

WP3 Data Management, Data Security and Interoperability

Thomas Ganslandt
Friedrich-Alexander-University Erlangen-Nuremberg

EC Evaluation 19.06.2024



This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP3 - Staff



**Thomas
Ganslandt**

Physician (MD)
Medical Informatics

PI UHEI/UKER

**Association of WP3 to
CUBiDA team:**

Core Unit for Bioinformatics,
Data Integration & Analysis

Unit at Erlangen University Hospital
establishing central research data
management services



Geun-Hyun Kim

Physics (B.A.)
Psychology (Dipl.)

Data Science &
Usability



**Tariq
Elahi**

Computer Scientist (PhD)

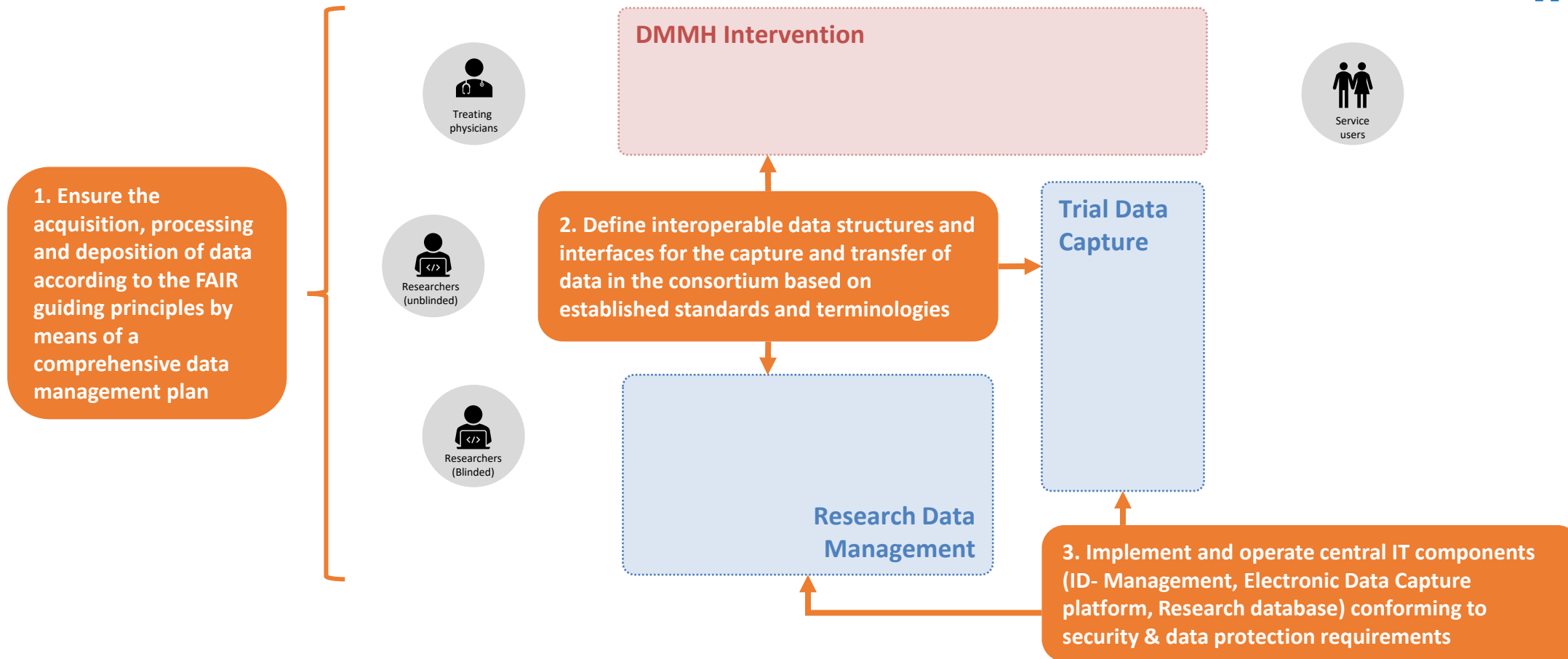
Privacy & IT Security

30 PM UHEI => UKER

2.4PM UEDIN



WP3 - Objectives



Task 3.1: FAIR Data Management (1)



- **Development of a Data Management Plan (DMP)**
 - Inventory of data sets, data items and potential artifacts
 - Suitable metadata annotations & terminologies
 - Data Sharing & integration of Data Governance Policy, suitable license
 - Contribution to patient data protection & security aspects
- **Maintenance of the DMP throughout the project**
- **Deposition & long-term archiving of datasets & artifacts**



Task 3.1: FAIR Data Management (2)



- **Status after second period**
 - second release of DMP has been prepared & will be submitted after revision of first release has been approved
 - addition of DataCite metadata schema for annotation of datasets & artifacts
 - addition of phase II data capture aspects
 - extension of inventory with phase II datasets
 - semantic annotation of sensor data completed
 - semantic annotation of forms data ongoing
 - annotation of forms complete
 - annotation of form items/valuesets at 25%

