

# **WP3 Data Management, Data Security and Interoperability**

**Thomas Ganslandt**  
**Friedrich-Alexander-University Erlangen-Nuremberg**

**General Assembly Bratislava, 04.05.2023**



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

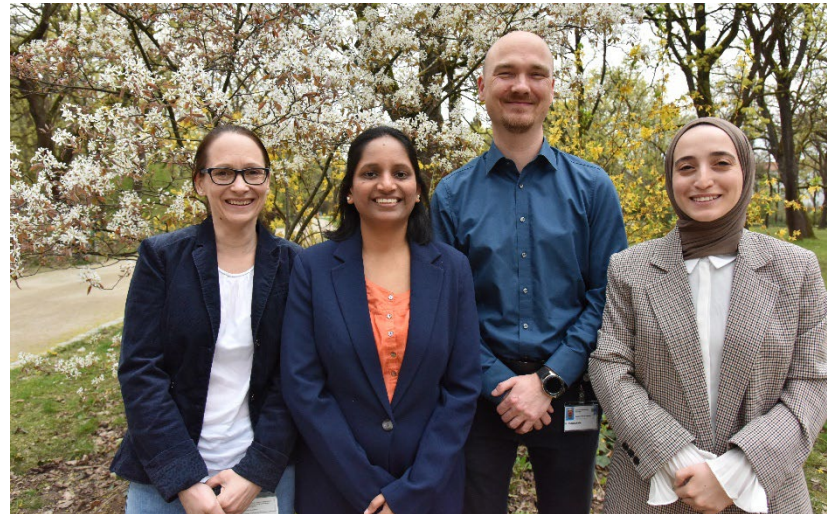
Unit at Erlangen University Hospital establishing  
central research data management resources



**Geun-Hyun Kim**

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

Data Science &  
Usability



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 (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
  - Patient data protection & security aspects, including threat analysis
- **Maintenance of the DMP throughout the project**
- **Deposition & long-term archiving of datasets & artifacts**



# Task 3.1: FAIR Data Management (2)



- **Current status & ongoing work**
  - second release of Data Management Plan submitted
    - addition of detailed phase I & II datasets to inventory
    - update on phase II data capture
  - liaising with WP5 for provision of phase I datasets
- **next steps**
  - servicing data use requests from phase I data
  - deposition of phase I data



# Task 3.2: Interoperable Data Structures (1)



- **Goal: Interoperability with international clinical/scientific data platforms**
  - e.g. German Medical Informatics Initiative, French Health Data Hub, SPHN
  - by use of internationally adopted data structures & terminologies
- **Development of an Implementation guide**
  - based on the data inventory & FAIR guidelines from Task 3.1
  - 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**

