the following codes:**
R: Document, report (excluding the periodic and final reports)
DEM: Demonstrator, pilot, prototype, plan designs
DEC: Websites, patents filing, press & media actions, videos, etc.
OTHER: Software, technical diagram, etc

Author list

Organisation	Name	Contact information
6 - UKER	Thomas Ganslandt	Thomas.ganslandt@medma.uni-heidelberg.de
6 - UKER	Geun Hyun Kim	GeunHyun.Kim@medma.uni-heidelberg.de
2 - CIMH	Anita Schick	anita.schick@zi-mannheim.de
8 - movisens GmbH	Simon Krause	simon.krause@movisens.com

Document history

Version	Date	Description
1	**11.04.2022**	**Initial release**
2	**30.04.2023**	**Update regarding data privacy of sensor data and changes to the central research IT infrastructure architecture**

List of abbreviations

- CSV: Comma-separated values
- DMMH: Digital Mobile Mental Health
- DMP: Data management plan
- eCRF: Electronic case report form
- FAIR: Findability, Accessibility, Interoperability, and Reuse
- GPS: Global Positioning System
- ESM: Experience Sampling Method
- FHIR: Fast Healthcare Interoperability Resources
- HL7: Health-Level 7
- ID: Identifier
- IMMERSE: Implementing Mobile MEntal health Recording Strategy for Europe
- LOINC: Logical Observation Identifiers Names and Codes
- SNOMED CT: Systematized Nomenclature of Medicine, Clinical Terms
- WIFI: Wireless Fidelity
- WP: Work package

Table of contents

1. Summary....4
2. Deliverable Report: D3.2 Implementation Guide.... 4
2.1. Approach for the Implementation Guide.... 4
2.2. Architecture....5
2.3. Data protection considerations.... 9
2.4. Data Elements for Study & Participants.... 10
2.4.1. IMMERSE_ResearchSubject.... 10
2.4.2. IMMERSE_Patient.... 12
2.5. Data Elements for Questionnaires & Responses.... 13
2.5.1. IMMERSE_Questionnaire....14
2.5.2. IMMERSE_QuestionnaireResponse.... 17
2.5.3. IMMERSE_QuestionnaireResponseObservation....19
2.6. Data Elements for Mobile Sensor Data.... 21
2.6.1. IMMERSE_GeolocationObservation.... 22
2.6.2. IMMERSE_ActivityObservation....25
2.6.3. IMMERSE_StepsObservation.... 28
2.6.4. IMMERSE_DeviceOnOffObservation.... 29
2.6.5. IMMERSE_DisplayOnOffObservation....31
2.6.6. IMMERSE_AppUsageObservation....33
2.6.7. IMMERSE_NotificationObservation.... 36
3. References.... 39

1. Summary

This implementation guide describes the architecture of the components as well as the data structures for the implementation of the “Phase II” trial of the IMMERSE project. The main objective is to provide a specification for the implementation of the IT infrastructure and data capture so that it complies to data protection requirements through pseudonymized data capture, and complies to the FAIR guiding principles [1] for data sharing by using standardized open data structures for all of the data elements collected in the trial as mandated by the IMMERSE data management plan. To this end, the HL7 FHIR standard has been selected to represent all data elements captured through eCRFs, the MoMent app or the movisensXS app in a standardized, well-structured and openly available format. FHIR is a standard that has gained wide international adoption in the field of healthcare, including mobile health apps and that is made available freely by HL7 International. The FHIR standard provides powerful means of adapting data structures to individual project needs, while still ensuring compatibility to the basic standard. In addition, all data elements, attributes and value sets are semantically annotated with concepts from a suitable terminology, if available. This approach will support annotation of research datasets with rich metadata before deposition into a long-term research data archive. While originating from the area of health care provision and routine healthcare IT, FHIR is gaining increasing adoption in the research context and is positioned to allow tight integration between clinical care and research. It is thus an ideal option for representation of medical research data using an open, structured and well-annotated standard to satisfy the FAIR guiding principles.[2] This approach will ensure findability of IMMERSE datasets through rich metadata annotations based on open terminologies as well as ensure interoperability and reusability through open standards and terminologies for the data structures.

In the Architecture section, the components, their intended use and the interfaces between them are described. In the following sections, the data elements to be generated in the trial are described in detail, regarding their intended use, their relationship to other study data elements, their attributes and semantic encoding, their FHIR representation, as well as data protection and implementation aspects.

2. Deliverable Report: D3.2 Implementation Guide

2.1. Approach for the Implementation Guide

The component architecture was designed in collaboration with partners from Movisens to include all components of the overall IMMERSE platform, their technical interactions and interfaces, and manual interactions. It takes into account the need to pseudonymize data throughout the platform while being able to link datasets generated from different source components together in the research database, as well as generate individually pseudonymized extracts for research users to protect against the unwanted combination of multiple partial exports.

The data elements are defined by

- the required “backbone” elements for representing study sites and participants
- the questionnaires to represent data captured through ESM forms or eCRFs
- the mobile sensing data captured through the movisensXS app

The mobile sensing data to be used were selected with representatives from WP4 and Movisens, taking into account both data protection aspects as well as the relevance of the individual data types for analysis in the project.

The HL7 FHIR Standard (Fast Healthcare Interoperability Resources) was chosen as an open, standardized representation of the IMMERSE data elements. HL7 has been the definitive standard for data transfer in the clinical context for many decades. HL7 FHIR is a relatively new addition to the HL7 family of standards, but has already gained wide adoption in clinical care, and is being taken up for research use as well, e.g. in the German Medical Informatics Initiative. HL7 FHIR offers basic "Resources" for data elements that provide a well-structured definition of attributes, value sets and references between the resources, including options for semantic annotation. In addition, FHIR provides a powerful mechanism of "profiling" these resources to adapt them to individual project needs. A profile "inherits" the structure of its basic resource, but can be adapted e.g. by removing unnecessary attributes, changing the cardinality of attributes, setting defined terminologies or semantic annotations, or extending elements with new attributes. This approach enables the fine-tuning of FHIR profiles to individual project needs, while still ensuring compatibility to the basic resources as defined by the HL7 organization. There is a large and vibrant community of organizations that use HL7 FHIR, provide tooling for the specification and dissemination of FHIR profiles as well as the exchange and adaptation of existing profiles.

For each data element, a suitable FHIR basic resource was identified. With the goal of data minimization, profiles were created for all basic FHIR resources that removed attributes which are not captured and/or required in the IMMERSE project. Additionally, semantic annotations were defined for all attributes and value sets. These were based in most cases on standardized terminologies, which were, however, not available in some cases. In these cases, the existing codes were used as supplied by the respective source systems. It was not necessary to add extensions to the FHIR profiles, as all required attributes could be represented by the structures of the basic resources.

2.2. Architecture

The MoMent platform contains several components that both interact internally as well as with study-specific services that are implemented for scientific evaluation during the IMMERSE project. Interfaces between the components serve to establish interoperability with other systems.

The following figure schematically shows the main components of the platform, study-specific components and interfaces as well as user interactions between them:

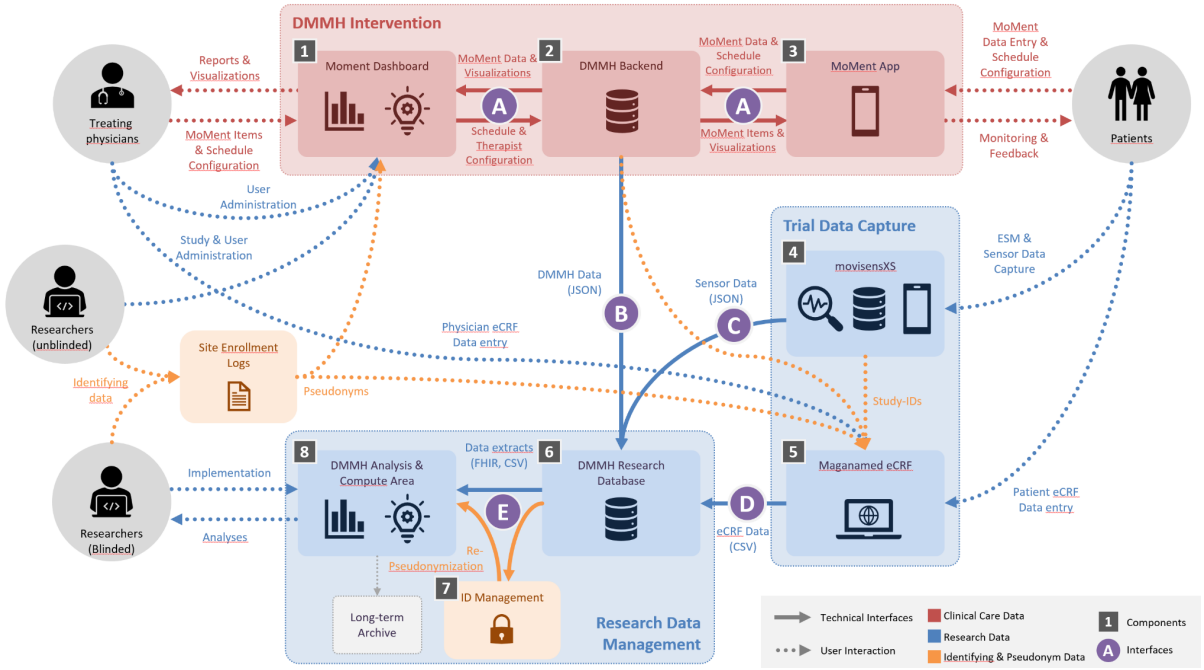

Figure 1: DMMH Platform components (boxes labeled clockwise with numbers 1 to 8), technical interfaces (arrows labeled clockwise with letters A to F) and manual user interactions (dotted arrows). DMMH Prototype components are shown in red, study-specific components for scientific use within the project are shown in blue, and the ID management component is shown in orange.

The technical components and their roles are briefly described in the following table. Numbers and letters refer back to the identifiers in Figure 1.