IMMERSE
Implementing Mental Health Recording Strategy for Europe
H2020-945263

Data Management Plan	
DL1	
Documentation level	PD
Contractual date of delivery	06.10.2023
Actual date of delivery	06.10.2023
Type	6
Work package	1
Work package lead	F. Dierker, Göttingen

Consortium

Organization	Name	Contact information
1 - LMU	F. Dierker, Göttingen	F.dierker@med.uni-goettingen.de
2 - ICI	Lucy Wipacit	lucy.wipacit@icimn.ie
3 - ICI	ICIMN, Birmingham	icimn@birmingham.ac.uk
4 - ICI	Maria Wolkosz	maria.wolkosz@icimn.ie
5 - ICI	Simon Krause	simon.krause@bham.ac.uk
6 - ICI	Simon Krause	simon.krause@bham.ac.uk

Data Summary

- Purpose of data collection
- Data types and formats
- Re-use of existing data
- Data utility

FAIR Data

- Making data findable and accessible
- Making data interoperable & reusable
- Increase data re-use

Allocation of resources

Data security

Ethical aspects

Conclusions

References

Appendices

- Inventory of datasets & artifacts



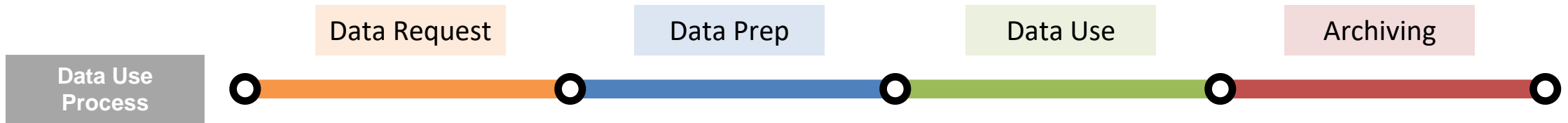
Task 3.1: FAIR Data Management (3)



- **Data Storage & Archiving**
 - UKER Cloud: storage of internally used datasets
 - FAU RADAR: storage of publicly available datasets, assignment of DOIs, registration in catalog of FAU library
 - encrypted archiving of datasets with person-related datasets
- **Provision of access to project datasets**
 - contribution to data governance process (WP6)
 - data use process shown in detail in the following slides



Task 3.1: FAIR Data Management (4)



