Session 4: Data Management Workshop

Thomas Ganslandt, Geun-Hyun Kim

Erlangen University Hospital (UKER), Dept. of Medical Information & Communication

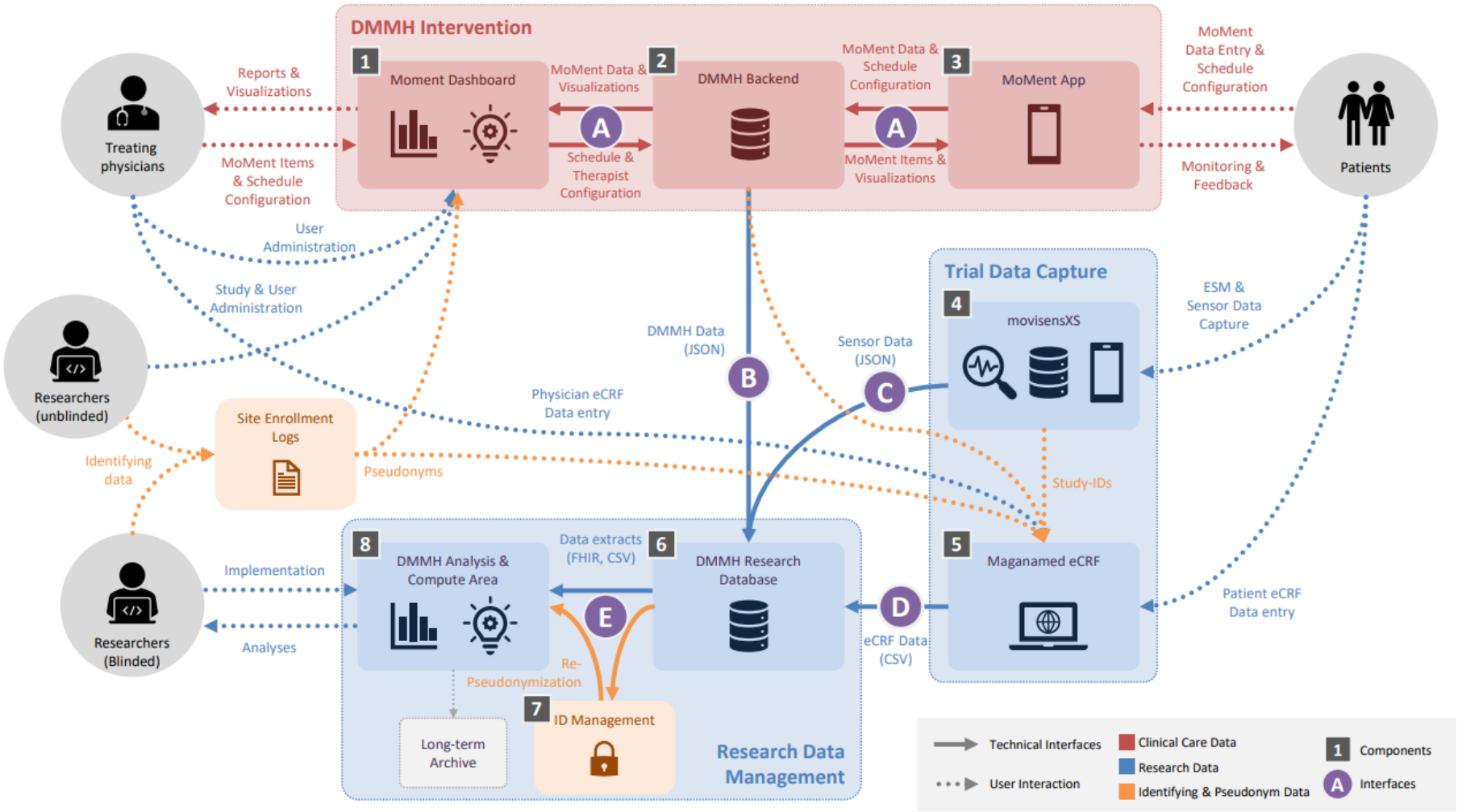
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Workshop Content

1. Data Management Architecture
2. Data Availability
3. Dummy Data
4. Data Use Process
5. Research Data Management: Next Steps

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

Data Management Architecture (1)

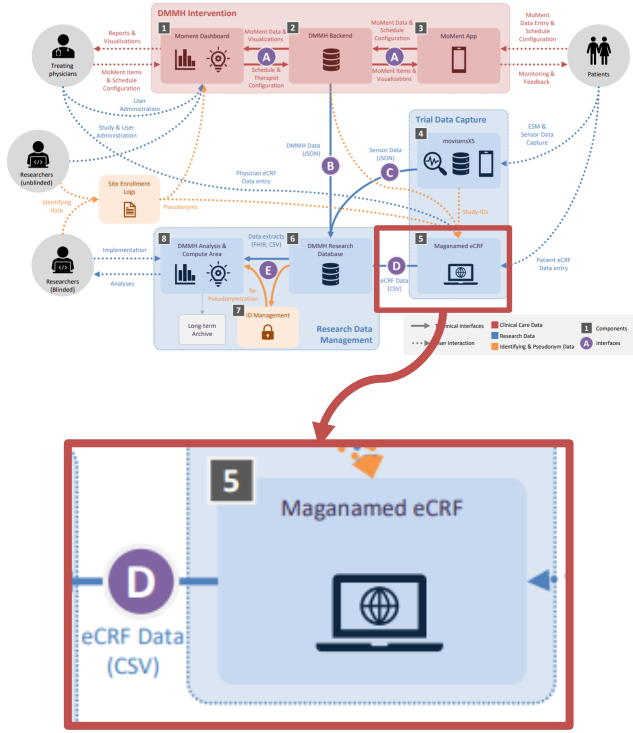


This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Data Management Architecture (2)

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)