Component	Role
1	MoMent Dashboard The MoMent Dashboard is used by treating physicians & unblinded researchers to configure ESM questionnaires and schedules for the participants, execute user & study administration and consume reports and visualizations based on collected ESM data. Users can export summarized reports of ESM data e.g. for manual import into the electronic health record of the treating organization.
2	DMMH Backend The DMMH Backend receives configuration data from treating physicians through the Dashboard as well as ESM data and schedule changes received from participants through the MoMent App. It stores the ESM configuration and captured pseudonymous data for use within the DMMH Intervention. On Enrollment of participants, unique identifiers are generated by the Backend which are used throughout the trial period. To enable linkage of DMMH data to the other datasets collected in the study, these identifiers are manually entered into the eCRF.
3	MoMent App

	The MoMent App pings participants in the intervention arm of the trial at scheduled times and provides forms for documenting pseudonymous ESM data by the participants. It also enables schedule changes by participants. Collected data and schedule changes are communicated to the DMMH Backend for storage and provision to other components.
4	movisensXS movisensXS is a mobile sensing platform that is deployed in parallel to the MoMent App for capturing pseudonymized physiological, geolocation and/or mobile device use data. It is also used to capture ESM data at specified time points during the trial for both the intervention and control arm participants. Appropriate measures are taken to pseudonymize potentially identifying data elements (e.g. geolocation) before transfer to other components. On activation of participants mobile phones, movisensXS generates distinct sequential identifiers within each observation period. To enable linkage of movisensXS data to the other datasets collected in the study, these identifiers are manually entered into the eCRF after activation within each observation period.
5	eCRF (Maganamed) The eCRF is used for collecting pseudonymized RCT data from study participants and treating physicians through electronic case report forms.
6	DMMH Research Database The DMMH Research Database serves as a central integrated data store for data collected through the DMMH intervention, the movisensXS sensing platform and the eCRF, and making it available to data requesters according to the Data Governance Policy and decisions of the Data Governance Board. Datasets are provided with project-specific pseudonyms to prevent cross-project merging of separate datasets.
7	ID-Management System The ID-Management system is used to re-assign central participant pseudonyms to individual pseudonym domains for each data export from the DMMH research database. Use of separate pseudonyms prevents unintended collation of datasets exported for separate data use projects. The ID management system will be operated by the Trust Center implemented at Erlangen University Hospital as an independent entity within the German Medical Informatics Initiative.
8	DMMH Analysis & Compute Area The DMMH Analysis & Compute Area is used by blinded researchers to implement and execute analysis and machine learning software on datasets received from the DMMH Research Database.

The technical interfaces are described in the following table:

Interface	Requirements
A	Interfaces between the MoMent Dashboard, the DMMH Backend and the MoMent App Interface B is internal to the DMMH prototype. There are no project-specific requirements regarding the data structures or API for communication between the Dashboard, Backend and App. Data transferred over interface B is pseudonymized. The communication must be secured with adequate encryption (e.g. TLS) to ensure that data exchanges cannot be eavesdropped or modified. Authentication (e.g. through certificates) must be implemented to ensure that communication occurs only between authorized services of the platform.
B	Interface between the DMMH Backend and the Research Database Interface C is study-specific and serves to transfer DMMH data from the DMMH Backend to the Research Database. To ensure an interoperable implementation that enables sustainable re-use and valorisation beyond the project, this interface will be implemented using a REST-based API documented with an OpenAPI specification. Data will be provided through the API in JSON format and converted to HL7 FHIR resources before storage in the Research database. For each data element to be made available through the interface, appropriate FHIR profiles will be selected either for re-use from established projects or defined according to IMMERSE requirements. Relevant sources for existing FHIR profile definitions are the International Patient Summary (IPS), the European EHR Exchange Format (EHRxF) and the German Medical Informatics Initiative (MII). Data exchanged through interface C is pseudonymized. Communication between components must be protected with adequate encryption (e.g. TLS) and authentication (e.g. certificate-based) between components.
C	Interface between movisensXS and the DMMH Research Database Interface D is study-specific and serves to transfer sensor data from movisensXS to the Research Database. Data is provided by movisensXS through its integrated REST API in a CSV format with accompanying metadata in XML format. To enable sustainable re-use and valorization beyond the project, data is converted to adequate FHIR profiles on the Research Database side before storage and provisioning. For each data element to be made available through the interface, appropriate FHIR profiles will be selected either for re-use from established projects or defined according to IMMERSE requirements. Relevant sources for existing FHIR profile definitions are the International Patient Summary (IPS), the European EHR Exchange Format (EHRxF) and the German Medical Informatics Initiative (MII). Data exchanged through interface D is pseudonymized. Communication between components must be protected with adequate encryption (e.g. TLS) and authentication (e.g. certificate-based) between components.
D	Interface between the eCRF and the Research database Interface E is study-specific and serves to integrate RCT data captured through the trial eCRFs with the DMMH data captured through the DMMH App in the Research

	Database. Data will be exported from the eCRF in CSV format. For integration with the FHIR-based Research Database, data will be transformed into FHIR questionnaire/questionnaireResponse profiles before storage. Data exchanged through interface E is pseudonymized. Communication between components must be protected with adequate encryption (e.g. TLS) and authentication (e.g. certificate-based) between components.
E	Interface between the DMMH Research Database, ID-Management System and DMMH Analysis & Compute Area Interface F is study-specific and serves to provide datasets to data use projects that have been granted access to subsets of Research Database content. Data will be provided either as FHIR bundles or in CSV format with appropriate annotation. Data exchanged through interface E will be pseudonymized, with a separate set of identifiers for each data use project to prohibit merging of separately provided datasets. Communication between components must be protected with adequate encryption (e.g. TLS) and authentication (e.g. certificate-based) between components.

2.3. Data protection considerations

All data elements collected through the IMMERSE data capture platforms (eCRF, MoMent App, movisensXS) are considered person-related as they are captured and stored in pseudonymized form. The capability to link pseudonyms to participant identities is available only to the site investigators which maintain their local participant identification logs. Data provided for analysis are re-pseudonymized to preclude unintended merging of partial datasets. Data captured for IMMERSE are considered a special category of personal data according to GDPR article 9. The pseudonymized use of this data is based on the informed consent given by participants upon enrollment into the phase II trial. A full anonymization of IMMERSE data is not planned for data stored in the central research database or long term archive due to the resulting significant reduction in data quality. Anonymization may however be carried out for data exports in the context of individual data use projects.

Sensor data items captured through the movisensXS platform present a potential risk regarding unintended re-identification of patients e.g. through geotracking data that would reveal home or work locations. During creation of this implementation guide each sensor data item potentially available through movisensXS was first evaluated regarding utility for the project, resulting in a limited set of data items selected for implementation. For these data items, re-identification risks were evaluated to decide whether additional measures needed to be implemented. For re-identification to occur, data elements must either divulge directly identifying data (e.g. a home address from geolocation data) or contain pseudo-identifiers that can be linked to datasets available to an adversary (e.g. birthdate and zip code available from a public voter registration). For evaluation of re-identification risks, the potential benefit to the adversary as well as the effort required to implement the attack need to be considered. Thus, data may be factually de-identified even if re-identification is theoretically possible, but the effort to implement the attack surpasses the potential benefit to the adversary.

De-identification measures described in this implementation guide refer only to the data stored in the IMMERSE central research database and the long-term archive. Further de-identification measures

may be applied to data exports taking into account the specific requirements of the individual data use projects. This could e.g. include the conversion of date/timestamps to durations relative to study enrollment, suppression, generalization or perturbation of potential pseudo-identifiers or the transformation of sensitive attributes to reduce risks (e.g. geocoordinates to distance covered). In addition to these technical measures, data protection will be addressed as well by organizational measures e.g. covering appropriate use and dissemination in the data use agreements of individual data use projects.

2.4. Data Elements for Study & Participants

In order to anchor the data elements captured during the course of the trial for each participant, a representation of the study participant is required. Please note that both patients as well as treating physicians are study participants that need to be differentiated in the research database.

For the study participant, FHIR provides the basic resource “ResearchSubject”, which was profiled as “IMMERSE_ResearchSubject” to fit IMMERSE requirements. The assignment of study participants into trial arms is expressed as an attribute of IMMERSE_ResearchSubject. Since ResearchSubject does not provide any demographic data, additionally the FHIR basic resource “Patient” was profiled as “IMMERSE_Patient”. It should be noted that treating physicians also contribute to the study as participants who document the perspective of a service user. Since FHIR currently does not permit to reference Practitioner resources from Observation resources, treating physicians will also be represented using IMMERSE_Patient resources. The differentiation between patients and treating physicians as study participants will be implemented using the IMMERSE_ResearchSubject.assignedArm attribute.

For representation in the IMMERSE research database, for each study participant a single instance of IMMERSE_ResearchSubject is required, which references a single instance of the IMMERSE_Patient profile. All data captured for the study participant will reference the IMMERSE_ResearchSubject instance.

2.4.1. IMMERSE_ResearchSubject

Role of data element and relationship to other data elements in the study

The IMMERSE_ResearchSubject resource is the structured representation of an IMMERSE study participant in the research database. All questionnaire responses and observations contain reference to the respective IMMERSE_Patient. The study and consent attributes are implemented as static strings, as they are not explicitly modeled as distinct structured resources in the research dataset.

Attributes of the data element

The following table lists the individual attributes of the FHIR resource, including short descriptions and semantic encodings for attributes and value sets, where applicable:

Attribute Name	Attribute Datatype	Attribute Description	Semantic encoding
identifier	integer	Unique pseudonymous identifier of the research subject within the study	-

status	value set (ResearchSubjectStatus)	Status of the research subject in the study	-
period	date range	Start and end of participation	-
study	identifier(Identifier)	Identifier of the study	-
individual	integer	reference to an IMMERSE_Patient resource providing demographic data for the research subject	-
assignedArm	string	Which arm of the study (intervention, control as well as patient/treating physician) the research subject was assigned to	-
actualArm	string	The arm of the study that the research study actually participated in	-
consent	identifier(Identifier)	Identifier of the project consent	-

FHIR Representation of the data element

The IMMERSE_ResearchSubject resource is a profile of the generic FHIR ResearchSubject resource. It is available online at Simplifier: https://simplifier.net/immerse-eu/immerse_researchsubject

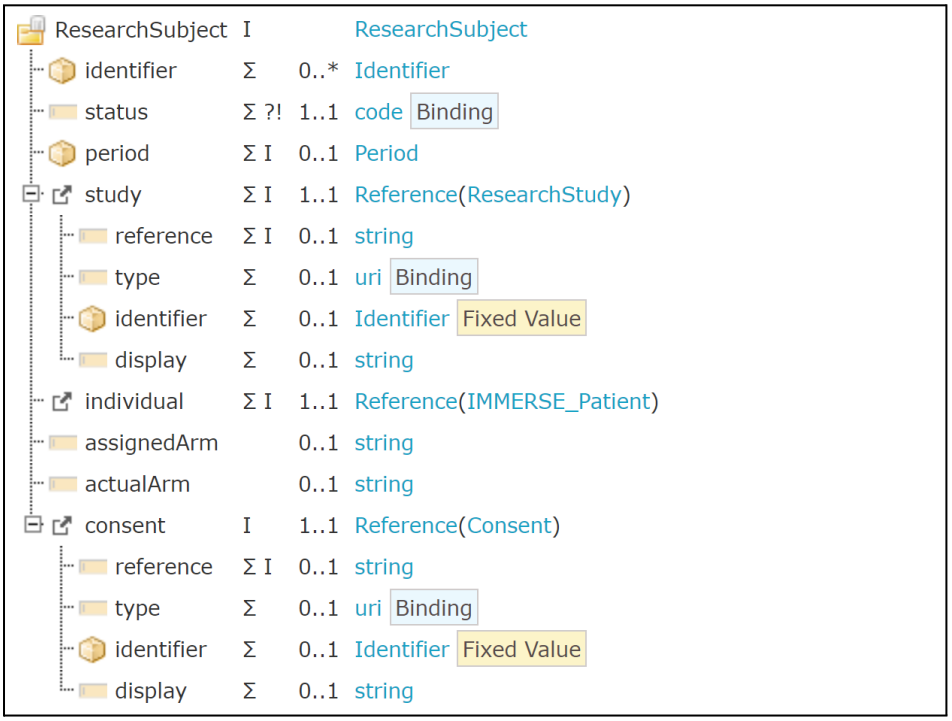

Fig. 2: Screenshot of the IMMERSE_ResearchSubject resource from Simplifier

Changes from the generic FHIR resource to the IMMERSE profile:

- the cardinality of the consent reference was changed from 0..1 to 1..1 as the IMMERSE research database only contains data from patients who have given informed consent

Data protection aspects of the FHIR profile

The IMMERSE_ResearchSubject resource is stored with a pseudonymous identifier. Besides the enrollment in the IMMERSE study, it does not contain any directly or indirectly identifying or medical attributes.