Pre-registration of draft proposal

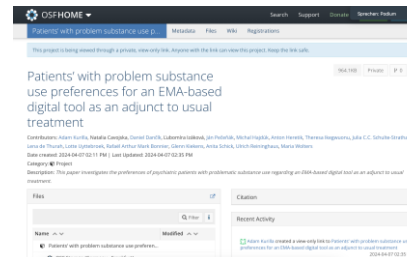
Pre-approval by DGB

Pre-registration of full proposal

Approval by DGB



- Commenting by participants
- preliminary data availability check by WP3

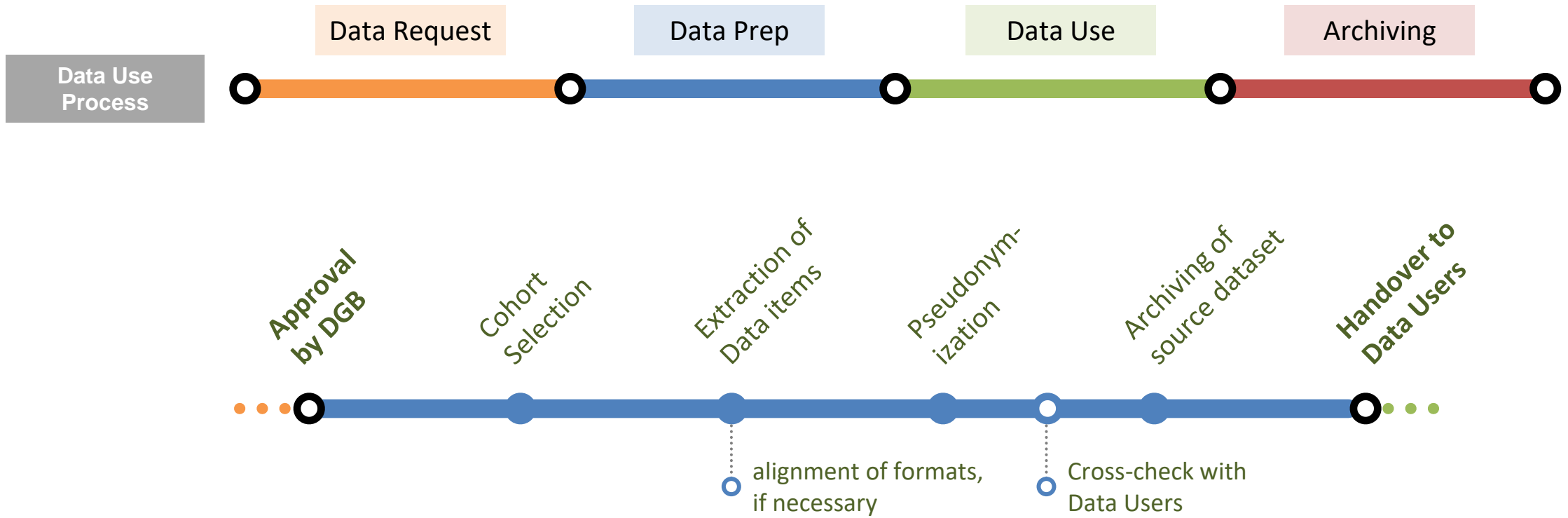


- Commenting by participants
- final data availability check by WP3



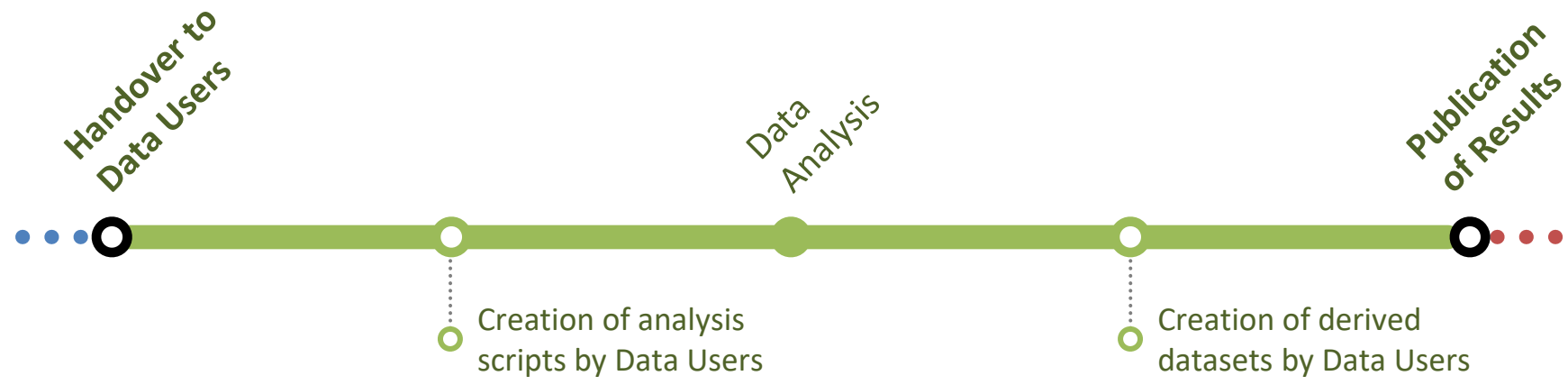
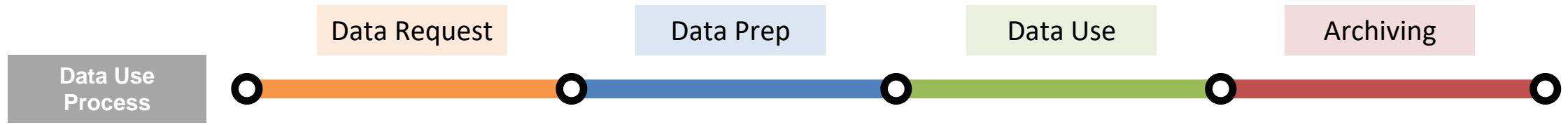
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Task 3.1: FAIR Data Management (4)



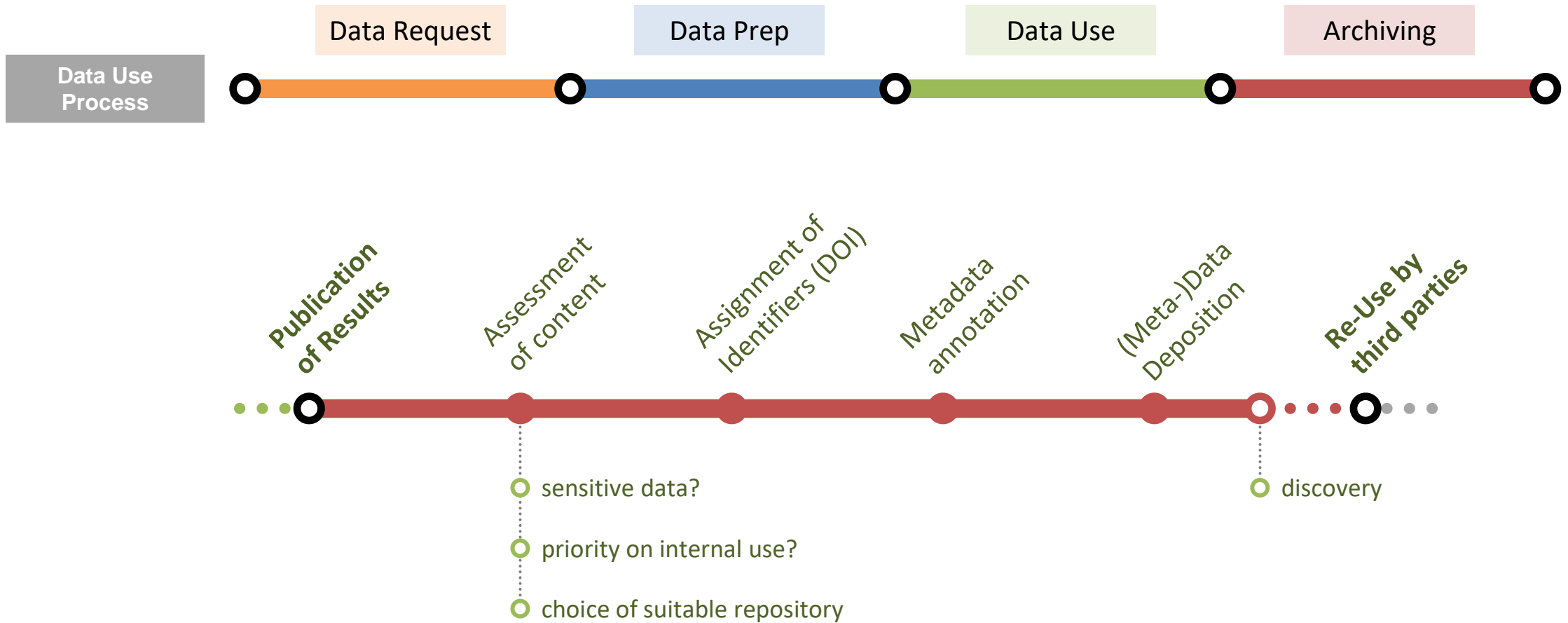
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Task 3.1: FAIR Data Management (4)