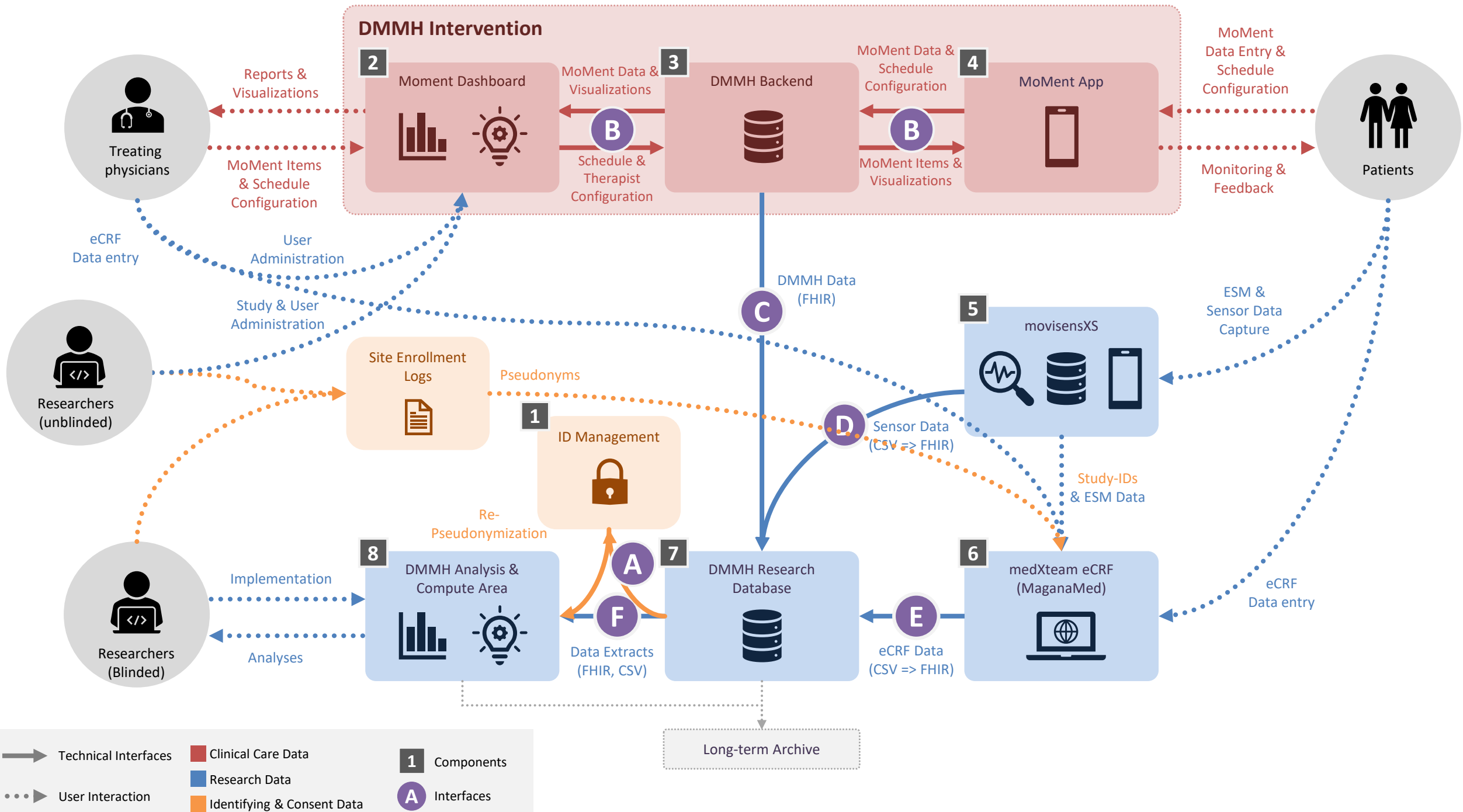


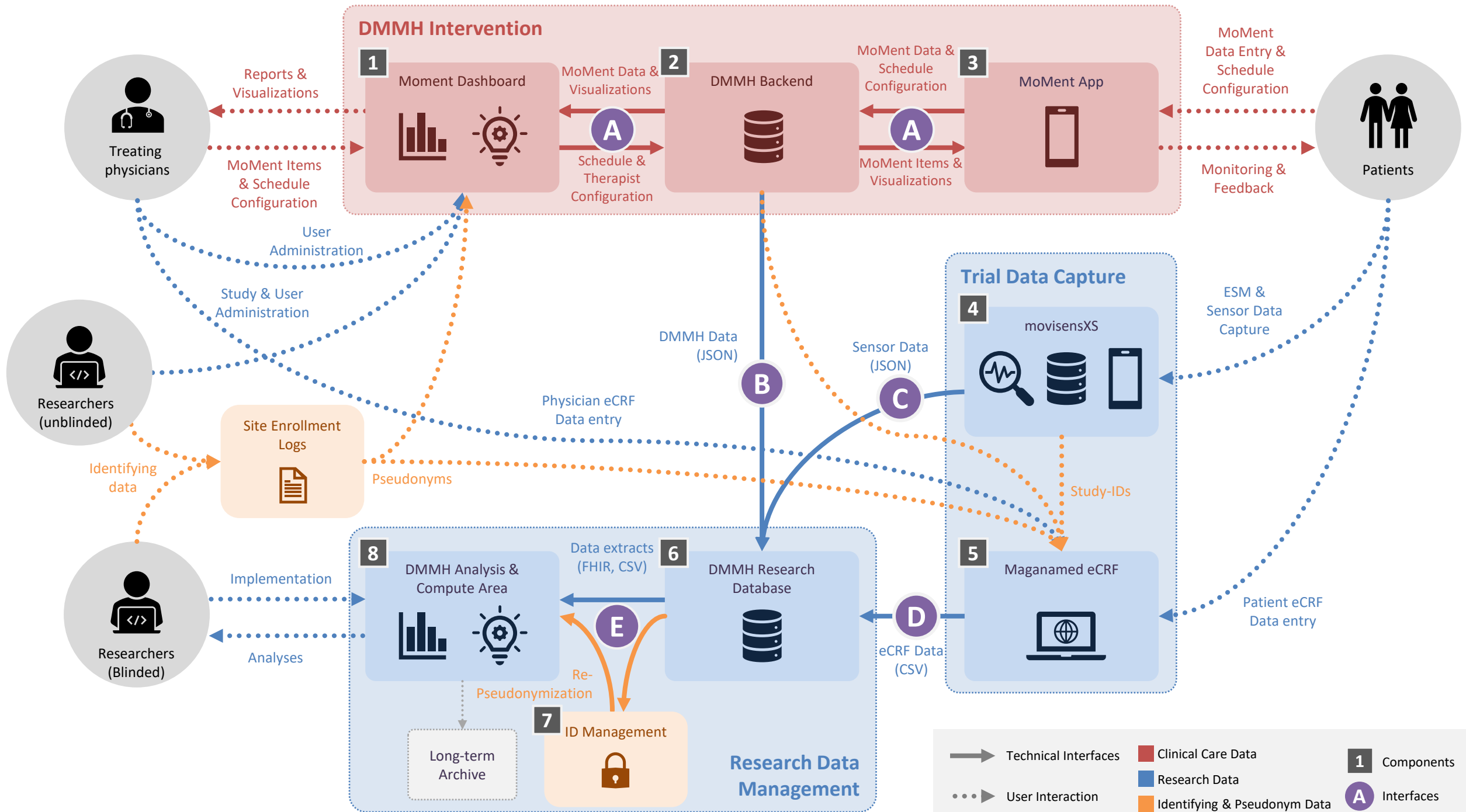
# Task 3.2: Interoperable Data Structures (2)