Implementation aspects of the FHIR profile

A single instance of the IMMERSE_ResearchSubject resource needs to be generated for each participant of the IMMERSE study based on data from the ID management system.

2.4.2. IMMERSE_Patient

Role of data element in the study

The IMMERSE_Patient resource is the structured representation of the demographics data of an IMMERSE study participant (patient or treating physician) in the research database.

Relationship to other data elements in the study

The IMMERSE_Patient resource is referenced by the IMMERSE_ResearchSubject resource.

Attributes of the data element

The following table lists the individual attributes of the FHIR resource, including short descriptions and semantic encodings for attributes and value sets, where applicable:

Attribute Name	Attribute Datatype	Attribute Description	Semantic encoding
identifier	integer	Unique pseudonymous identifier of the research subject within the study	-
gender	code (AdministrativeGender)	gender of the research subject	-
birthDate	date	partial birthdate (YYYY-MM) of the research subject	-
deceased.deceasedBoolean	boolean	Indicator whether the study subject is deceased or not	-
deceased.deceasedDateTime	date	partial death date (YYYY-MM) of the research subject, if applicable	-
communication language	codeableConcept (CodeableConcept)	Language to communicate with the research subject	FHIR languages value set

FHIR Representation of the data element

The IMMERSE_Patient resource is a profile of the generic FHIR Patient resource. It is available online at Simplifier: https://simplifier.net/immerse-eu/immerse_patient

A static screenshot of the IMMERSE_Patient FHIR profile is shown in Figure 3:

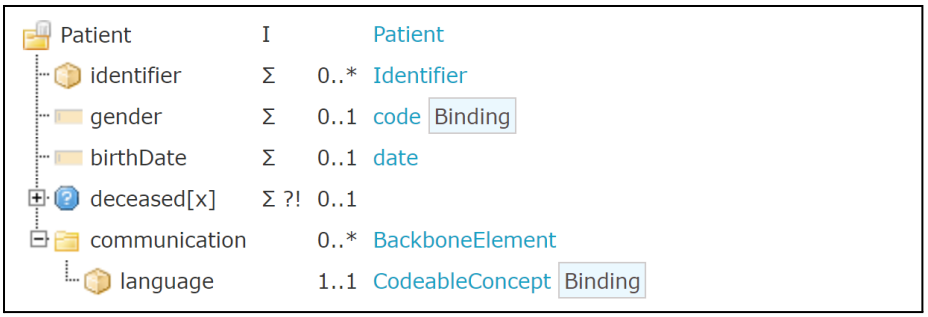

Fig. 3: Screenshot of IMMERSE_Patient FHIR profile

Changes from the generic FHIR resource to the IMMERSE profile:

- irrelevant data elements (e.g. related to clinical care) were removed.

Data protection aspects of the FHIR profile

The IMMERSE_Patient resource is stored with a pseudonymous identifier. The birthdate is stored as a partial date with just the year and month to reduce re-identification risks due to comparison with a full birthdate. Also, a potential death date is stored partially with year and month only. No other personally identifiable data is stored in this resource.

Implementation aspects of the FHIR profile

A single instance of the IMMERSE_Patient resource needs to be generated for each participant of the IMMERSE study based on data from the ID management system.

2.5. Data Elements for Questionnaires & Responses

Questionnaires are captured in IMMERSE both in the form of self-entered ESM forms in the DMMH or movisensXS components as well as through eCRFs and stored in the eCRF platform. FHIR provides the basic resource “Questionnaire” for the representation of forms with sections and items, as well as the resource “QuestionnaireResponse” for responses entered into the questionnaires, referenced back to the structure of the respective Questionnaire. For IMMERSE, an “IMMERSE_Questionnaire” as well as an “IMMERSE_QuestionnaireResponse” profile was created. Since the QuestionnaireResponse basic resource does not allow for semantic encoding of responses, an additional “IMMERSE_QuestionnaireResponseObservation” profile was created that replicates the content of the QuestionnaireResponse, but adds semantic annotations.

The semantic annotation of individual questionnaires is not covered by this version of the implementation guide. It will be added in a later revision once all ESM forms and eCRF questionnaires have been defined and semantic annotation carried out.

2.5.1. IMMERSE_Questionnaire

Role of data element in the study

The IMMERSE_Questionnaire profile is the structured representation of an ESM form or eCRF questionnaire in the IMMERSE research database. A single instance of this resource is required for each ESM form or questionnaire used in the study.

Relationship to other data elements in the study

IMMERSE_Questionnaire provides the structure of questionnaires referenced by the responses stored in instances of the IMMERSE_QuestionnaireResponse resources.

Attributes of the data element

Attribute Name	Attribute Datatype	Attribute Description	Semantic encoding
identifier	integer	identifier of the questionnaire	-
version	string	version of the questionnaire	-
name	string	computer-friendly name for the questionnaire	-
title	string	human-friendly name for the questionnaire	-
status	value set (PublicationStatus)	status of the questionnaire	-
subjectType	codeableConcept (ResourceType)	Subject that the Questionnaire is applied to (fixed value “IMMERSE_ResearchSubject”)	-
date	datetime	date (and optionally time) when the questionnaire was published	-
publisher	string	name of the the organization or individual that published the questionnaire	-
description	string	description of the questionnaire in free form text	-

useContext	UsageContext (UsageContext)	semantic encoding of the usage context of the questionnaire	applicable terminology will be chosen individually for each questionnaire
copyright	string	copyright statement relating to the questionnaire	-
approvalDate	date	date on which the resource content was approved by the publisher	-
lastReviewDate	date	date on which the resource content was last reviewed	-
effectivePeriod	Period (Period)	period during which the questionnaire content is planned to be in active use	-
code	codeableConcept (CodeableConcept)	Semantic encoding for the overall topic of the questionnaire	applicable terminology will be chosen individually for each questionnaire
item	BackboneElement	A questionnaire can contain multiple items; items can also be nested to form sections or groups within the questionnaire. The following attributes can be set for each item.	-
item.linkId	string	unique ID for item in questionnaire	-
item.code	codeableConcept (CodeableConcept)	Semantic encoding of the content of the item in the questionnaire	applicable terminology will be chosen individually for each questionnaire item
item.prefix	string	short label for a question item or item group	-
item.text	string	text description of a question item	-
item.type	QuestionnaireItemType (QuestionnaireItem Type)	type of questionnaire item	-
item.enableWhen	BackboneElement	Optional constraint indicating that this item should only be enabled when the specific condition is true	-
item.enableBehavior	code (EnableWhenBehavior)	controls how multiple enableWhen constraints are interpreted - whether all or any must be true	-
item.required	boolean	indication, if the item must be present in a "completed" QuestionnaireResponse or not	-
item.repeats	boolean	indication, if the item may occur multiple times in the response or not	-
item.maxLength	integer	maximum number of characters that are permitted in the answer to be considered a valid QuestionnaireResponse	-
item.answerValueSet	value set	value set containing a list of codes representing permitted answers for a choice question	-
item.answerOption	BackboneElement	optional list of selectable answers to an item	applicable terminology will be chosen individually for each questionnaire item
item.item	BackboneElement	nested item	-

FHIR Representation of the data element

The IMMERSE_Questionnaire is a profile of the generic FHIR Questionnaire resource. Attributes irrelevant for use in the IMMERSE Project were removed during profiling. The resulting FHIR profile is available online on Simplifier:

https://simplifier.net/immerse-eu/immerse_questionnaire

A static screenshot of the IMMERSE_Questionnaire FHIR profile is shown in Figure 4:

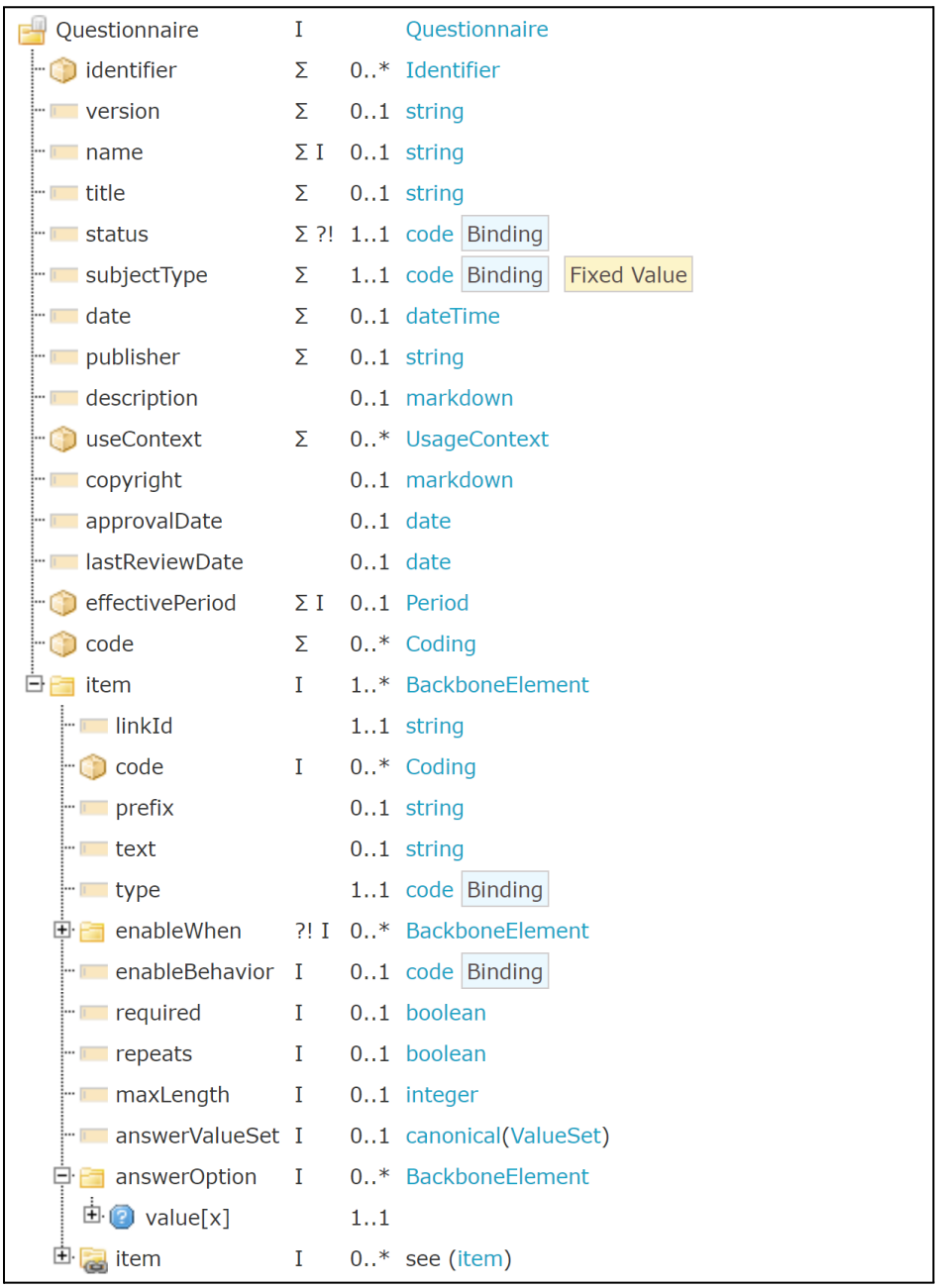

Fig. 4: Screenshot of IMMERSE_Questionnaire FHIR profile

Data protection aspects of the FHIR profile

The IMMERSE_Questionnaire profile does not contain any personally identifiable data.

Implementation aspects of the FHIR profile

A single instance of the IMMERSE_Questionnaire profile needs to be created for each ESM form or eCRF questionnaire based on the form and/or eCRF definitions made available by WP7. Each form needs to be semantically encoded with regard to the context and topic of the questionnaire. Each item needs to be semantically annotated with regards to the topic of the item, as well as each entry of value sets used in these items. If available, concepts from a suitable openly available terminology are to be used. Alternatively, original coding from the form can be used.

2.5.2. IMMERSE_QuestionnaireResponse

Role of data element in the study

The IMMERSE_QuestionnaireResponse profile is the structured representation of a study participant's response to an ESM form or eCRF questionnaire. An instance of IMMERSE_QuestionnaireResponse contains all answers for each item of the referenced IMMERSE_Questionnaire resource. An instance of IMMERSE_QuestionnaireResponse is required for each time that a study participant fills out an ESM form or provides answers to an eCRF questionnaire. An IMMERSE_QuestionnaireResponse resource contains all the responses that were to the items in the referenced Questionnaire in a single execution of the questionnaire.

Relationship to other data elements in the study

An IMMERSE_QuestionnaireResponse resource references the study participant through their IMMERSE_Patient resource as well as the IMMERSE_Questionnaire resource for which it contains the responses.

Attributes of the data element

Attribute Name	Attribute Datatype	Attribute Description	Semantic encoding
identifier	integer	unique id for this set of answers	-
questionnaire	integer	the questionnaire being answered	
status	value set (QuestionnaireResponseStatus)	status of the questionnaire response such as in-progress, completed, amended, entered-in-error or stopped	
subject	integer	study participant who gave the questionnaire response (reference to an IMMERSE_Patient resource)	
authored	datetime	date the answers were gathered	
item	BackboneElement	Like the questionnaire, a questionnaireResponse can contain multiple items; items can also be nested to form sections or groups within the questionnaire. The following attributes can be set for each item. The structure of items in the response is identical to the structure in the referenced questionnaire.	
item.linkId	string	pointer to specific question item from Questionnaire	

		that corresponds to this response item	
item.text	string	text of the question being answered.	
item.answer.value	datatype depends on definition in the referenced questionnaire item	answer to the question of the research subject	
item.answer.item	BackboneElement	Nested questions found within this answer.	
item.item	BackboneElement	Nested questions	

FHIR Representation of the data element

The IMMERSE_QuestionnaireResponse profile was created based on the generic FHIR questionnaireResponse resource. Data elements irrelevant for use in the IMMERSE project were removed during profiling. The resulting FHIR profile is available online on Simplifier:

https://simplifier.net/immerse-eu/immerse_questionnaireresponse

A static screenshot of the IMMERSE_QuestionnaireResponse FHIR profile is shown in Figure 5:

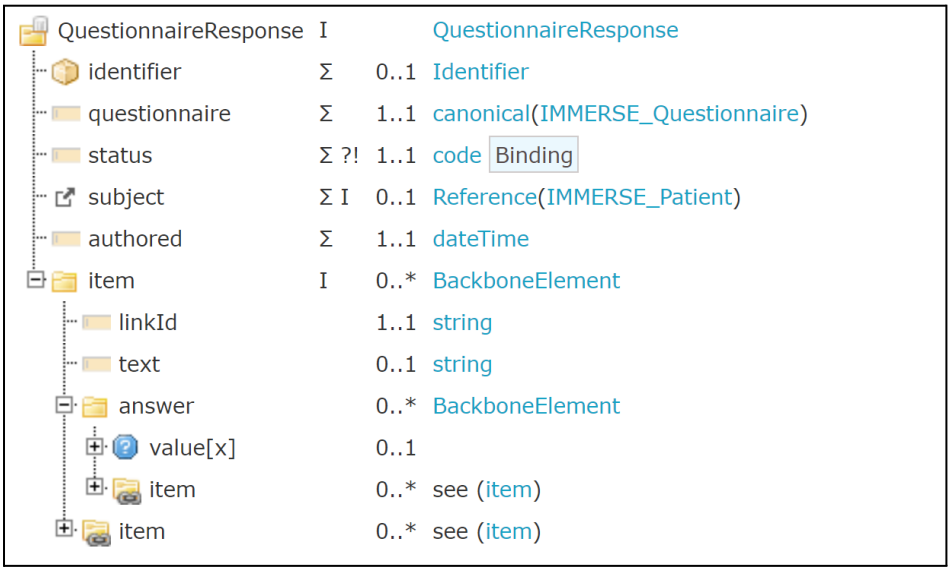

Fig. 5: Screenshot of IMMERSE_QuestionnaireResponse FHIR profile

Data protection aspects of the FHIR profile

The IMMERSE_QuestionnaireResponse object is stored with a pseudonymous identifier of the study participant. The backbone part of the resource does not contain any personally identifying data of the study participant, however the response “payload” of the resource could contain sensitive items in case of free text data entry. It therefore must be enforced through organizational measures that no questionnaire items containing personally identifiable information (e.g. name, full birthdate) be used in the IMMERSE study. Free text items that could contain personally identifiable data entered erroneously by the study participant must be redacted before transfer into the IMMERSE research database. It should be noted that even when applying these measures, questionnaire responses

could still contain information that may indirectly expose a study participants identity through a unique combination of traits that an attacker may be able to link to other data sources available to them. IMMERSE_QuestionnaireResponse data should therefore be considered sensitive, and organizational measures should be taken to reduce re-identification risks (e.g. data minimization, contractual prohibitions on re-identification). Also, technical measures like anonymization of exported data should be considered in the context of the specific data items, risks and data quality requirements of the individual data use projects.

Implementation aspects of the FHIR profile

A single instance of the IMMERSE_QuestionnaireResponse resource needs to be created for each ESM form or eCRF questionnaire filled out by or for a study subject, based on information available from either the MoMent app or the eCRF platform. Besides the reference to the IMMERSE_Questionnaire resource, versions and linkIds to the individual items between the response and questionnaire must be aligned.

2.5.3. IMMERSE_QuestionnaireResponseObservation

Role of data element in the study

The IMMERSE_QuestionnaireResponseObservation is an auxiliary profile to provide a structured representation of semantically annotated responses to data entered into IMMERSE ESM forms or eCRF questionnaires. It mirrors the response data in the IMMERSE_QuestionnaireResponse resource, but adds semantic annotations to the responses. A single instance of this resource is required for each individual response to an item of the referenced QuestionnaireResponse.

Relationship to other data elements in the study

The IMMERSE_QuestionnaireResponseObservation references the study participant through their IMMERSE_Patient resource, as well as the IMMERSE_QuestionnaireResponse from which it was derived, and the linkId of the individual response item in it.

Attributes of the data element

Attribute Name	Attribute Datatype	Attribute Description	Semantic encoding
identifier	integer	Identifier of the resource	-
status	value set (ObservationStatus)	status of the questionnaire response observation	-
code	codeableConcept (CodeableConcept)	semantic encoding of the questionnaire item (taken from the “item.code” attribute of the reference IMMERSE_Questionnaire resource	defined for individual instances of the resource depending on content
subject	integer	unique pseudonymous identifier of the study participant within the study (reference to the IMMERSE_Patient resource)	-
effective	datetime	date and time when the questionnaire response was entered (taken from the “authored” attribute of the referenced IMMERSE_QuestionnaireResponse resource)	-
derivedFrom	integer	Reference to the IMMERSE_QuestionnaireResponse resource the the data was taken from	-

component.itemLinkId	string	LinkId of the questionnaire item for which the resource contains the answer (taken from the “item.linkId” attribute of the referenced IMMERSE_QuestionnaireResponse resource)	-
component.itemAnswer	datatype depends on definition in the referenced questionnaire item	response value (taken from the “item.answer” attribute of the referenced IMMERSE_QuestionnaireResponse resource); for questions with predefined value sets, the semantic encoding is taken from the referenced item of the IMMERSE_Questionnaire resource	defined for individual instances of the resource depending on content

FHIR Representation of the data element

The IMMERSE_QuestionnaireResponseObservation was profiled based on the generic FHIR Observation resource. Attributes irrelevant in the IMMERSE project were removed during profiling. To represent both the linkId to the individual questionnaire item as well as the response value, the component attribute was sliced.