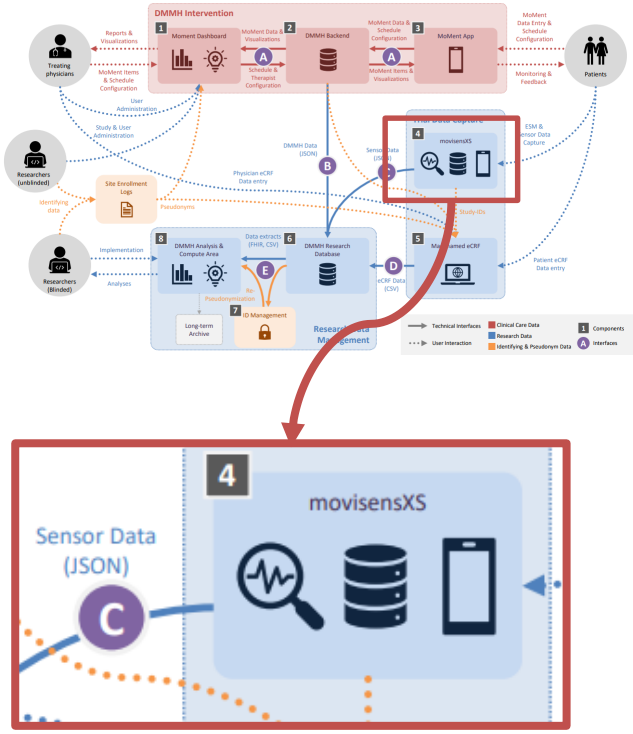


This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Data Management Architecture (3)

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)

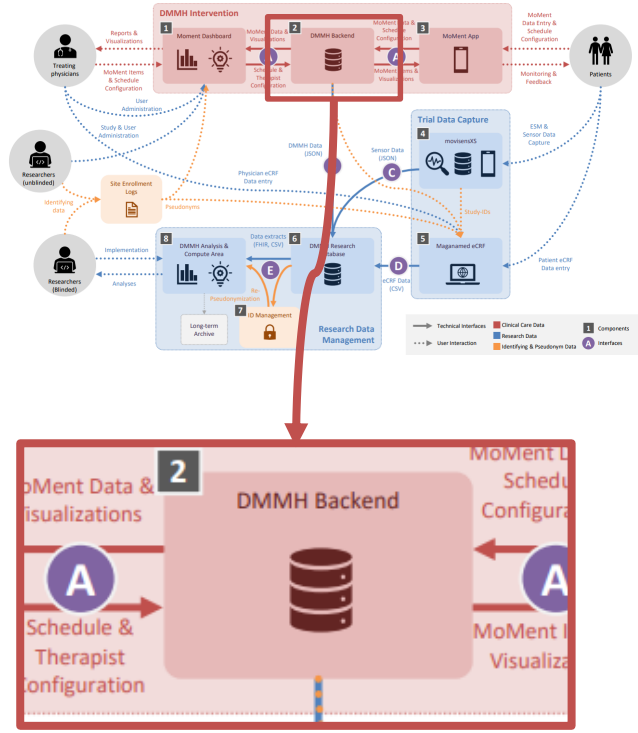


This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Data Management Architecture (4)

TherapyDesigner/MoMent

- Content:
  - primary ESM forms
  - prompt & usage data
- Status: interface still to be implemented
  - manual exports available on request by movisens
- formats:
  - source: JSON (metadata file + 1 file per site)
  - available: CSV
  - dummy data: not yet implemented

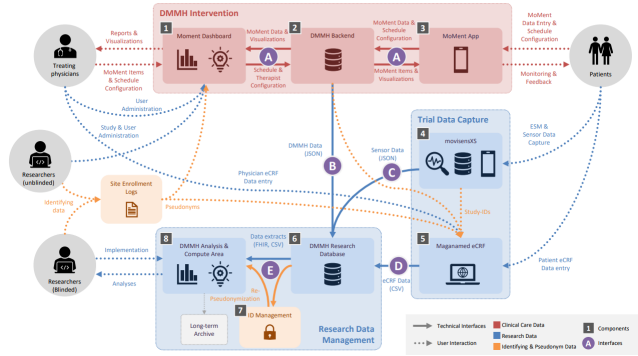


This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Data Management Architecture (5)

Additional Data Sources

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

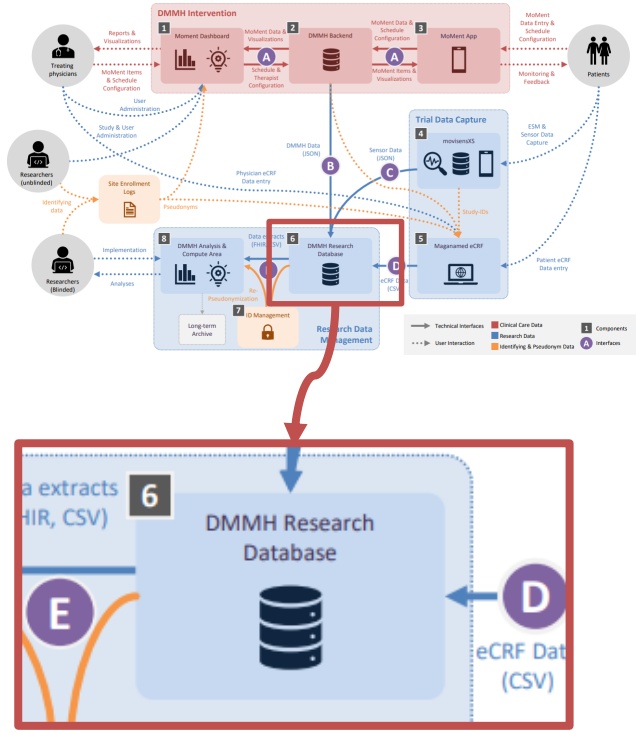


This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Data Management Architecture (6)

Research Database

- Content:
  - collection of data from all IMMERSE sources
- Status:
  - file-based exports from source systems
  - import of MaganaMed data into SQL database in production
  - import of TherapyDesigner data in development
- formats:
  - source: JSON (metadata file + 1 file per site)
  - available: CSV
  - dummy data: not yet implemented