This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Task 3.1: FAIR Data Management (4)



This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Task 3.2: Interoperable Data Structures (1)



- **Goal: Interoperability with international clinical/scientific data platforms**
 - by use of internationally adopted data structures & terminologies
 - e.g. German Medical Informatics Initiative
- **Development of an Implementation guide**
 - based on the data inventory
 - specification of data structures based on HL7 FHIR standard
 - semantic annotation with terminologies like LOINC & SNOMED CT
- **Design of an architecture for the central IT platform**
- **Maintenance of the Implementation guide throughout the project**

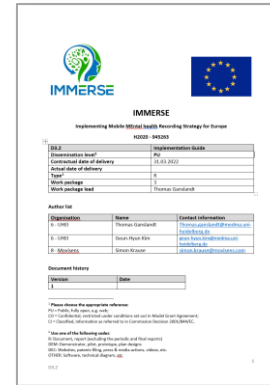


Task 3.2: Interoperable Data Structures (2)



- **Status after first period**

- Implementation Guide deliverable released on time
- covers architecture, data structures for questionnaires (from eCRF, movisensXS & MoMent) & sensor data
- explicitly covers sensor data protection aspects



- **Ongoing work**

- second release with updated architecture and detailed description of data flows has been completed and will be submitted once the revision of the first release has been accepted
- transformation of raw data to interoperable FHIR representation

Approach for the Implementation Guide Architecture

Data Elements for Study & Participants

- IMMERSE_ResearchSubject
- IMMERSE_Patient
- IMMERSE_Organization

Data Elements for Questionnaires & Responses

- IMMERSE_Questionnaire
- IMMERSE_QuestionnaireResponse
- IMMERSE_QuestionnaireResponseObservation

Data Elements for Mobile Sensor Data

- IMMERSE_GeolocationObservation
- IMMERSE_ActivityObservation
- IMMERSE_StepsObservation
- IMMERSE_DeviceOnOffObservation
- IMMERSE_DisplayOnOffObservation
- IMMERSE_AppUsageObservation
- IMMERSE_NotificationObservation



