

---

# Semantische Datenintegration heterogener medizinischer Daten

– ein Beispiel der Interaktion zwischen



**Prof. Dr. Juliane Fluck**

---



# Integrative Data Semantics for Neurodegeneration Research

 **Fraunhofer**

SCAI



**Coordinator:**

Prof. Dr. Juliane Fluck



Prof. Dr. Stefan Bonn  
Director for Medical Systems  
Biology  
University Clinic Hamburg-  
Eppendorf



Prof. Dr. Thomas Klockgether  
Director Neurology UKB  
Director of Clinical Research  
DZNE



Research Data Management and Infrastructure  
FAIR Data

Information Centre Life Sciences

ZB MED

Funded by



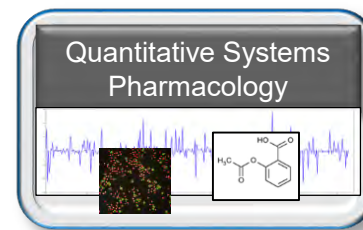
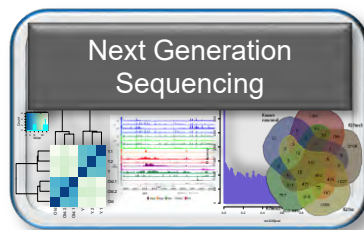
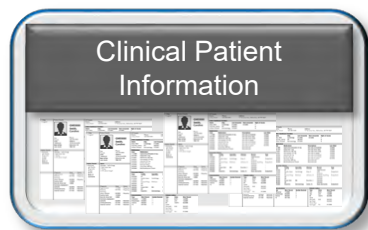
Federal Ministry  
of Education  
and Research