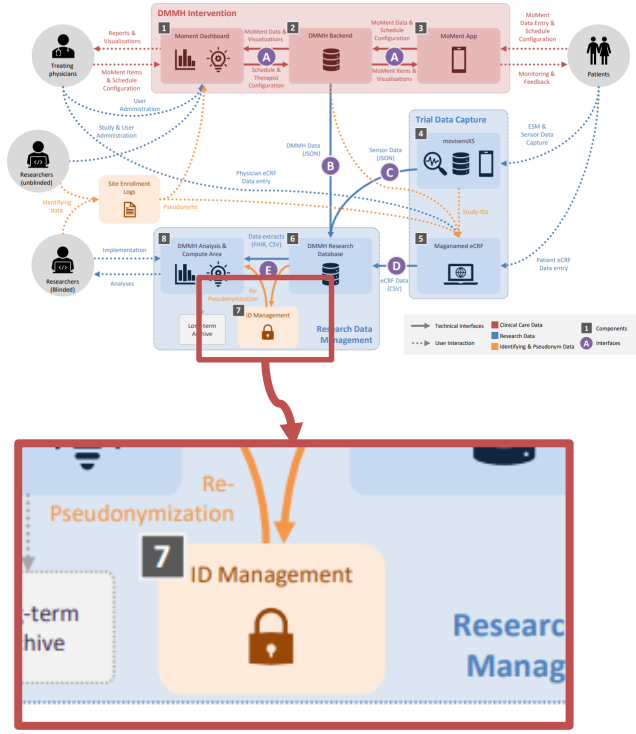


This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Data Management Architecture (7)

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

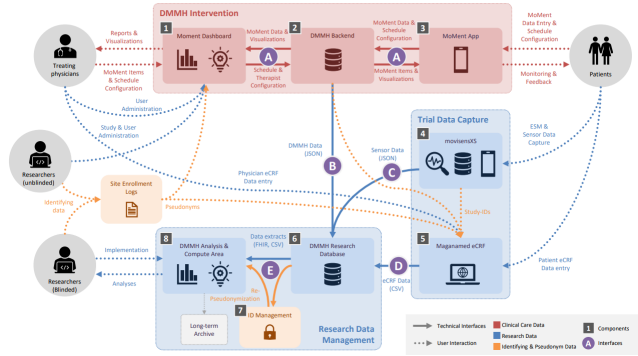


This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Data Management Architecture (8)

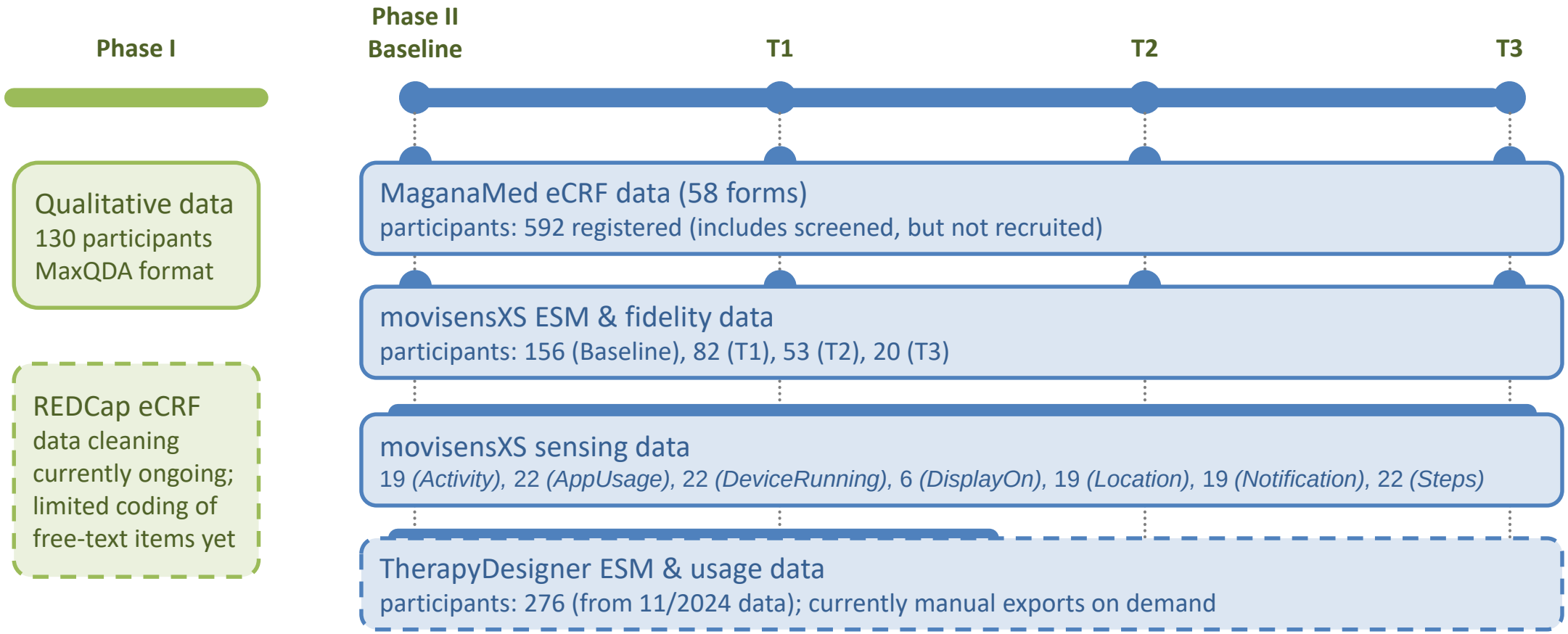
Additional Infrastructure

- 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
  - “fake“ datasets to be used for analysis script development
  - based on trial metadata (e.g. MaganaMed data dictionary)