- **Current status & ongoing work**
  - second release of Implementation guide submitted
    - updated phase II architecture
    - extended description of data protection aspects of movisensXS datasets
  - semantic annotation of datasets
- **next steps**
  - inclusion of MoMent ESM & usage data into the implementation guide
    - as soon as the interface has been completed & data is available
  - completing metadata annotation









# Task 3.2: Interoperable Data Structures (5)



- **Focus on semantic annotation of datasets**
  - goal within FAIR framework: make datasets **findable** by assigning labels from standardized terminologies
  - relevant terminologies:
    - LOINC: Logical Observation Identifiers, Names and Codes
    - SNOMED CT: Systematic Nomenclature of Medicine
    - CDISC: Clinical Data Interchange Standards Consortium
  - nontrivial task
    - IMMERSE datasets often capture complex psychological concepts
    - terminologies often provide no exact match



# Task 3.2: Interoperable Data Structures (6)



- **Example of a simple 1:1 match: EuroQOL Five Dimension Three Level Scale**

- designation in Maganamed eCRF: EQ-5D-5L
- no match in LOINC
- good match in SNOMED CT:
  - 73041000052103: EuroQoL five dimension five level questionnaire (assessment scale)
- perfect match in CDISC controlled terminology
  - C100136: EuroQOLFive Dimension Three Level Scale
  - even provides codes for the individual items within EuroQOL
  - e.g. C100392: Question 1 – however does not describe content of item

● EuroQoL five dimension five level questionnaire (assessment scale) ★ 📄

SCTID: 73041000052103

73041000052103 | EuroQoL five dimension five level questionnaire (assessment scale) |

C100136	EuroQOLFive Dimension Three Level Scale Test Code	The EuroQOL (European Quality of Life) Five Dimension three level scale (EQ-5D-3L) (copyright 1990 EuroQol Group EQ-5D test code. All Rights Reserved) test code.
C100392	EQ-5D-3L-Question 1	The EuroQol (European Quality of Life) Five Dimension Three Level Scale (EQ-5D-3L) (copyright 1990 EuroQol Group EQ-5D. All Rights Reserved). Question 1.



# Task 3.2: Interoperable Data Structures (7)



- **Example of a complex match: Childhood Trauma Questionnaire**
  - designation in Maganamed eCRF: CTQ
  - no direct match in LOINC, SNOMED CT or CDISC
  - solution: use post-coordination to combine several terms from SNOMED CT
    - 10197000 | Psychiatric interview and evaluation (procedure) |
      - : { 363702006 | Has focus (attribute) |
      - = ( 25809009 | Victim of trauma (finding) |
      - : 246454002 | Occurrence (attribute) | = 255398004 | Childhood (qualifier value) | ) }



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



- **Based on Implementation Guide (Task 3.2) and Privacy Threat Analysis (Task 3.1)**
- **Scope:**
  - ID- & Consent Management
  - Research Database
- **Interfaces for data import**
  - Maganamed & REDCap eCRF data
  - movisensXS
  - MoMent App
- **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**



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



- **Current status & ongoing work**
  - operation of REDCap instances for secondary randomization
  - implementation of interfaces to IMMERSE data sources
    - movisensXS: automated extraction implemented (REST API, JSON & XML format)
    - Maganamed: manual export (Excel/CSV files)
      - extraction of codebook implemented
      - extraction of eCRF content ongoing
    - REDCap: manual export (CDISC ODM file)
      - extraction of codebook & eCRF content not started yet
    - DMMH: automated export (REST API, JSON format)
      - ongoing, concurrent with movisens implementation of interface
      - received API specification, currently experimenting with it
  - **next steps**
    - completion of Maganamed data extraction, implementation of quality reports
    - later on: dummy datasets, FHIR transformation of data



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



- **focus on data quality**
  - proposal to use dataquieR framework from Greifswald university
    - established for SHIP epidemiological cohort study
    - allows definition of structured data quality requirements for each data item
      - e.g. datatype & permitted values, ranges, missingness, expected distributions
      - including plausibility checks across items
    - pro: standardized approach, tool is actively developed, direct collaboration with team
    - con: very detailed report; complex configuration & report content, limited “big picture” overview
- **for discussion**
  - what level of detail is needed, and how should datasets be bundled to reduce the number of reports?
    - also: are there formal issues to check (e.g. unintended access to detailed data, unblinding)?
  - what should be covered in an (additional) „big picture“ dashboard?



# WP3 – Deliverables & Milestones



- **Deliverables**

- **D3.1: Data Management Plan (M6)**
- **D3.2: Implementation Guide for interoperable data structures and interfaces (M12)**
- **D3.3: Implementation report for 3rd platform release (M36)**

- **Milestones**

- **MS9: First central IT platform release (M18)**
- **MS17: Second central IT platform release (M30)**
- **MS25: Third central IT platform release (M42)**