This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

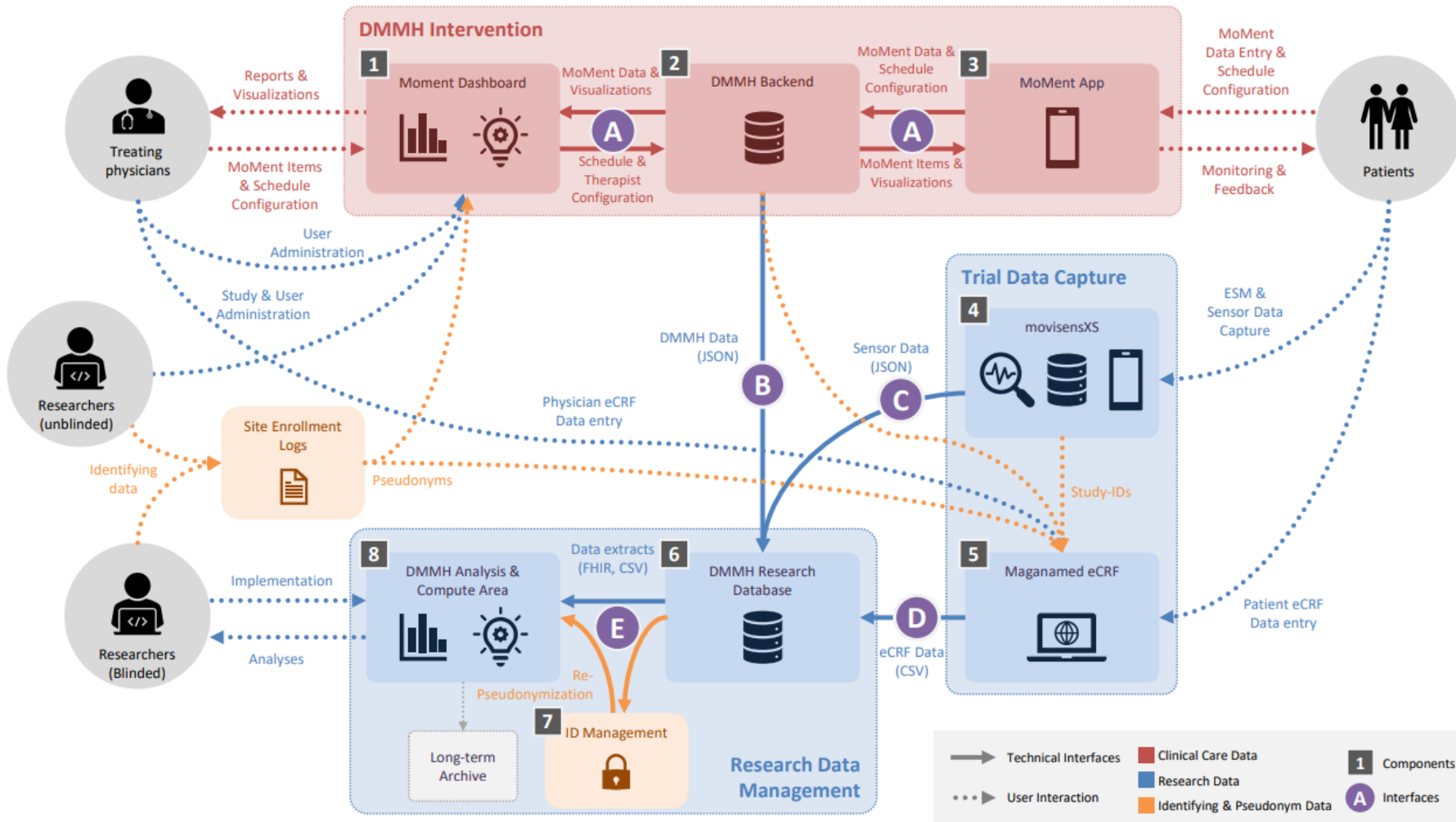
Task 3.3: Implementation & Operation of Central IT (1)



- **Scope:**
 - Integration of Data Capture platforms, ID-Management, Research Database
- **Interfaces for data import**
 - Movisens platform (movisensXS, MoMent), Maganamed eCRF
- **Interfaces for data use**
 - data provision for analysis, including execution (optional) of analysis code
 - data deposition to longterm archiving
- **Provision/Maintenance of platform in 3 releases throughout project**
 - (1) extraction of eCRF & sensor data
 - (2) initial version of research database (eCRF + sensor data) & re-pseudonymization
 - (3) second version of research database (+DMMH data)



Task 3.3: Implementation & Operation of Central IT (2)



This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

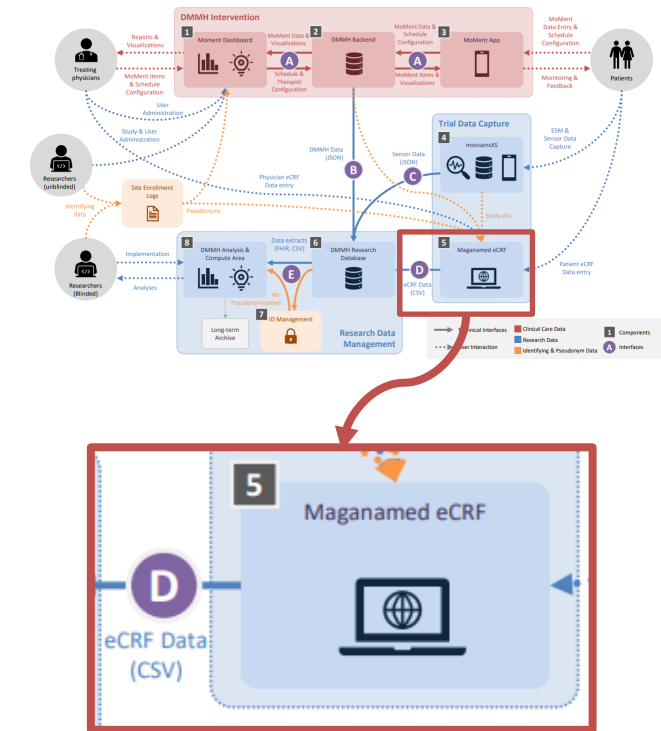


Task 3.3: Implementation & Operation of Central IT (3)



MaganaMed

- Content: Participant eCRFs
- Status: interface in production use
 - weekly manual exports & backup at UKER
 - extraction & transformation of data dictionary
- Formats:
 - source: CSV (1 file per questionnaire)
 - available: CSV
 - dummy data available (CSV)