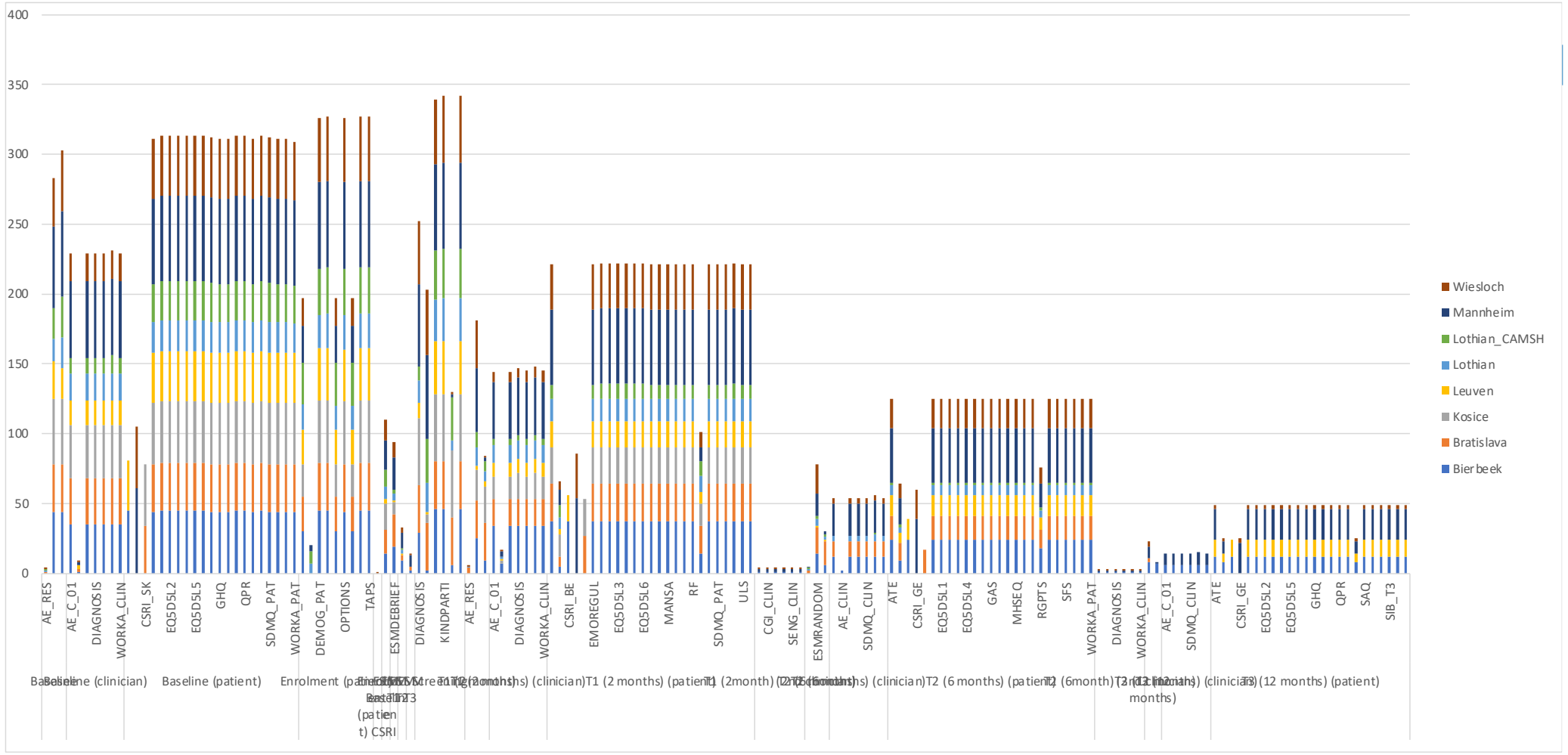
This project has received funding from the European Union’s Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Data Availability (1)



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

Data Availability (2): MaganaMed Visite/Form vs Site



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

Data Availability (3): Structure of MaganaMed Exports

	A	B	C	D	E	F	G	H	I	J	K
1	participant_i	center_name	created_at	started_at	finished_at	visit_name	diary_date	MANSA_01	MANSA_02	MANSA_03	MANSA_04
2	I-MA-C-653	Mannheim	2023-07-23T1	2023-07-23T1	2023-07-23T1	Baseline (patient)			7	3	2
3	I-MA-C-653	Mannheim	2023-10-12T1	2023-10-12T1	2023-10-12T1	T1 (2 months) (patient)		5	3	7	1
4	I-LO-F-349	Lothian	2023-07-20T0	2023-07-20T0	2023-07-20T1	Baseline (patient)		2	3	3	
5	I-LO-F-349	Lothian	2023-03-07T1	2023-03-07T1	2023-03-07T1	T2 (6 months) (patient)			4		2
6	I-LE-C-506	Leuven	2023-06-05T0	2023-06-05T0	2023-06-05T1	Baseline (patient)			5	3	1
7	I-LE-C-506	Leuven	2023-06-03T1	2023-06-03T1	2023-06-03T1	T1 (2 months) (patient)		6	5	3	
8	I-LE-C-506	Leuven	2023-09-26T0	2023-09-26T0	2023-09-26T1	T2 (6 months) (patient)		7		1	1
9	I-LE-C-506	Leuven	2023-12-07T1	2023-12-07T1	2023-12-07T1	T3 (12 months) (patient)				3	1
10	I-LE-F-137	Leuven	2023-08-09T1	2023-08-09T1	2023-08-09T1	Baseline (patient)		5	3	4	1
11	I-LE-F-137	Leuven	2023-04-04T0	2023-04-04T0	2023-04-04T1	T2 (6 months) (patient)		3	4	7	
12	I-LE-P-512	Leuven	2023-03-19T1	2023-03-19T1	2023-03-19T1	Baseline (patient)		1	5		2
13	I-LE-P-512	Leuven	2023-12-04T0	2023-12-04T0	2023-12-04T1	T2 (6 months) (patient)		4	3	5	1
14	I-LE-P-512	Leuven	2023-03-06T1	2023-03-06T1	2023-03-06T1	T3 (12 months) (patient)		2		1	
15	I-WI-T-338	Wiesloch	2023-08-22T1	2023-08-22T1	2023-08-22T1	Baseline (patient)		3		6	2
16	I-WI-T-338	Wiesloch	2023-07-11T1	2023-07-11T1	2023-07-11T1	T1 (2 months) (patient)		7	6		1

participant ID & site

time informations (creation / completion)