The resulting IMMERSE_QuesitonnaireResponseObservation FHIR profile is available online on Simplifier:

https://simplifier.net/immerse-eu/immerse_questionnaireresponseobservation

A static screenshot of the IMMERSE_QuestionnaireResponseObservation FHIR profile is shown in Figure 6:

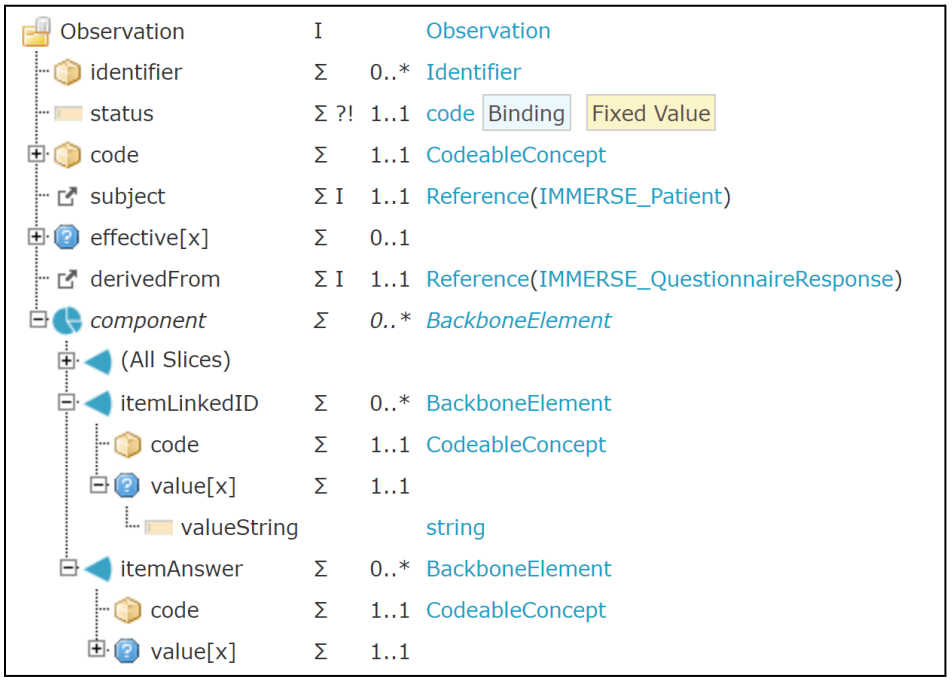

Fig. 6: Screenshot of IMMERSE_QuestionnaireResponseObservation FHIR profile

Data protection aspects of the FHIR profile

The same caveats apply as for the IMMERSE_QuestionnaireResponse profile described above.

Implementation aspects of the FHIR profile

A single instance of the IMMERSE_QuestionnaireResponseObservation needs to be created for each item of a response to an ESM form or eCRF questionnaire, based on data available from the referenced IMMERSE_QuestionnaireResponse and IMMERSE_Questionnaire resources.

2.6. Data Elements for Mobile Sensor Data

Mobile sensing data is captured in IMMERSE through the movisensXS platform. It utilizes sensors in the mobile phone of the study participant to capture direct sensor or derived data elements which will be analyzed in WP4.

The following mobile sensor data types were selected to be used in IMMERSE:

- Geolocation
- Physical activity
- Steps taken
- Device status (on/off)
- Display status (on/off)
- Categorized App usage
- Notification usage

movisensXS will also be used to capture ESM data during specific observation periods throughout the study (both in the intervention as well as the control group). ESM data will be represented as IMMERSE_QuestionnaireResponse and IMMERSE_QuestionaireResponseObservation resources as described above.

Data is captured in movisensXS in pseudonymized form with timestamps given as relative duration in seconds since enrollment into the app. Data can be exported in CSV flat file format accompanied by an XML file giving definitions of the CSV file columns as well as the enrollment timestamp.

Sensor data will be transformed into appropriate structured and semantically annotated FHIR formats before storage in the IMMERSE research database, including the conversion of relative timestamps to absolute date/time data as well as addition of semantic annotations where available. Geolocation data is additionally preprocessed to reduce risks of re-identification due to the leakage of home or work locations from the tracked geolocations.

Data protection aspects are discussed in detail for each of data types selected for capture in IMMERSE in the sections below. A relevant risk was identified for geolocation data that would permit an attacker to re-identify a subject without the need for linkage to a separate dataset. Accordingly, measures have been implemented to address this risk by modifying geolocation data before importing it into the IMMERSE research database. For all other sensor data types, a comparison dataset would be required to match the IMMERSE dataset to, coming either from an app installed in parallel to movisensXS on the trial subjects mobile device or (in one case) from a fitness tracker worn by the subject. Considerations regarding availability of such comparison datasets, efforts required for implementing a re-identification attack and possible benefits are described in detail for each data type in the following sections.

2.6.1. IMMERSE_GeolocationObservation

Role of data element in the study

Geolocation is one of the sensing data elements that can be acquired through the movisensXS app if configured on a study subject's compatible mobile phone. Recording of subject location is triggered whenever the subject moves more than a specified distance from the previous location. Geolocation is detected through either the GPS (global positioning system) satellite service or through proximity to known WIFI access points. Each measurement is timestamped based on seconds relative to the enrollment of the study subject into the movisensXS app. In addition to the latitude, longitude and

altitude measurement, an accuracy measure is given, describing a confidence circle around the location given in the measurement Geolocation data is captured in the study to determine mobility of the study subjects.

Relationship to other data elements in the study

Geolocation data is collected in relation to each study subject, encoded in the IMMERSE_Patient FHIR profile. Each geolocation data instance is linked to its study participant via a FHIR reference to the respective IMMERSE_Patient instance. There are no explicit references to other IMMERSE data elements, but timestamps can be used to relate geolocation instances to instances of other IMMERSE data instances.

Attributes of the data element

The following table lists the individual attributes of the FHIR resource, including short descriptions and semantic encodings for attributes and value sets, where applicable:

Attribute Name	Attribute Datatype	Attribute Description	Semantic encoding
identifier	integer	identifier of the sensing measurement within the study	-
status	value set (ObservationStatus)	status of the measurement in the study	-
code	codeableConcept (CodeableConcept)	type of measurement	snomedct:68130003
subject	integer	unique pseudonymous identifier of the research subject within the study (reference to the IMMERSE_Patient resource)	
effective	datetime	date and time of measurement	
latitude	float	geographic latitude of the study subject at the timepoint of measurement	loinc:91588-4
longitude	float	geographic longitude of the study subject at the timepoint of measurement	loinc:915789-2
altitude	float	altitude of the study subject above sea level	loinc:41905-1
accuracy	float	accuracy of lat/lon measurement in meters around the stated geolocation	snomedct:371884006

FHIR Representation of the data element

For the FHIR representation of geolocation data, the FHIR Observation resource was chosen. In the course of profiling this resource, irrelevant data elements (e.g. related to clinical care) were removed. To cover the individual geolocation attributes, the components branch of the resource was sliced to explicitly represent values for latitude, longitude, altitude and accuracy. Adequate semantic encodings were searched in LOINC and SNOMED and could be retrieved for all attributes.

The resulting IMMERSE_Geolocation FHIR profile is available online on Simplifier:

https://simplifier.net/immerse-eu/immerse_geolocationobservation

A static screenshot of the IMMERSE_GeolocationObservation FHIR profile is shown in Figure 7:

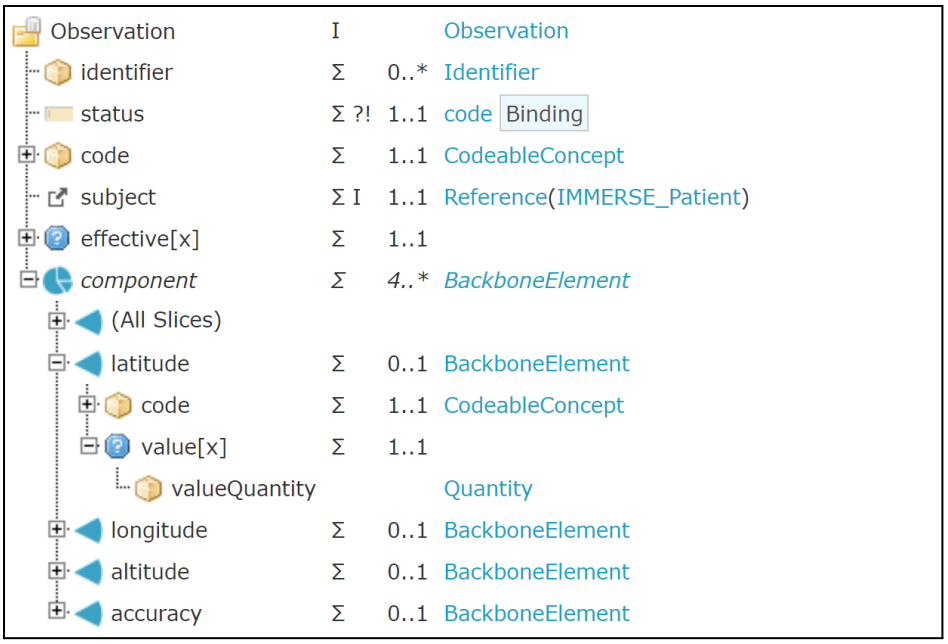

Fig. 7: Screenshot of IMMERSE_Geolocation FHIR profile

Data protection aspects of the FHIR profile

Geolocation data is highly sensitive in the sense that the geographic locations can easily reveal the home and/or work locations of study subjects. Because of this, geolocation data is anonymized by shifting all geographic measurements to a latitude, longitude and altitude (0,0,0) location based on the average of all latitude/longitude/altitude measurements recorded in the full geolocation track. This approach centers the geo track to an identical location for all study participants, so only relative movements between measurements are available for analysis, which is sufficient to assess the mobility of study participants.

Re-identification from the geotracks shifted to a (0,0,0) location could in theory still be possible by trying to match tracks to roadway patterns e.g. from the OpenStreetMap database. This would however require a very large effort for the implementation of an algorithm that would need to match tracks against a very large set of roadways (as of 04/2023 OpenStreetMap is listing >900.000.000 ways and relations). In view of the benefits of re-identification and related effort, this approach is not considered a relevant privacy risk. However, in the context of data provision for data use projects, additional measures (e.g. replacement of individual coordinates with distance covered) will be considered in the context of the specific requirements of the project.

Implementation aspects of the FHIR profile

A single instance of the IMMERSE_Geolocation profile needs to be created for each individual geolocation measurement available from movisensXS. Anonymization of geotracks needs to be carried out based on the data of the full track before data is imported into the IMMERSE research database.

Source data for this data element is produced by the movisensXS application. The following aspects

need to be taken into account during implementation of the transformation from movisensXS raw data to FHIR resources:

- Attribute “effective”: movisensXS provides a relative timestamp for the measurement in seconds since enrollment into the study (integer) which needs to be converted to a regular absolute date & time datatype during transformation to FHIR.
- Anonymization of geolocation data needs to be carried out before import into the research database

movisensXS source attribute	FHIR profile attribute	Transformation
geo_rel_timestamp	effective	conversion to absolute date/timestamps
latitude	component.latitude	-
longitude	component.longitude	-
altitude	component.altitude	-
accuracy	component.accuracy	-

2.6.2. IMMERSE_ActivityObservation

Role of data element in the study

Physical activity is on the sensing data elements that can be acquired through the movisensXS app if configured on a study subject’s compatible mobile phone. Detection of subject physical activity is based on Google Play Services, and it categorizes physical activity into 5 groups (IN_VEHICLE, ON_BICYCLE, ON_FOOT, STILL, UNKNOWN, and TILTING). Besides the category of activity, a confidence measure is generated, giving the likelihood of the activity being correct. Each measurement is timestamped based on seconds relative to the enrollment of the study subject into the movisensXS app. Physical activity data is captured in the study to determine activity level of the study subjects.