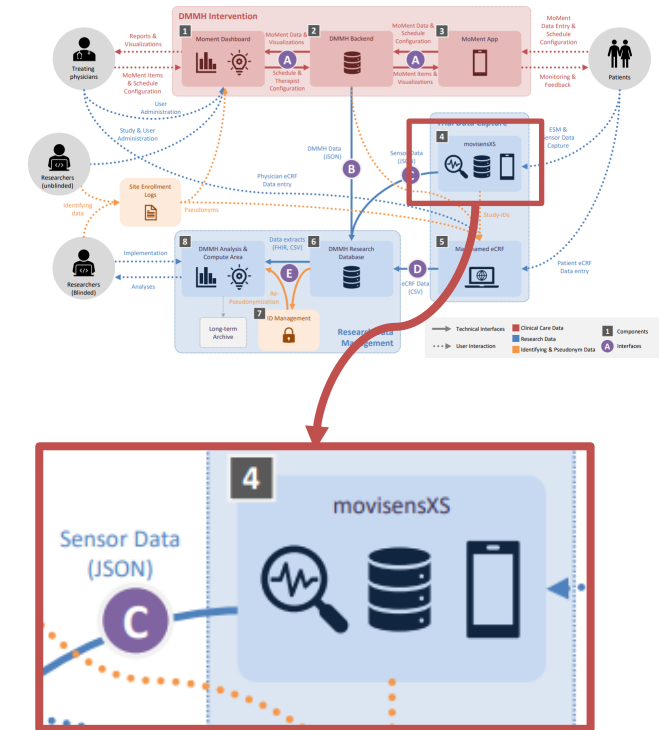


Task 3.3: Implementation & Operation of Central IT (4)



movisensXS

- Content:
 - sensing data (7 modalities)
 - ESM & fidelity forms
- Status: interface in production use
 - weekly manual exports & backup at UKER
- formats:
 - source: CSV (sensing, 1 file per modality/participant), JSON & Excel (ESM & fidelity)
 - available: CSV, Excel
 - dummy data: available for ESM forms (CSV)

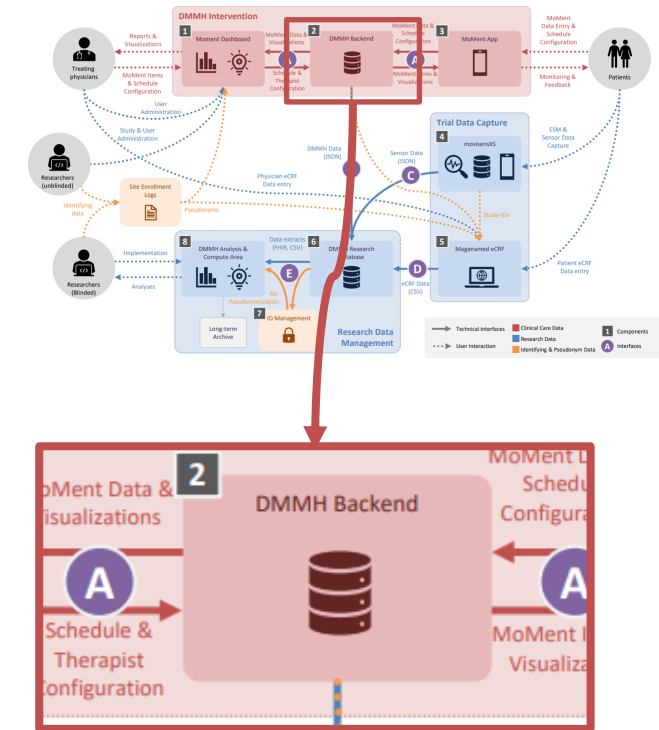


Task 3.3: Implementation & Operation of Central IT (5)



TherapyDesigner/MoMent

- Content:
 - primary ESM forms
 - prompt & usage data
- Status: interface implemented, but currently in revision
 - manual exports available on request by movisens
- formats:
 - source: JSON (metadata file + 1 file per site)
 - available: CSV
 - dummy data: not required

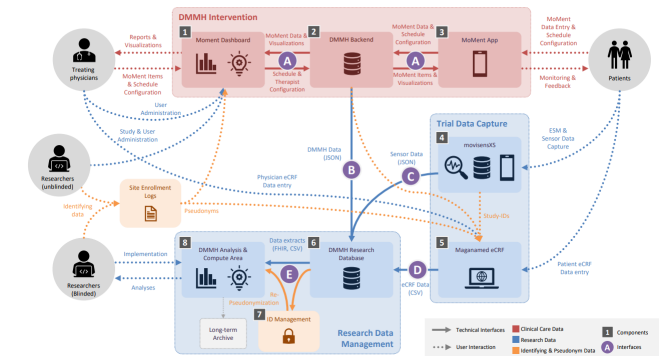


Task 3.3: Implementation & Operation of Central IT (6)



Additional Data Sources

- Phase I REDCap eCRF data
 - available in REDCap source systems
- Phase I Qualitative Data
 - available in Erlangen (de-identified transcripts, QPDX files)
- Erlangen REDCap
 - used for trial coordination
 - weekly manual export
 - CSV files (1 per site & eCRF)