visit name

Variable names for each question & their answers

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

Data Availability (4): Structure of movisensXS ESM/Fidelity Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Participant	Trigger	rigger_dat	rigger_tim	gger_coun	Form	m_start_d	m_start_ti	n_finish_d	n_finish_t	n_upload_	n_upload_	Missing	item_815	EMA_happ	EMA_bore	MA_insecu	EMA_angr	EMA_sad	MA_irritat
2	1	Initial	2024-07-0	07:30	1	Initial	2024-07-0	07:34:48	2024-07-0	07:36:05	2024-07-0	07:37:12								
3	1	Button Pre	2024-07-0	07:35	2	EMA	2024-07-0	07:37:40	2024-07-0	07:38:37	2024-07-0	07:42:37			1	2	6	2	5	3
4	1	Random T	2024-07-0	08:00	3	Missing	2024-07-0	08:03:31	2024-07-0	08:03:31	2024-08-0	08:03:31	Ignored							
5	1	Random T	2024-07-0	09:30	4	EMA	2024-07-0	09:33:12	2024-07-0	09:36:15	2024-07-0	09:39:03			1	2	3	4	4	2
6	1	Random T	2024-07-0	11:00	5	EMA	2024-07-0	11:04:51	2024-07-0	11:08:04	2024-07-0	11:09:25			6	6	3	5	3	1
7	1	Random T	2024-07-0	12:30	6	EMA	2024-07-0	12:32:25	2024-07-0	12:35:00	2024-07-0	12:39:57			2	1	2	6	3	5
8	1	Random T	2024-07-0	14:00	7	EMA	2024-07-0	14:01:31	2024-07-0	14:04:21	2024-07-0	14:06:04			6	3	5	3	3	5
9	1	Random T	2024-07-0	15:30	8	EMA	2024-07-0	15:34:04	2024-07-0	15:38:30	2024-07-0	15:39:58			6	7	1	4	7	4
10	1	Random T	2024-07-0	17:00	9	EMA	2024-07-0	17:04:30	2024-07-0	17:06:04	2024-07-0	17:08:10			1	3	4	7	3	6
11	1	Random T	2024-07-0	18:30	10	EMA	2024-07-0	18:34:45	2024-07-0	18:38:25	2024-07-0	18:42:21			3	6	5	4	3	5
12	1	Random T	2024-07-1	08:00	11	EMA	2024-07-1	08:03:49	2024-07-1	08:06:53	2024-07-1	08:07:26			1	1	3	2	1	4
13	1	Random T	2024-07-1	09:30	12	EMA	2024-07-1	09:34:08	2024-07-1	09:36:58	2024-07-1	09:38:31			2	6	1	3	6	5
14	1	Random T	2024-07-1	11:00	13	EMA	2024-07-1	11:04:04	2024-07-1	11:05:39	2024-07-1	11:06:44			2	1	4	6	3	7
15	1	Random T	2024-07-1	12:30	14	EMA	2024-07-1	12:34:33	2024-07-1	12:37:20	2024-07-1	12:40:20			4	3	2	7	6	1
16	1	Random T	2024-07-1	14:00	15	EMA	2024-07-1	14:04:48	2024-07-1	14:06:02	2024-07-1	14:10:00			6	7	4	2	2	7
17	1	Random T	2024-07-1	15:30	16	EMA	2024-07-1	15:33:21	2024-07-1	15:37:28	2024-07-1	15:41:40			1	3	1	6	5	2
18	1	Random T	2024-07-1	17:00	17	EMA	2024-07-1	17:02:39	2024-07-1	17:05:07	2024-07-1	17:09:09			6	4	7	2	4	4

participant ID

Trigger information (time, count)

form (Initial, EMA, Missing)

time (response)

Missing form (Ignored, Incompleted)

Variable names for each question & their answers

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

Data Availability (5): Structure of movisensXS Geotrack data

rticipant_	timestampStart	time	latitude	longitude	altitude	accuracy	oding_osm	coding_osm_	coding_osm_v	ocoding_ty
RE9G5L	2023-08-10T22:	573920	0,020643	0,027398	183,4	7,162	way	amenity	townhall	house
RE9G5L	2023-08-10T22:	574221	0,020028	0,027691	190,5109	69,802	node	amenity	school	house
RE9G5L	2023-08-10T22:	574440	0,020057	0,02739	190,8212	135,6	node	highway	bus_stop	house
RE9G5L	2023-08-10T22:	574735	0,020353	0,028634	191,4753	57,392	way	building	apartments	house
RE9G5L	2023-08-10T22:	576100	0,020262	0,027047	182,1	75,834	node	highway	bus_stop	house
RE9G5L	2023-08-10T22:	577416	0,020612	0,027521	183,8	135,6	way	amenity	townhall	house
RE9G5L	2023-08-10T22:	577770	0,019964	0,027483	181,9	66,406	node	amenity	school	house
RE9G5L	2023-08-10T22:	577787	0,020126	0,028	181,9	58,993	node	highway	bus_stop	house
RE9G5L	2023-08-10T22:	579547	0,020019	0,027065	181,9	59,644	node	highway	bus_stop	house
RE9G5L	2023-08-10T22:	580528	0,020702	0,027391	183,8	72,9	way	amenity	townhall	house
RE9G5L	2023-08-10T22:	583102	0,020184	0,02675	181,9	61,503	way	building	apartments	house
RE9G5L	2023-08-10T22:	584573	0,020701	0,0274	183,8	77,6	way	amenity	townhall	house

participant id

Relative time from the enrollment

anonymized location data

Meta data about the location information extracted using OpenStreetMap (osm type, osm key, osm value, geotype)