Relationship to other data elements in the study

Physical activity data is collected in relation to each study participant, encoded in the IMMERSE_Patient FHIR profile. Each physical activity data instance is linked to its study participant via a FHIR reference to the respective IMMERSE_Patient instance. There are no explicit references to other IMMERSE data elements, but timestamps can be used to relate physical activity instances to instances of other IMMERSE data instances.

Attributes of the data element

The following table lists the individual attributes of the FHIR resource, including short descriptions and semantic encodings for attributes and value sets, where applicable:

Attribute Name	Attribute Datatype	Attribute Description	Semantic encoding
identifier	integer	Identifier of the sensing measurement within the study	-

status	value set (ObservationStatus)	Status of the measurement in the study	-
code	codeableConcept (CodeableConcept)	Type of measurement	snomedct:68130003
subject	integer	Unique pseudonymous identifier of the research subject within the study (reference to the IMMERSE_Patient resource)	
effective	datetime	date and time of measurement	
activityCategory	codableConcept	estimated physical activity of the study subject	value set: ● IN_VEHICLE = IN_VEHICLE ● ON_BICYCLE = snomedct:300648001 ● ON_FOOT = snomedct:300615001 ● STILL = snomedct:300610006 ● UNKNOWN = snomedct:261665006 ● TILTING = TILTING
activityConfidence	integer	the likelihood that the study subject is performing this activity (0~100)	Attribute: snomedct:118582008

FHIR Representation of the data element

For the FHIR representation of Physical activity data, the FHIR Observation resource was chosen. In the course of profiling this resource, irrelevant data elements (e.g. related to clinical care) were removed. To cover the individual Physical activity attributes, the components branch of the resource was sliced to explicitly represent values for activityCategory and activityConfidence. Adequate semantic encodings were searched in LOINC and SNOMED and could be retrieved for all attributes. For the activityCategory value set, only a partial mapping could be achieved, covering 4 out of 6 values. For the remaining 2 options, the coding provided by movisensXS was kept in place.

The resulting IMMERSE_Activity FHIR profile is available online on Simplifier: https://simplifier.net/immerse-eu/immerse_activityobservation

A static screenshot of the IMMERSE_Activity FHIR profile is shown in Figure 8:

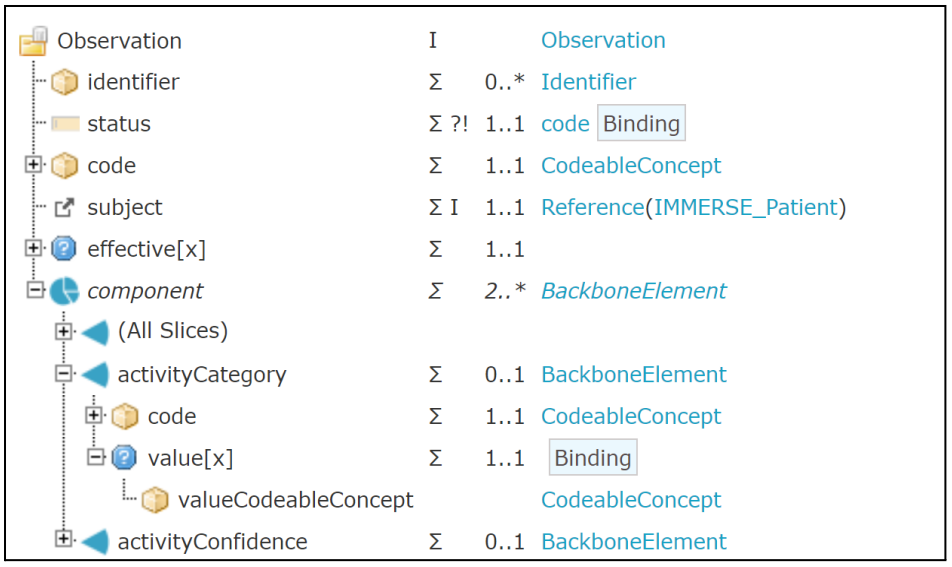

Fig. 8: Screenshot of IMMERSE_Activity FHIR profile

Data protection aspects of the FHIR profile

Activity data is collected by means of timestamped broad categories (e.g. subject is on bicycle or in vehicle). For re-identification, an attacker would need to have access to a dataset with similar activity and timestamp data (e.g. from a fitness tracker device worn by the subject, or an app installed on the same device at the same time) and develop an algorithm to match it to the IMMERSE steps data. In view of the benefits of re-identification and related effort, this approach is not considered a relevant privacy risk. However, in the context of data provision for data use projects, additional measures (e.g. conversion of timestamps to relative durations since study enrolment) will be considered in the context of the specific requirements of the project.

Implementation aspects of the FHIR profile

Source data for this data element is produced by the movisensXS application. The following aspects need to be taken into account during implementation of the transformation from movisensXS raw data to FHIR resources:

- Attribute “effective”: movisensXS provides a relative timestamp for the measurement in seconds since enrollment into the study (integer) which needs to be converted to a regular absolute date & time datatype during transformation to FHIR.
- Attribute “activityCategory”: the value set could not be semantically encoded fully based on available standardized terminologies. Suitable concepts were located in SNOMED CT for 4 of 6 values. The other 2 values were kept as-is.

movisensXS source attribute	FHIR profile attribute	Transformation
activity_rel_timestamp	effective	conversion to absolute date/timestamps
activityCategory	component.activityCategory	mapping to value set
activityConfidence	component.activityConfidence	-

2.6.3. IMMERSE_StepsObservation

Role of data element in the study

Steps is on the sensing data elements that can be acquired through the movisensXS app if configured on a study subject's compatible mobile phone. This data is estimated based on 5 seconds of measurement of the Smartphone accelerometer in a minute to determine the steps taken in that duration. Each measurement is timestamped based on seconds relative to the enrollment of the study subject into the movisensXS app. Steps data is captured in the study to determine activity level of the study subjects besides Physical activity data.

Relationship to other data elements in the study

Steps data is collected in relation to each study participant, encoded in the IMMERSE_Patient FHIR profile. Each Steps data instance is linked to its study participant via a FHIR reference to the respective IMMERSE_Patient instance. There are no explicit references to other IMMERSE data elements, but timestamps can be used to relate Steps instances to instances of other IMMERSE data instances.

Attributes of the data element

The following table lists the individual attributes of the FHIR resource, including short descriptions and semantic encodings for attributes and value sets, where applicable:

Attribute Name	Attribute Datatype	Attribute Description	Semantic encoding
identifier	integer	Identifier of the sensing measurement within the study	-
status	value set (ObservationStatus)	Status of the measurement in the study	-
code	codeableConcept (CodeableConcept)	Type of measurement	loinc:55423-8
subject	integer	Unique pseudonymous identifier of the research subject within the study (reference to the IMMERSE_Patient resource)	
effective	datetime	date and time of measurement	
value	integer	steps of the study subject	

FHIR Representation of the data element

For the FHIR representation of Steps data, the FHIR Observation resource was chosen. In the course of profiling this resource, irrelevant data elements (e.g. related to clinical care) were removed. Adequate semantic encoding for the attribute was retrieved in LOINC.

The resulting IMMERSE_Steps FHIR profile is available online on Simplifier:

https://simplifier.net/immerse-eu/immerse_stepsobservation

A static screenshot of the IMMERSE_Steps FHIR profile is shown in Figure 9:

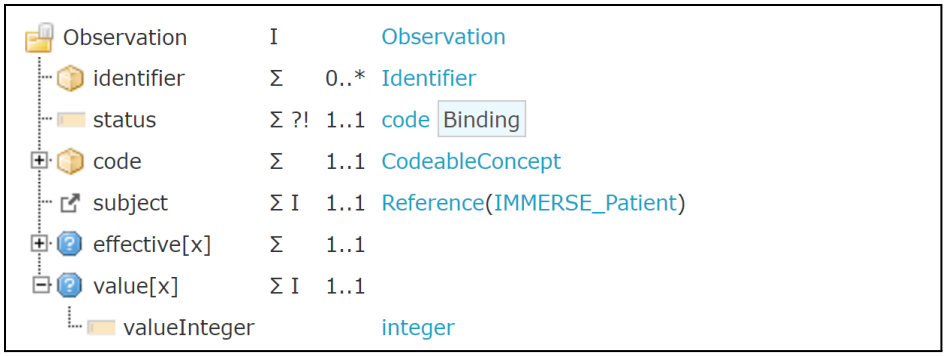

Fig. 9: Screenshot of IMMERSE_Steps FHIR profile

Data protection aspects of the FHIR profile

Steps data is collected by means of a timestamped number of steps taken by the subject. For re-identification, an attacker would need to have access to a dataset with similar step count and timestamp data (e.g. from a fitness tracker device worn by the subject) and develop an algorithm to match it to the IMMERSE activity data. In view of the benefits of re-identification and related effort, this approach is not considered a relevant privacy risk. However, in the context of data provision for data use projects, additional measures (e.g. conversion of timestamps to relative durations since study enrolment) will be considered in the context of the specific requirements of the project.

Implementation aspects of the FHIR profile

Source data for this data element is produced by the movisensXS application. The following aspect needs to be taken into account during implementation of the transformation from movisensXS raw data to FHIR resources:

- Attribute “effective”: movisensXS provides a relative timestamp for the measurement in seconds since enrollment into the study (integer) which needs to be converted to a regular absolute date & time datatype during transformation to FHIR.

movisensXS source attribute	FHIR profile attribute	Transformation
steps_rel_timestamp	effective	conversion to absolute date/timestamps
steps	value	-

2.6.4. IMMERSE_DeviceOnOffObservation

Role of data element in the study

Device On/Off is on the sensing data elements that can be acquired through the movisensXS app if configured on a study subject’s compatible mobile phone. Recording of device state is triggered whenever the state of the study subjects mobile phone changes, for example from ‘on’ to ‘off’, or from ‘off’ to ‘on’. Each measurement is timestamped based on seconds relative to the enrollment of the study subject into the movisensXS app. Device On/Off data is captured in the study to control possible technical issues so that it could be helpful to track reasons for missings.

Relationship to other data elements in the study

Device On/Off data is collected in relation to each study participant, encoded in the IMMERSE_Patient FHIR profile. Each Device On/Off data instance is linked to its study participant via a FHIR reference to the respective IMMERSE_Patient instance. There are no explicit references to other IMMERSE data elements, but timestamps can be used to relate Device On/Off instances to instances of other IMMERSE data instances.