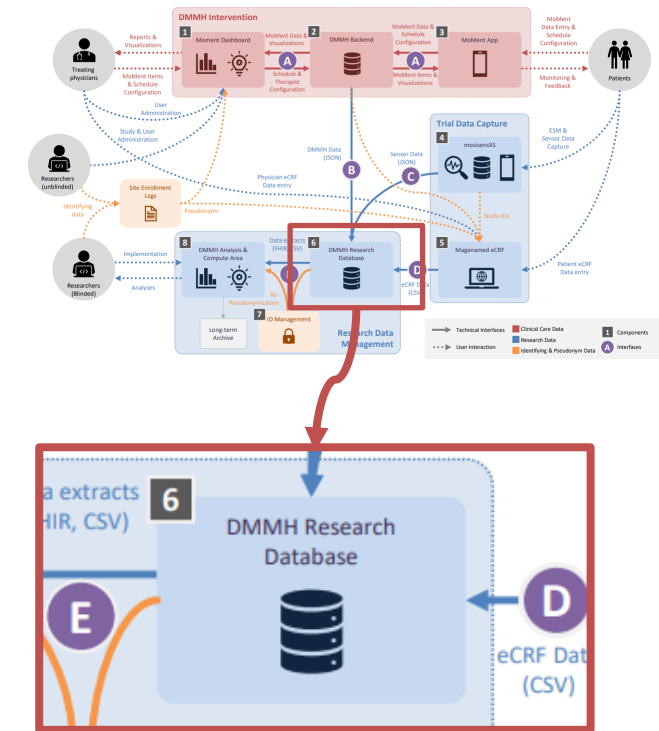


Task 3.3: Implementation & Operation of Central IT (7)



Research Database

- Content:
 - collection of data from all IMMERSE sources
 - to be used for cohort selection and data extraction across all IMMERSE phase II data sources
- Status:
 - file-based exports from source systems
 - import of MaganaMed & movisensXS data into SQL database in production
 - import of TherapyDesigner data in development

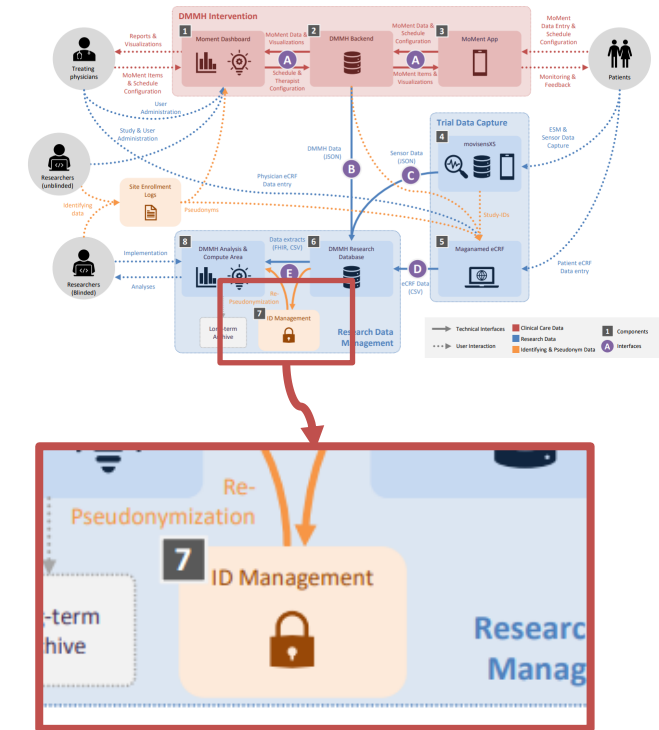


Task 3.3: Implementation & Operation of Central IT (8)



Re-Pseudonymisation

- Purpose:
 - secondary pseudonymisation of datasets for analysis
- Status: in production use
- concept:
 - carried out by independent trust center of UKER using gPAS software
 - replacement of original study pseudonyms with individually generated set of random IDs for each data use project
 - assignment list is used by WP3 only during data preparation
 - assignment list is archived by trust center for legitimized access

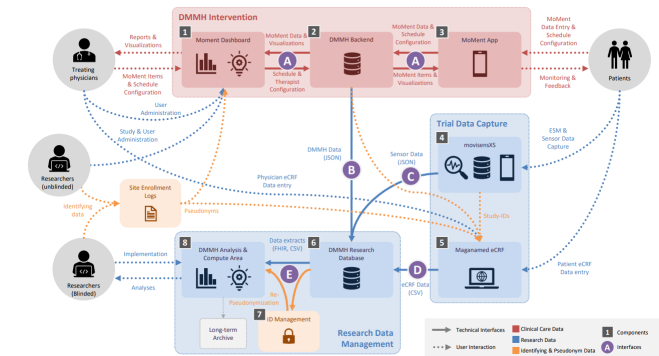


Task 3.3: Implementation & Operation of Central IT (9)