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

Data Availability (6): Structure of Tabular TherapyDesigner Data

lt_period_	lt_period_	pseudonym	eraction_t	igger_nam	ger_time_s	ger_time_	raction_st	id	status	value	id.1	status.1	value.1	id.2	status.2	value.2	id.3
2022-11-0	2023-11-2	78H8EJ	interactio	Initialization			INITIALIZA	Therapy G	ANSWERED		Rating of (	ANSWERE	4	Feeling of	ANSWERE	4	Level of re
2022-11-0	2023-11-2	78H8EJ	interactio	Configura	2023-02-2	2023-02-2	COMPLETI	d46e18b5-	ANSWERE	Europe/Brussels							
2022-11-0	2023-11-2	78H8EJ	interactio	Requeste	2023-02-2	2023-02-2	COMPLETI	870d2cd7-	ANSWERE	1	b27097d4-	ANSWERE	6	088403e9-	ANSWERE	7	521994d1-
2022-11-0	2023-11-2	78H8EJ	interactio	TriggerDai	2023-02-2	2023-02-2	COMPLETI	Activity cc	ANSWERE	9	Irritation l	ANSWERE	1	Relaxatio	ANSWERE	6	Enthusias
2022-11-0	2023-11-2	78H8EJ	interactio	Requeste	2023-02-2	2023-02-2	COMPLETI	Want to cl	ANSWERE	0							
2022-11-0	2023-11-2	78H8EJ	interactio	Requeste	2023-02-2	2023-02-2	COMPLETI	Want to cl	ANSWERE	0							
2022-11-0	2023-11-2	78H8EJ	interactio	Requeste	2023-02-2	2023-02-2	COMPLETI	Want to cl	ANSWERE	0							
2022-11-0	2023-11-2	78H8EJ	interactio	TherapyManager			MANAGEF	Key Probl	ANSWERE	Relatie m	Craving	ANSWERE	WAHR	Interpers	ANSWERE	WAHR	Therapy G
2022-11-0	2023-11-2	78H8EJ	interactio	TriggerDai	2023-02-2	2023-02-2	COMPLETI	Content L	ANSWERE	6	Happiness	ANSWERE	4	Sadness L	ANSWERE	1	Insecurity
2022-11-0	2023-11-2	78H8EJ	subintera	TriggerDai	2023-02-2	2023-02-2	COMPLETI	Difficulty	ANSWERE	5							
2022-11-0	2023-11-2	78H8EJ	subintera	TriggerDai	2023-02-2	2023-02-2	COMPLETED		ANSWERE	1							
2022-11-0	2023-11-2	78H8EJ	subintera	TriggerDai	2023-02-2	2023-02-2	COMPLETI	Level of c	ANSWERE	4	Level of s	ANSWERE	2				
2022-11-0	2023-11-2	78H8EJ	subintera	TriggerDai	2023-02-2	2023-02-2	COMPLETI	Level of TI	ANSWERE	5							
2022-11-0	2023-11-2	78H8EJ	subintera	TriggerDai	2023-02-2	2023-02-2	COMPLETI	Level of e	ANSWERE	4							
2022-11-0	2023-11-2	78H8EJ	interactio	TriggerDai	2023-02-2	2023-02-2	COMPLETI	Sadness L	ANSWERE	1	Activity cc	ANSWERE	5	Lonelines	ANSWERE	1	Rating of s

participant id

Response information (COMPLETED, INCOMPLETE, IGNORED)

items, status & answers

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

Data Availability (7): Structure of TherapyDesigner Forms Metadata

The core module for daily beeps	interaction	interactio	37215c57-	Physical present persons (5a)		Family	PersonScale
The core module for daily beeps	interaction	interactio	2a5b7ad8-	Online present persons (5b)		Family	PersonScale
The core module for daily beeps	interaction	interactio	bbc22d78-	Prefer being alone (6c)		4	SevenPointLikerScale
The core module for daily beeps	interaction	interactio	7e6f126e-	Connection with people (6d)		4	SevenPointLikerScale
The core module for daily beeps	interaction	interactio	2aab1289-	Rating of first key problem area		4	SevenPointLikerScale
The core module for daily beeps	interaction	interactio	c3d1019f-	Rating of second key problem area		4	SevenPointLikerScale
The core module for daily beeps	interaction	interactio	699e8d42-	Rating of third key problem area		4	SevenPointLikerScale
The core module for daily beeps	interaction	interactio	e2f60219-	TriggerDailyAtRandomInTime	Daily 8 times with minimum distance of 20 minutes a		
The core module for daily beeps	interaction	interactio	16b3718d-	WeekdayConditionConfig	On one of the following weekdays: MONDAY, TUESD		
The core module for daily beeps	interaction	interactio	66c10f43-	TriggerDailyAtRandomInTime	Daily 8 times with minimum distance of 20 minutes a		
The core module for daily beeps	interaction	interactio	dbd69cdd	WeekdayConditionConfig	On one of the following weekdays: SATURDAY		
The core module for daily beeps	interaction	interactio	7c65264c-	TriggerDailyAtRandomInTime	Daily 8 times with minimum distance of 20 minutes a		
The core module for daily beeps	interaction	interactio	c1c56f62-	WeekdayConditionConfig	On one of the following weekdays: SUNDAY		
Additional Mood Items	subinteraction	configurat	2ecb0087-	bored.		false	
Additional Mood Items	subinteraction	configurat	63ea9c44-	tired.		false	
Additional Mood Items	subinteraction	configurat	3de0d58f-	lonely.		false	
Additional Mood Items	subinteraction	configurat	082bf7d4-	exhilerated.		false	
Additional Mood Items	subinteraction	configurat	2449d48d-	stressed.		false	
Additional Mood Items	subinteraction	configurat	fc45f56d-	angry.		false	
_AddOn Modules:Transdiagnostic:Behaviou	subinteraction	configurat	765eed7a-	Threat Anticipation	Please thi	false	
_AddOn Modules:Transdiagnostic:Behaviou	subinteraction	interactio	fddc4ffa-	Rating of Threat in the near future		4	SevenPointLikerScale

Interaction & configuration informations

Interaction ids & question name

Default value of answer

Question type

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

Dummy Data (1)

Intention

- provide early access to similarly structured datasets to enable analysis script implementation
  - circumvents data protection & trial pre-clouseout issues

Method

- based on metadata of actual datasets (e.g. MaganaMed data dictionary)
- data items are filled with randomly chosen values from the respective set or missings

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

Dummy Data (2)

What it can & can‘t do

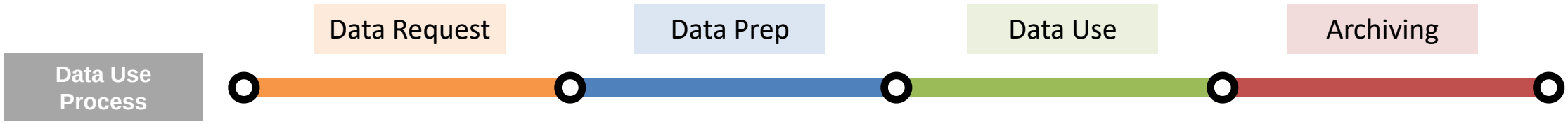
- data is identically structured to the real data & syntactically correct
- however, the data is not at all medically plausible
  - it does not reflect frequencies or correlations from the original data
- data can be used to develop & test analysis scripts regarding correct implementation
  - analysis results of dummy data will of course not reflect content of the actual data

