

## WP3 Data Management, Data Security and Interoperability

Thomas Ganslandt; Wolfgang Krebs  
Friedrich-Alexander-University Erlangen-Nuernberg  
Erlangen University Hospital

Steering Committee Assembly Edinburgh, 17.11.2023



# WP3 - Staff



**Thomas  
Ganslandt**

Physician (MD)  
Medical Informatics

PI UHEI/UKER



**Geun-Hyun Kim**

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

Data Science &  
Usability

**Wolfgang Krebs**

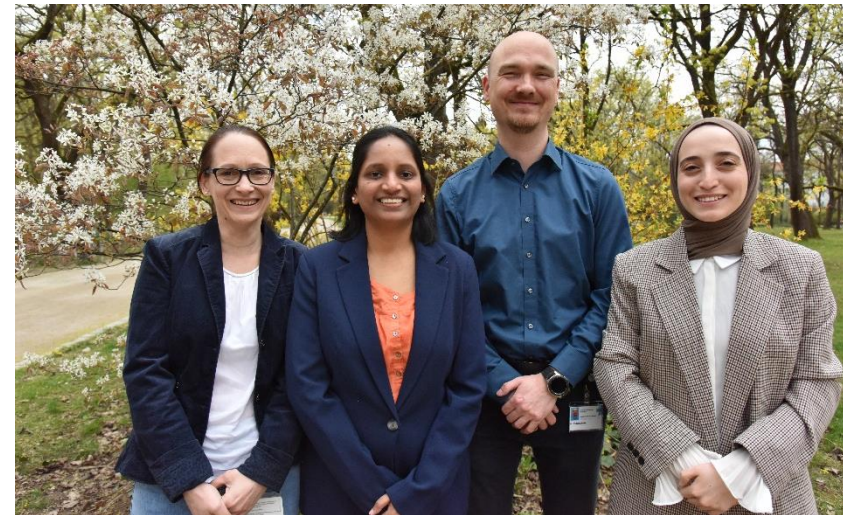
Biomedical Scientist (Phd)  
Software Developer

RDM / Bioinformatics  
at UKER

## Association of WP3 to CUBiDA team:

Core Unit for Bioinformatics, Data Integration & Analysis

Unit at Erlangen University Hospital (UKER) establishing  
central research data management resources



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

# WP3 - Objectives



- 1. Ensure the acquisition, processing and deposition of data according to the FAIR guiding principles by means of a comprehensive data management plan**
- 2. Define interoperable data structures and interfaces for the capture and transfer of data in the consortium based on established standards and terminologies**
- 3. Implement and operate central IT components (ID- and Consent Management, Electronic Data Capture platform, Research database) conforming to security & data protection requirements**



# 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**
  - Deposition of phase I data (now possible)
    - On a secure and password protected UKER storage folder
  - supported Mannheim with SOP documentation & participation in the inspection
- **next steps**
  - Servicing data use requests for 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**