# Data produced within DZNE

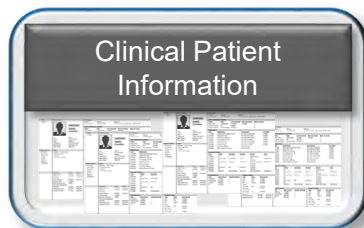
---

# Data produced within DZNE + Routine Data + Public Data

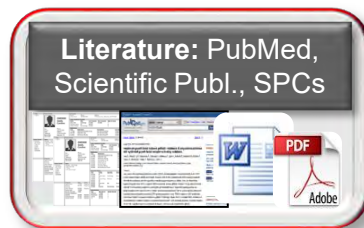
## DZNE Primary Data



## + UKB Routine Data



## + Public Data

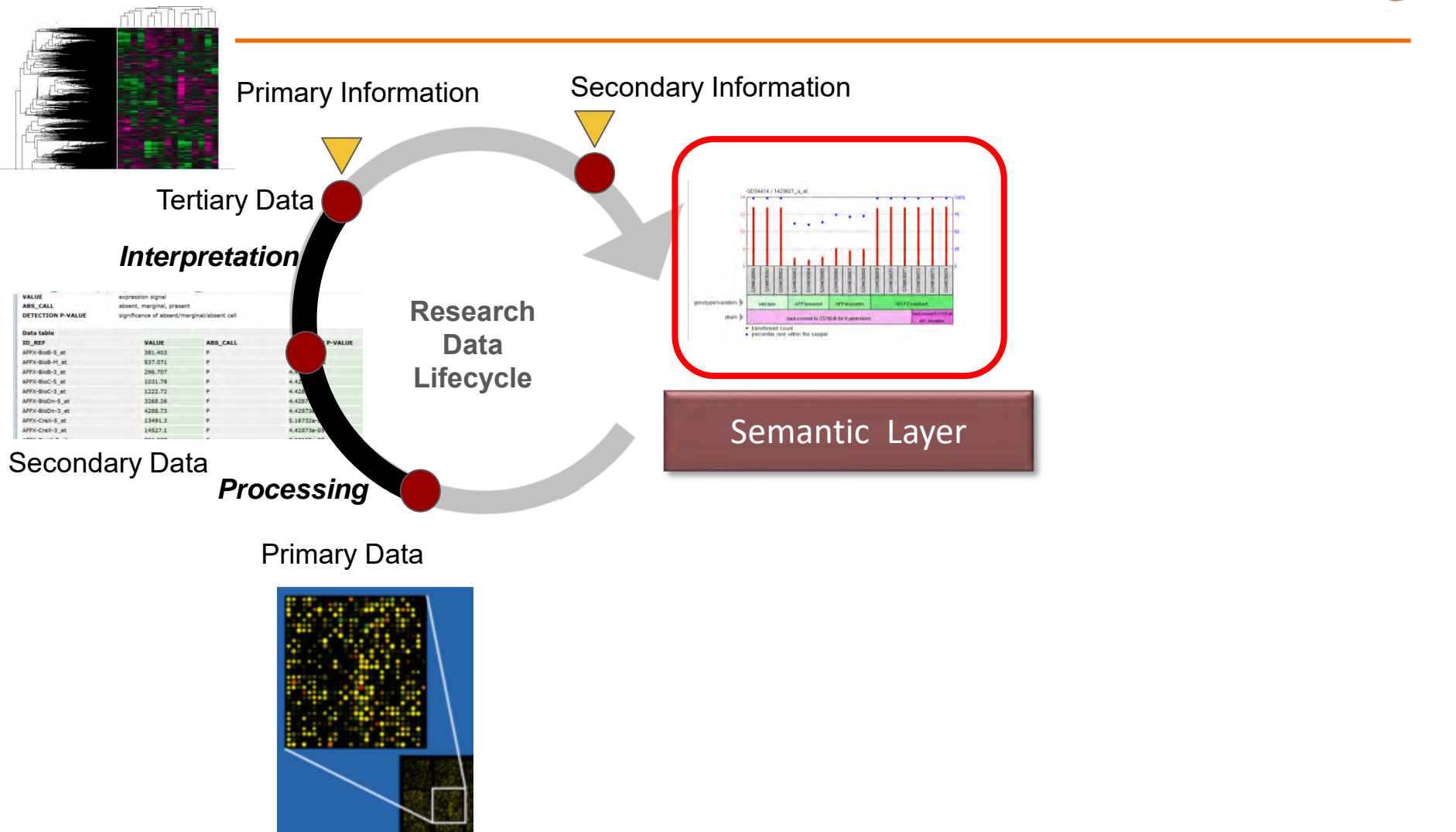


# Key Objectives

---

- Enable combined analysis
- Easily extendible to new data sets

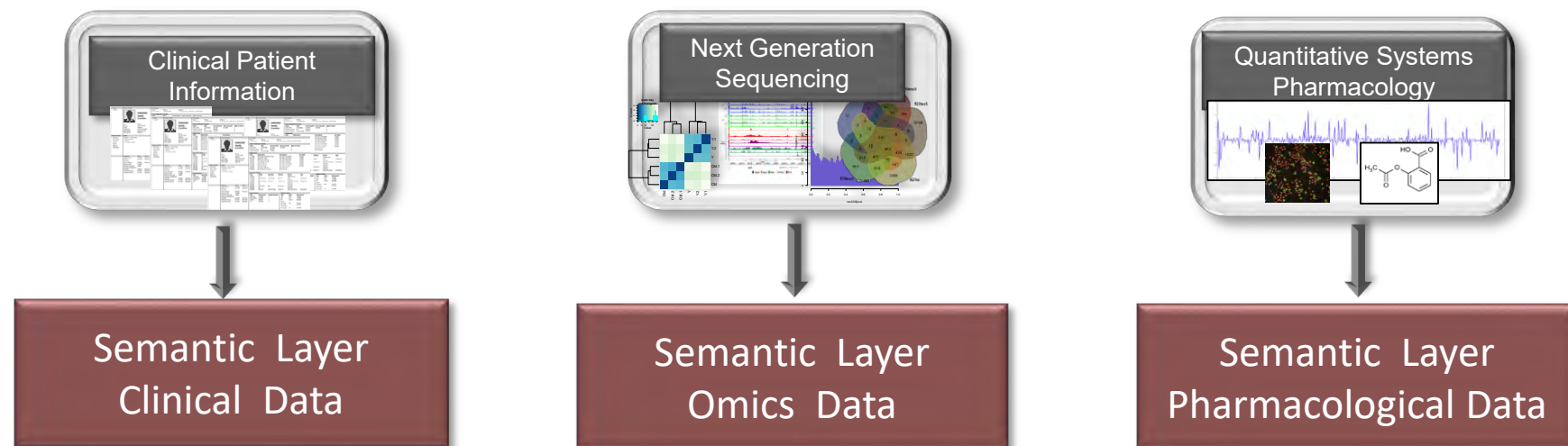
# Research Data Lifecycle



Adapted from 'Recent ORKG developments' by Markus Stocker DILS2018

# Semantic Layers

- Relevant data types and their annotations (= metadata)
- Shared Semantics → Data types/annotations using common, community-accepted vocabulary
- Analyzed/interpreted Data (often no raw data)
- Data integration and links between data elements and with external data



# Semantic Layer: NGS Data

---

## Relevant data types:

- small RNA
- RNA
- CAGE (cap analysis gene expression) for promoter activity detection
- WES (whole exome sequencing) for DNA variant detection