Availability

- currently available for ManaMed eCRF and movisensXS ESM data

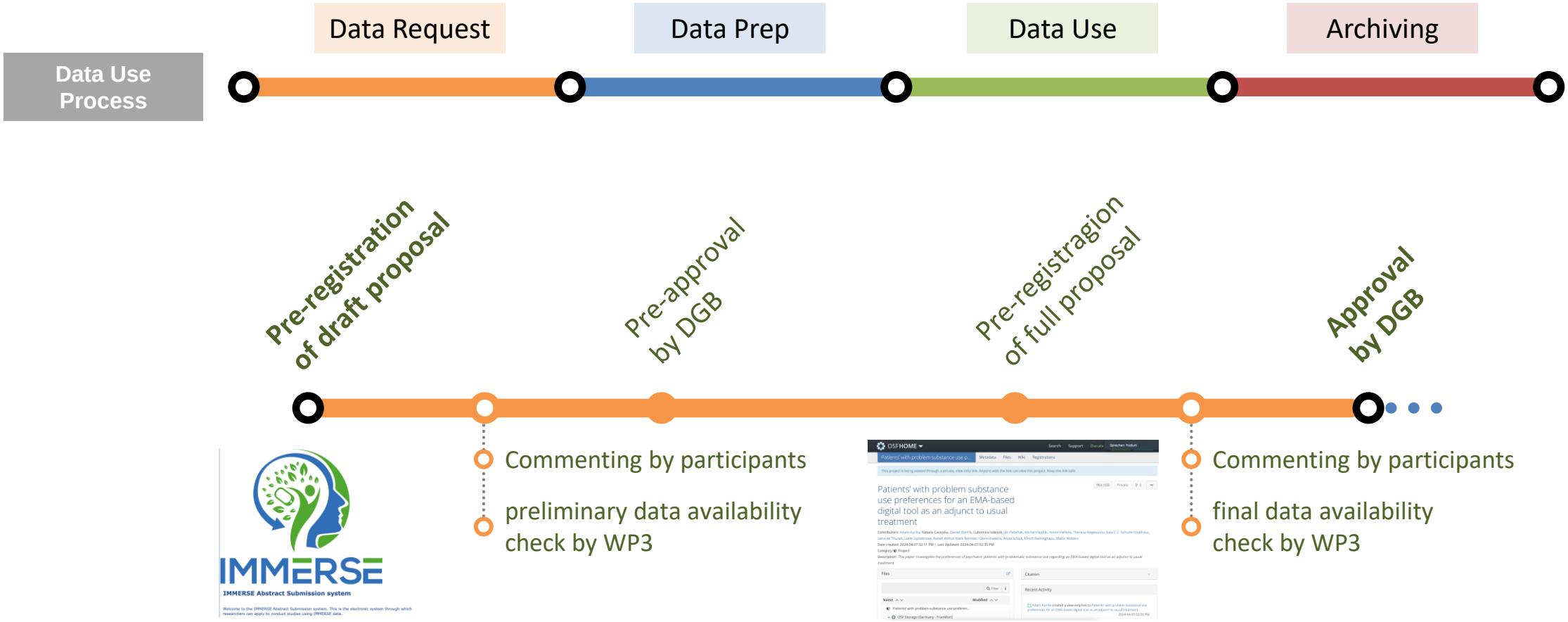
This project has received funding from the European Union’s Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Data Use Process (1)



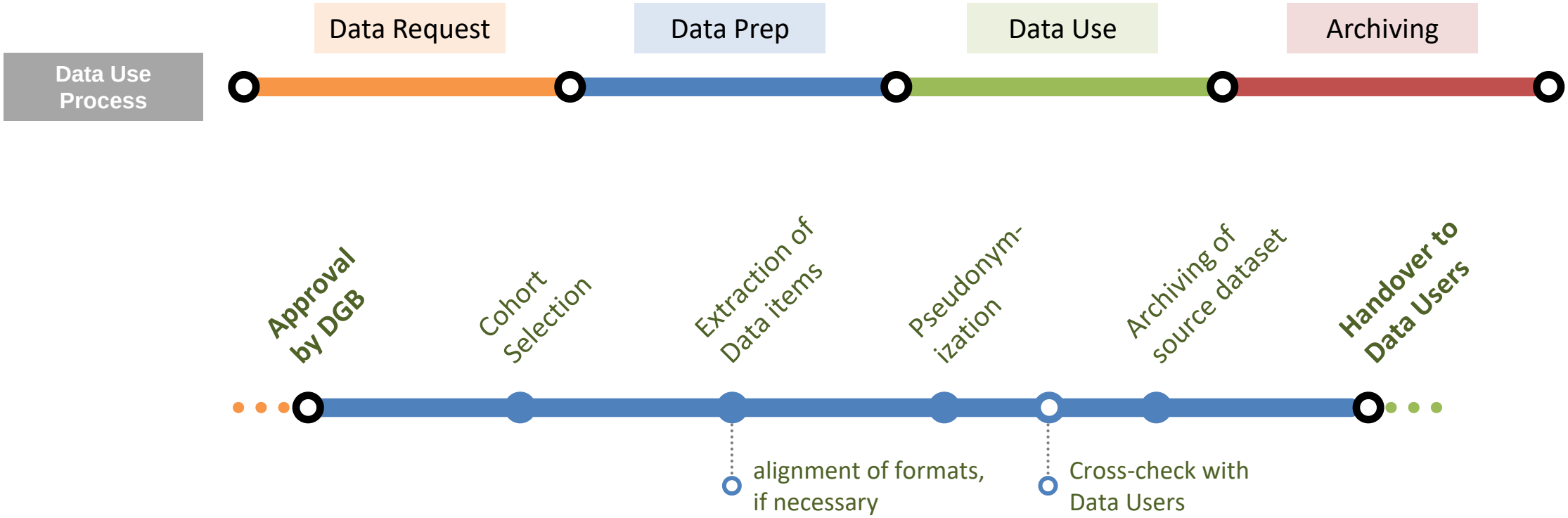
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Data Use Process (2)