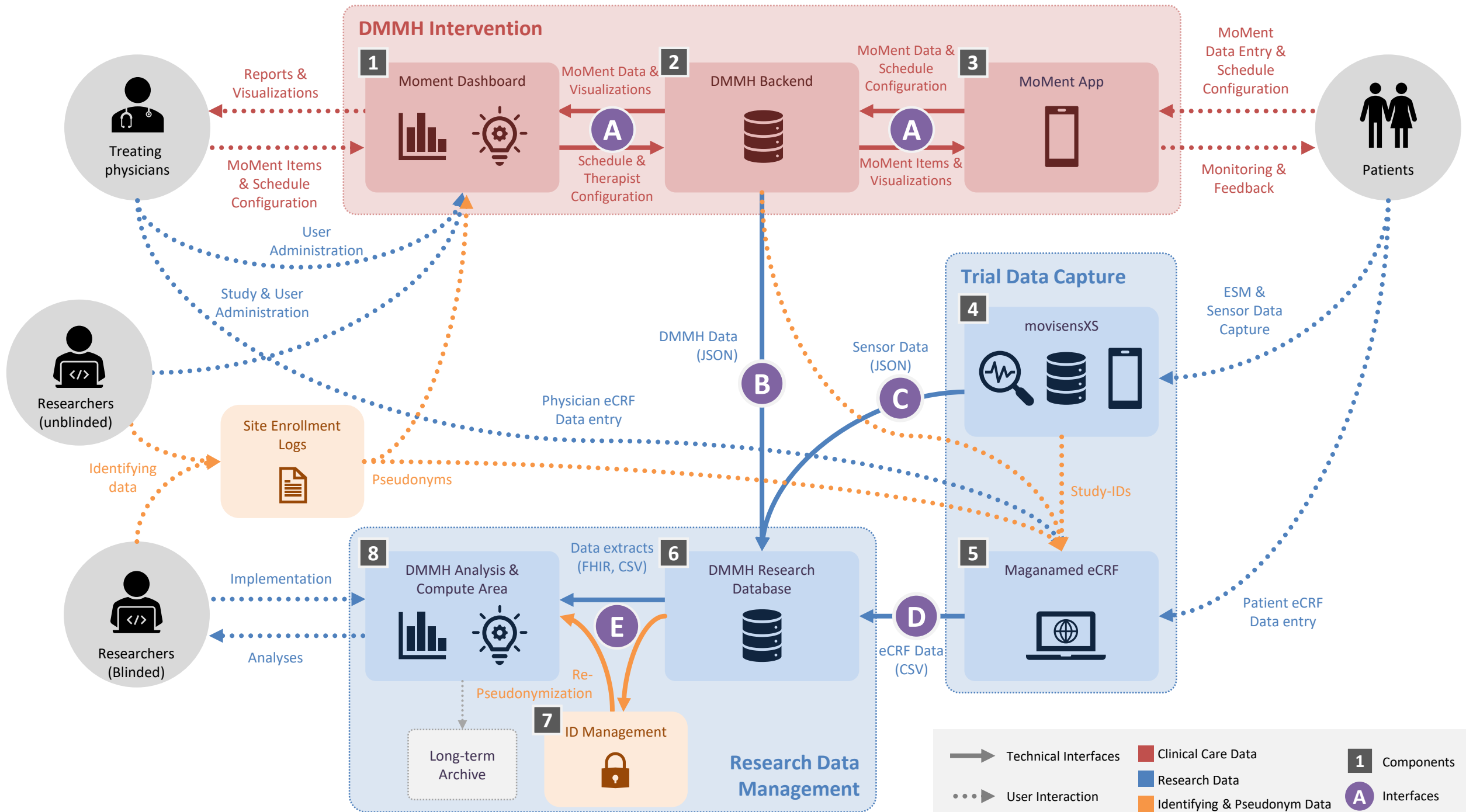


# Task 3.2: Interoperable Data Structures (2)



- **Current status & ongoing work**
  - Semantic annotation of datasets
    - On questionnaire level: Finished for all 56 questionnaires
    - On item level: More complex questionnaires still in preparation; semantic annotation for low complexity questionnaires like “Demographics” already finished
- **Next steps**
  - Inclusion of MoMent ESM & usage data into the implementation guide
    - As soon as the interface (API) has been tested & data is available
  - Completing metadata annotation







# 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 long-term archiving
- **Provision/Maintenance of platform in 3 releases throughout project**



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



- **Current status & ongoing work**
  - 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 implemented
    - REDCap: manual export (CDISC ODM file)
      - Extraction of codebook & eCRF content not started yet
    - MoMent App: automated export (REST API, JSON format)
      - Ongoing
      - Access to staging API for test database now possible, currently experimenting with it



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



- **Current status & ongoing work**
  - Dummy datasets
    - Maganamed: First Python program version available to generate dummy data
    - MovisensXS: First Python program version available to generate ESM dummy data
    - MoMent App: Ongoing
  - Establishment of quality reports (with custom programming scripts)
    - Maganamed: Completion status and fillout time (visualizations + lists)
    - MovisensXS: Completion status of questionnaires
  - Establishment of DMMH Research database
    - Custom programming script to automate import of Maganamed eCRF data into database
    - Automated integration of MovisensXS data with Maganamed data



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



- **Current status & ongoing work**
  - Data storage / sharing of IMMERSE data
    - Individually accessible folders are established on a secure UKER server to upload and share data by all locations
    - IMMERSE data and shared data is stored on a secure UKER server on a regular basis
- **Next steps**
  - Further implementation of additional quality reports
  - Integration of Maganamed, MovisensXS and MoMent App data in DMMH database
  - later on: FHIR transformation of data



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



- **status dashboard on data entry/completeness**
- **focus on data quality**
  - proposal to use dataquieR framework from Greifswald university
    - suitable for eCRF and ESM data
    - partially suitable for sensor data: will e.g. capture individual out-of-bounds data elements, but not implausible trends over time
  - implementation to be started after prioritized tasks (i.e. interface to MoMent, dummy datasets) have been finalized
- **still 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)?



# 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)**