Attributes of the data element

The following table lists the individual attributes of the FHIR resource, including short descriptions and semantic encodings for attributes and value sets, where applicable:

Attribute Name	Attribute Datatype	Attribute Description	Semantic encoding
identifier	integer	Identifier of the sensing measurement within the study	-
status	value set (ObservationStatus)	Status of the measurement in the study	-
code	codeableConcept (CodeableConcept)	Type of measurement	**DEVICE_ON_OFF**
subject	integer	Unique pseudonymous identifier of the research subject within the study (reference to the IMMERSE_Patient resource)	-
effective	datetime	date and time of measurement	-
value	integer	state of the smartphone device (running or not)	**value set:** ● **DEVICE_OFF** ● **DEVICE_ON**

FHIR Representation of the data element

For the FHIR representation of Device On/Off data, the FHIR Observation resource was chosen. In the course of profiling this resource, irrelevant data elements (e.g. related to clinical care) were removed. Since no adequate semantic encoding for this attribute and its value set could be retrieved from LOINC or SNOMED, an alphanumeric code “DEVICE_ON_OFF” was designated for the attribute, and the original coding from movisensXS was maintained for the value set.

The resulting IMMERSE_Device On/Off FHIR profile is available online on Simplifier:

https://simplifier.net/immerse-eu/immerse_deviceonoffobservation

A static screenshot of the IMMERSE_DeviceOnOff FHIR profile is shown in Figure 10:

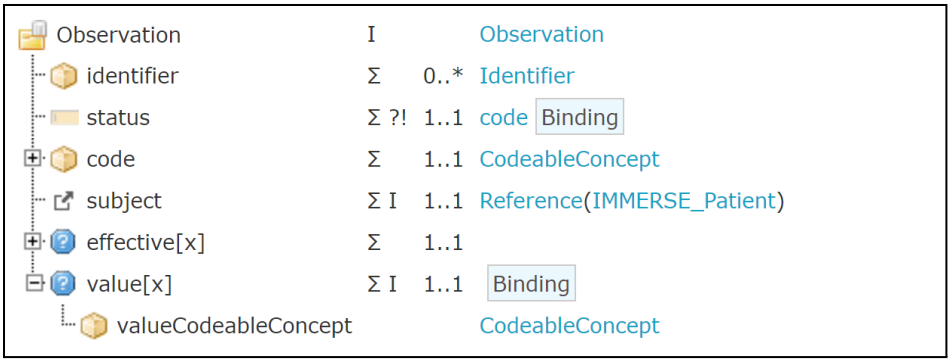

Fig. 10: Screenshot of IMMERSE_DeviceOnOff FHIR profile

Data protection aspects of the FHIR profile

Device on/off data is collected by means of timestamped status when the subject's mobile phone was switched on or off. For re-identification, an attacker would need to have access to a dataset with similar device on/off and timestamp data (e.g. from an app installed at the same device at the same time) and develop an algorithm to match it to the IMMERSE data. In view of the benefits of re-identification and related effort, this approach is not considered a relevant privacy risk. However, in the context of data provision for data use projects, additional measures (e.g. conversion of timestamps to relative durations since study enrolment) will be considered in the context of the specific requirements of the project.

Implementation aspects of the FHIR profile

Source data for this data element is produced by the movisensXS application. The following aspects need to be taken into account during implementation of the transformation from movisensXS raw data to FHIR resources:

- Attribute "effective": movisensXS provides a relative timestamp for the measurement in seconds since enrollment into the study (integer) which needs to be converted to a regular absolute date & time datatype during transformation to FHIR.
- Attribute "value": the value set could not be semantically encoded based on available standardized terminologies. The movisensXS values were kept as-is.

movisensXS source attribute	FHIR profile attribute	Transformation
device_rel_timestamp	effective	conversion to absolute date/timestamps
device_onoff	value	mapping to value set

2.6.5. IMMERSE_DisplayOnOffObservation

Role of data element in the study

Display On/Off is on the sensing data elements that can be acquired through the movisensXS app if configured on a study subject's compatible mobile phone. Recording of display state is triggered whenever the state of the mobile phones display changes, for example from 'on' to 'off', or from 'off' to 'on'. Each measurement is timestamped based on seconds relative to the enrollment of the study subject into the movisensXS app. Display On/Off data is captured in the study to measure how much time the study participant spends on their phone.

Relationship to other data elements in the study

Display On/Off data is collected in relation to each study participant, encoded in the IMMERSE_Patient FHIR profile. Each Display On/Off data instance is linked to its study participant via a FHIR reference to the respective IMMERSE_Patient instance. There are no explicit references to other IMMERSE data elements, but timestamps can be used to relate Display On/Off instances to instances of other IMMERSE data instances.

Attributes of the data element

The following table lists the individual attributes of the FHIR resource, including short descriptions and semantic encodings for attributes and value sets, where applicable:

Attribute Name	Attribute Datatype	Attribute Description	Semantic encoding
identifier	integer	Identifier of the sensing measurement within the study	-
status	value set (ObservationStatus)	Status of the measurement in the study	-
code	codeableConcept (CodeableConcept)	Type of measurement	**DISPLAY_ON_OFF**
subject	integer	Unique pseudonymous identifier of the research subject within the study (reference to the IMMERSE_Patient resource)	-
effective	datetime	date and time of measurement	-
display_onoff	integer	state of display (on or off)	**value set:** ● **DISPLAY_OFF** ● **DISPLAY_ON**

FHIR Representation of the data element

For the FHIR representation of Device On/Off data, the FHIR Observation resource was chosen. In the course of profiling this resource, irrelevant data elements (e.g. related to clinical care) were removed. Since no adequate semantic encoding for this attribute and its value set could be retrieved from LOINC or SNOMED, an alphanumeric code "DISPLAY_ON_OFF" was designated for the attribute, and the original coding from movisensXS was maintained for the value set.

The resulting IMMERSE_DisplayOnOff FHIR profile is available online on Simplifier:

https://simplifier.net/immerse-eu/immerse_displyonoffobservation

A static screenshot of the IMMERSE_DisplayOnOFF FHIR profile is shown in Figure 11:

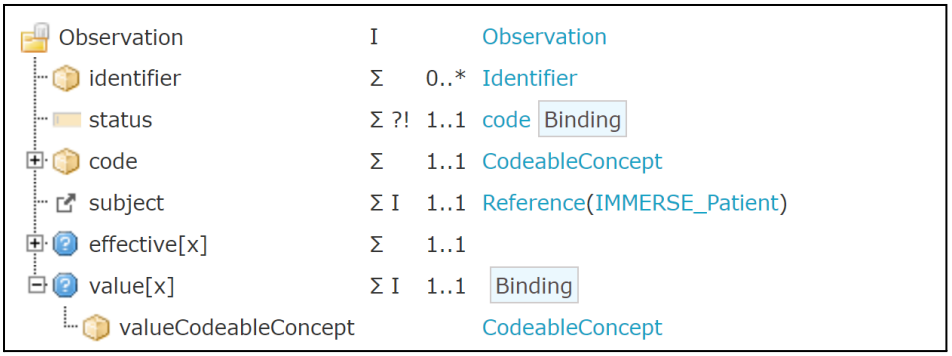

Fig. 11: Screenshot of IMMERSE_DisplayOnOff FHIR profile

Data protection aspects of the FHIR profile

Display on/off data is collected by means of timestamped status when the subject's mobile phone display was switched on or off. For re-identification, an attacker would need to have access to a dataset with similar display on/off and timestamp data (e.g. from an app installed at the same device at the same time) and develop an algorithm to match it to the IMMERSE data. In view of the benefits of re-identification and related effort, this approach is not considered a relevant privacy risk. However, in the context of data provision for data use projects, additional measures (e.g. conversion of timestamps to relative durations since study enrolment) will be considered in the context of the specific requirements of the project.

Implementation aspects of the FHIR profile

Source data for this data element is produced by the movisensXS application. The following aspects need to be taken into account during implementation of the transformation from movisensXS raw data to FHIR resources:

- Attribute "effective": movisensXS provides a relative timestamp for the measurement in seconds since enrollment into the study (integer) which needs to be converted to a regular absolute date & time datatype during transformation to FHIR.
- Attribute "value": the value set could not be semantically encoded based on available standardized terminologies. The movisensXS values were kept as-is.

movisensXS source attribute	FHIR profile attribute	Transformation
display_rel_timestamp	effective	conversion to absolute date/timestamps
display_onoff	value	mapping to value set

2.6.6. IMMERSE_AppUsageObservation

Role of data element in the study

App Usage is on the sensing data elements that can be acquired through the movisensXS app if configured on a study subject's compatible mobile phone. Recording of App Usage is triggered whenever each app of study subject's mobile phone is logged and shown up on the screen. Each measurement is timestamped based on seconds relative to the enrollment of the study subject into the movisensXS app. movisensXS does not log all apps that a study participant uses, but needs a specific "whitelist" of apps to be logged, which is configured on the device upon enrollment. WP7 needs to provide a list of app categories as well as country-specific app identifiers within these categories to allow usable data collection for this data element. Assignment of categories will occur at App Usage data is captured in the study to estimate the purpose of using the mobile phone which could be useful to detect categories of apps the study participant often used.

Relationship to other data elements in the study

App Usage data is collected in relation to each study participant, encoded in the IMMERSE_Patient FHIR profile. Each App Usage data instance is linked to its study participant via a FHIR reference to the respective IMMERSE_Patient instance. There are no explicit references to other IMMERSE data elements, but timestamps can be used to relate AppUsage instances to instances of other IMMERSE data instances.

Attributes of the data element

The following table lists the individual attributes of the FHIR resource, including short descriptions and semantic encodings for attributes and value sets, where applicable:

Attribute Name	Attribute Datatype	Attribute Description	Semantic encoding
identifier	integer	Identifier of the sensing measurement within the study	-
status	value set (ObservationStatus)	Status of the measurement in the study	-
code	codeableConcept (CodeableConcept)	Type of measurement	snomedct:706689003
subject	integer	Unique pseudonymous identifier of the research subject within the study (reference to the IMMERSE_Patient resource)	-
effective	datetime	date and time of measurement	-
Component.appCategory.value	string	Category of the app the study subject used	APP_CATEGORY
Component.appName.value	string	name of app the study subject used	APP_NAME

FHIR Representation of the data element

For the FHIR representation of app usage data, the FHIR Observation resource was chosen. In the course of profiling this resource, irrelevant data elements (e.g. related to clinical care) were removed. Adequate semantic encoding for the attribute was retrieved in SNOMED, but could not be found in

LOINC or SNOMED for the values (App category, App name), so alphanumeric codes “APP_CATEGORY” and “APP_NAME” were designated.