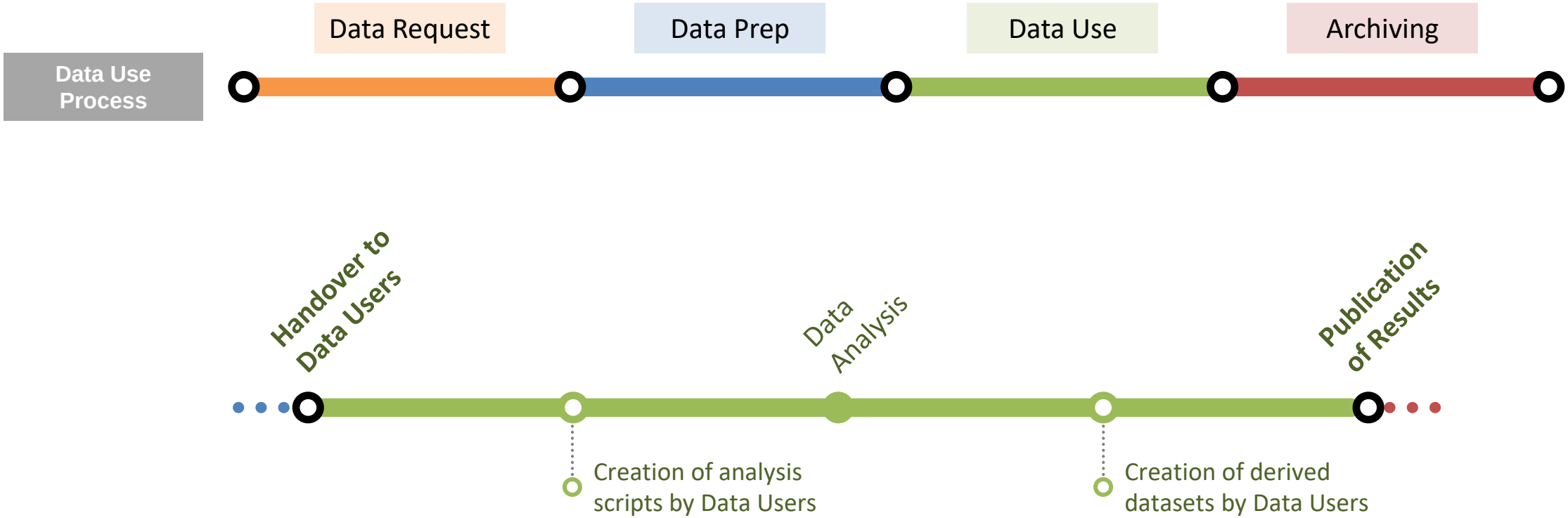
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Data Use Process (3)



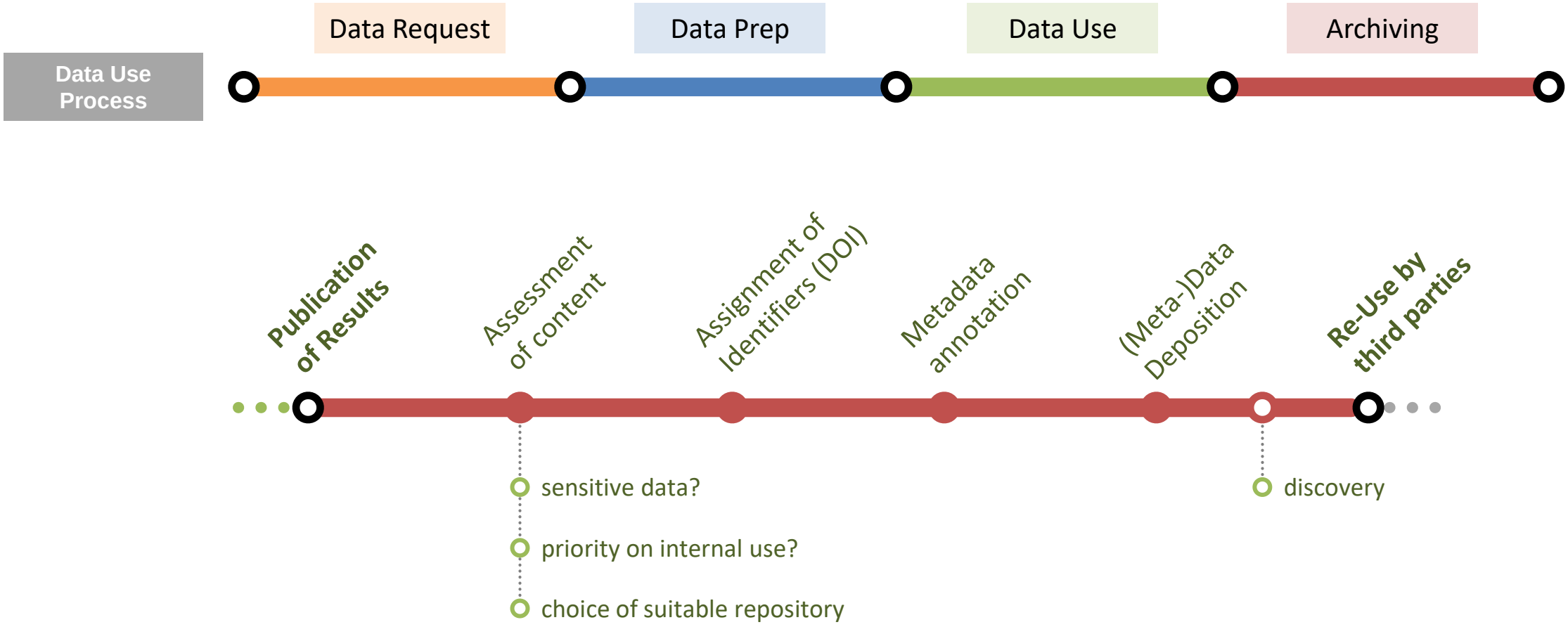
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

Data Use Process (4)



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

Data Use Process (5)



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

Research Data Management: Next Steps

Priority from our side

- finalization of import of all source data into SQLite database
- implementation of interface to TherapyDesigner
- finalization of metadata annotation & FHIR conversion

Requirements from the consortium?

- e.g. extension of dummy data?
- e.g. data quality aspects?

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