Additional Infrastructure

- UKER Cloud
 - shared data space hosted securely at UKER
 - ingest of data provided by other partners (e.g. Phase I data)
 - provision of exports for data use projects
- Status Dashboard
 - graphical overview of data entry status by site & eCRF
- Dummy data
 - “synthetic” datasets to be used for analysis script development
 - based on trial metadata (e.g. MaganaMed data dictionary)

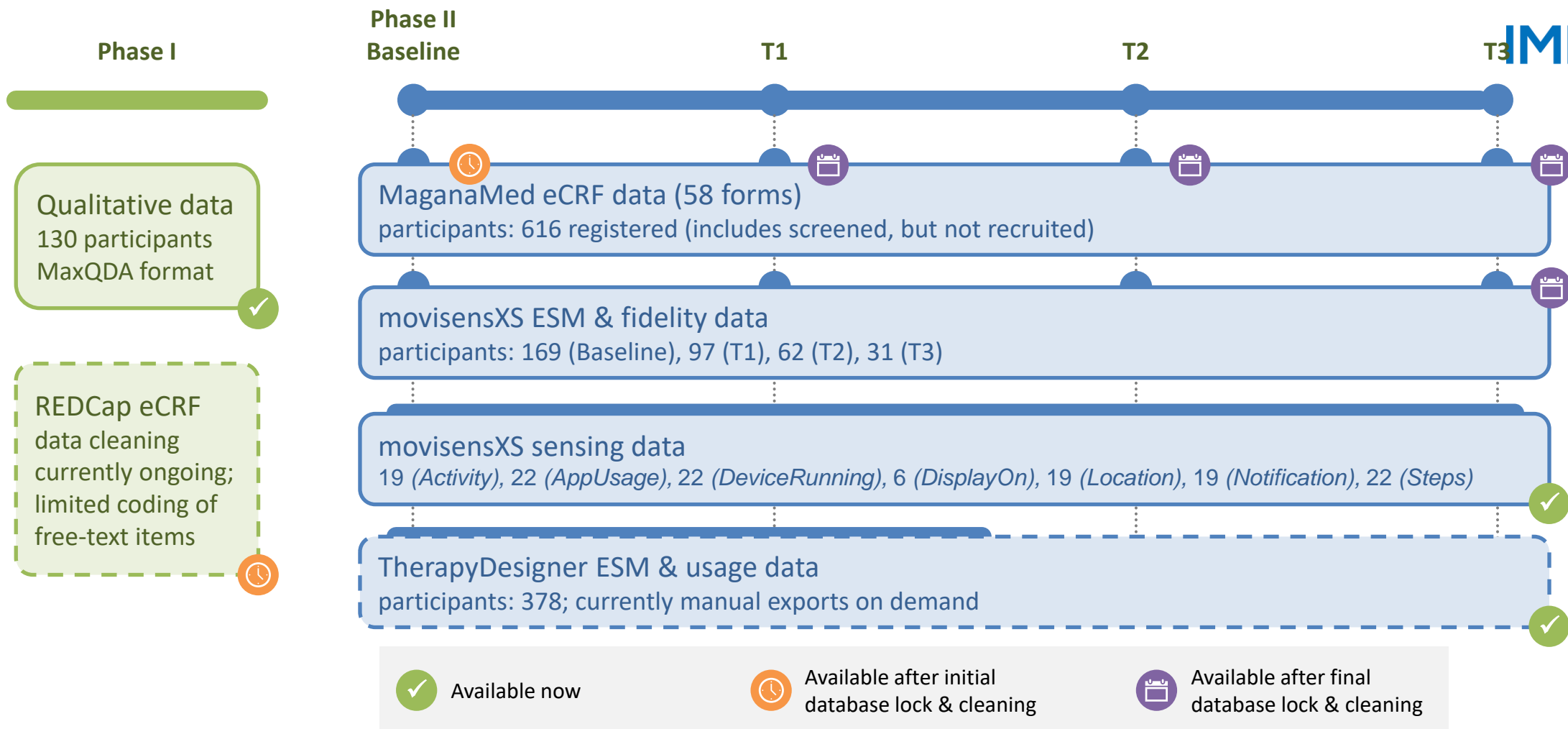


Task 3.3: Implementation & Operation of Central IT (10)

Data Availability



IMMERSE



This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP3 – Deliverables & Milestones



Milestone / deliverable	Title	Original deadline	Status
D3.1	Data Management Plan	10/2021	✓
D3.2	Implementation Guide	04/2022	✓
D3.3	Implementation Report	03/2024	✓
MS9	First central IT platform release	10/2022	✓
MS17	Second central IT platform release	09/2023	✓
MS25	Third central IT platform release	09/2024	ongoing



This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP3 – Next steps



- **T3.1**
 - second release of DMP
 - metadata annotation & deposition of upcoming result datasets
- **T3.2**
 - second release of implementation guide
 - semantic annotation of phase II trial data elements
- **T3.3**
 - collaboration with WP7 for database lock & data cleaning
 - finalization of research database (=> 3rd platform release)
 - Transformation of phase I & II data elements to FHIR