The resulting IMMERSE_AppUsage FHIR profile is available online on Simplifier:

https://simplifier.net/immerse-eu/immerse_appusageobservation

A static screenshot of the IMMERSE_AppUsage FHIR profile is shown in Figure 12:

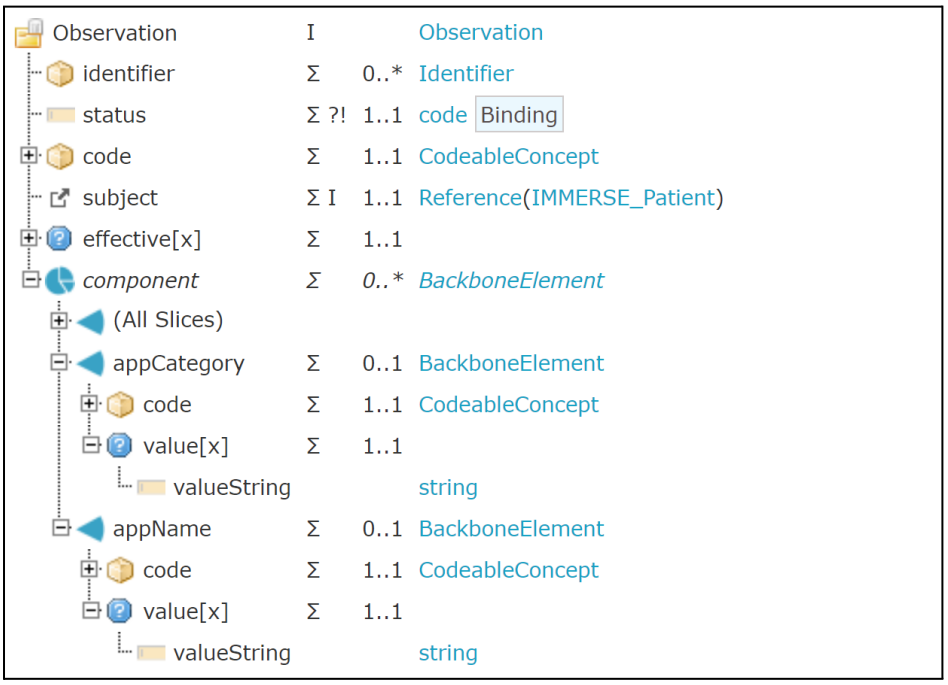

Fig. 12: Screenshot of IMMERSE_AppUsage FHIR profile

Data protection aspects of the FHIR profile

App usage data is collected by means of timestamped broad categories of applications used by the subject. For re-identification, an attacker would need to have access to a dataset with similar app usage and timestamp data (e.g. from an app installed at the same device at the same time) as well as the mapping of country-specific apps to the categories used in IMMERSE and develop an algorithm to match it to the IMMERSE data. In view of the benefits of re-identification and related effort, this approach is not considered a relevant privacy risk. However, in the context of data provision for data use projects, additional measures (e.g. conversion of timestamps to relative durations since study enrolment) will be considered in the context of the specific requirements of the project.

Implementation aspects of the FHIR profile

Source data for this data element is produced by the movisensXS application. The following aspect needs to be taken into account during implementation of the transformation from movisensXS raw data to FHIR resources:

- Attribute “effective”: movisensXS provides a relative timestamp for the measurement in seconds since enrollment into the study (integer) which needs to be converted to a regular absolute date & time datatype during transformation to FHIR.
- Attribute “appCategory.value”: the category is not available from movisensXS, but needs to

be assigned at conversion time from the WP7 table of app categories/app names to be logged.

movisensXS source attribute	FHIR profile attribute	Transformation
steps_rel_timestamp	effective	conversion to absolute date/timestamps
-	Component.appCategory.value	Derived from WP7 app category/name list at conversion time
appUsage_name	Component.appName.value	-

2.6.7. IMMERSE_NotificationObservation

Role of data element in the study

Notification is on the sensing data elements that can be acquired through the movisensXS app if configured on a study subject's compatible mobile phone. Recording of Notifications is triggered whenever an app sends a notification to study subject's mobile phone. Notifications could be triggered both by receiving a text message or by alerts generated by apps. Like the AppUsageObservation, apps needs to be explicitly whitelisted and categorized to have their notifications logged. Each measurement is timestamped based on seconds relative to the enrollment of the study subject into the movisensXS app. Notification data is captured in the study to estimate the study participants use habits of their mobile phone besides App Usage data.

Relationship to other data elements in the study

Notification data is collected in relation to each study participant, encoded in the IMMERSE_Patient FHIR profile. Each Notification data instance is linked to its study participant via a FHIR reference to the respective IMMERSE_Patient instance. There are no explicit references to other IMMERSE data elements, but timestamps can be used to relate Notification instances to instances of other IMMERSE data instances.

Attributes of the data element

The following table lists the individual attributes of the FHIR resource, including short descriptions and semantic encodings for attributes and value sets, where applicable:

Attribute Name	Attribute Datatype	Attribute Description	Semantic encoding
identifier	integer	Identifier of the sensing measurement within the study	-
status	value set (ObservationStatus)	Status of the measurement in the study	-
code	codeableConcept (CodeableConcept)	Type of measurement	snomedct:706689003
subject	integer	Unique pseudonymous identifier of the research subject within the study (reference to the	

		IMMERSE_Patient resource)	
effective	datetime	date and time of measurement	
notificationAppCategory.value	string	category of the app as supplied by WP7	
Component.notificationAppName.value	string	name of the app that sent notification	
Component.notificationLength.value	integer	Length of the content of the notification	snomedct:410668003

FHIR Representation of the data element

For the FHIR representation of Notification data, the FHIR Observation resource was chosen. In the course of profiling this resource, irrelevant data elements (e.g. related to clinical care) were removed. To cover the individual Physical activity attributes, the components branch of the resource was sliced to explicitly represent values for notificationAppName and notificationLength. Adequate semantic encodings were searched in LOINC and SNOMED and could be retrieved for the attribute and the length value, but not for the NotificationAppCategory and NotificationAppName values. For these, the alphanumeric codes “APP_CATEGORY” and “APP_NAME” were designated.

The resulting IMMERSE_Notification FHIR profile is available online on Simplifier:

https://simplifier.net/immerse-eu/immerse_notificationobservation

A static screenshot of the IMMERSE_Notification FHIR profile is shown in Figure 13:

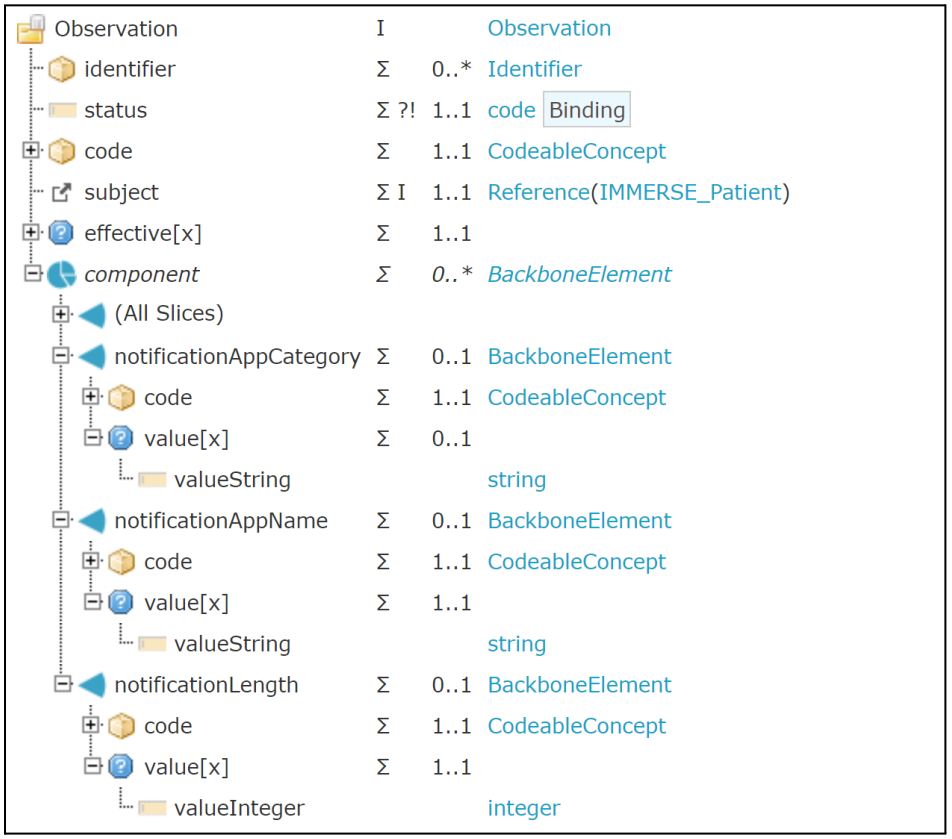

Fig. 13: Screenshot of IMMERSE_Notification FHIR profile

Data protection aspects of the FHIR profile

Notification data is collected by means of timestamped app and message lengths of notifications sent the subject. For re-identification, an attacker would need to have access to a dataset with similar notification app, message length and timestamp data (e.g. from an app installed at the same device at the same time) and develop an algorithm to match it to the IMMERSE data. In view of the benefits of re-identification and related effort, this approach is not considered a relevant privacy risk. However, in the context of data provision for data use projects, additional measures (e.g. conversion of timestamps to relative durations since study enrolment) or generalizing the message length variable will be considered in the context of the specific requirements of the project.

Implementation aspects of the FHIR profile

Source data for this data element is produced by the movisensXS application. The following aspects need to be taken into account during implementation of the transformation from movisensXS raw data to FHIR resources:

- Attribute “effective”: movisensXS provides a relative timestamp for the measurement in seconds since enrollment into the study (integer) which needs to be converted to a regular absolute date & time datatype during transformation to FHIR.
- Attribute “NotificationAppCategory.value”: the category is not available from movisensXS, but needs to be assigned at conversion time from the WP7 table of app categories/app names to be logged.
- Attribute “notificationAppName”: movisensXS provides this data combined with attribute “notificationLength” which needs to be pre-processed by dividing both attributes and converting to integer data type.
- Attribute “notificationLength”: movisensXS provides this data as string data type combined with attribute “notificationAppName”, which needs to be pre-processed by dividing both attributes and converting to integer data type.

movisensXS source attribute	**FHIR profile attribute**	**Transformation**
activity_rel_timestamp	effective	conversion to absolute date/timestamps
-	component.notificationAppCategory.value	
Notification	component.notificationAppName	Preprocessing (split of attribute at pipe symbol)
notification	component.notificationLength	preprocessing (split of attribute at pipe symbol) and conversion to integer

3. **References**

1. Wilkinson MD, Dumontier M, Aalbersberg I, Appleton G et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018. doi: 10.1038/sdata.2016.18.
2. Sinaci AA, Nunez-Benjumea F, Gencturk M, et al. From Raw Data to FAIR Data: The FAIRification Workflow for Health Research. Methods Inf Med. 2020;59(S 01):e21-e32. doi: 10.1055/s-0040-1713684.