## Community accepted terms:

- HGNC, EntrezGene, MGI, UniProt terms
- dbSNP
- miRBase



# Semantic Layer: NGS Data

---

## Data Interpretation

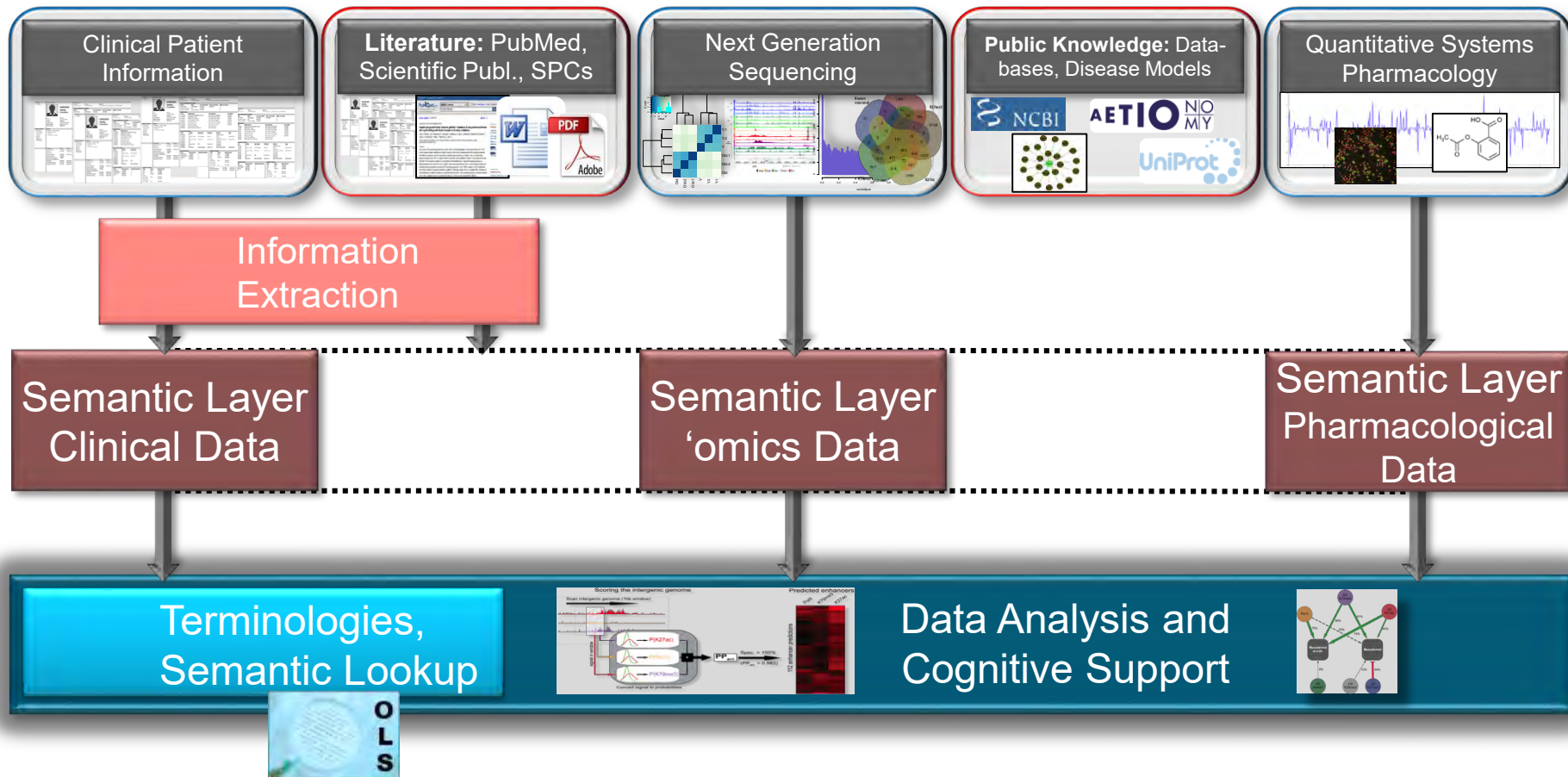
- Differential expression (DE) for all RNA expressions
- CADD (Combined Annotation Dependent Depletion)

## Data integration (links between data elements and external data)

- SNP – Gene relations (dbSNP)
- Small RNA – Gene/Protein relations (Text Mining and mirTarBase)
- Gene – Disease relations (DisGeNet)
- SNP – Disease relations (DisGeNet)
- Link to (disease-specific) network information

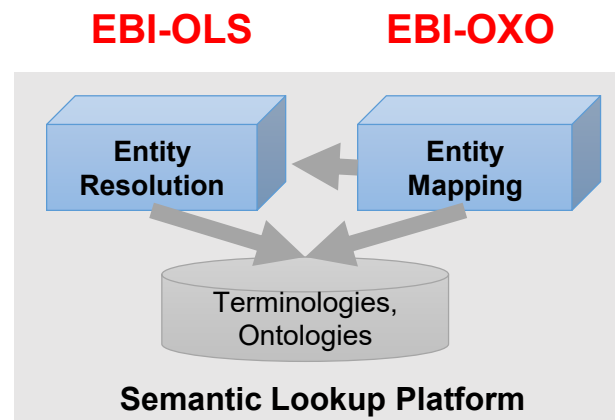
# IDSN Platform

Primary Data + UKB Routine Data + Public data



# Semantic Lookup Platform

- Provide a single access point to heterogeneous terminological resources and common vocabulary
- Suggest for a term a concept from controlled vocabularies and in such a way facilitate annotation of data
- Provide cross-references for concepts
- Query, browse and navigate terminologies/ontologies via UI/API
- Visualize & analyze the hierarchy



OLS: <https://github.com/EBISPOT/OLS>  
 OXO: <https://github.com/EBISPOT/OXO>



**Allow for query expansion and semantic data access**

# Ontology Lookup Service

Welcome to the EMBL-EBI Ontology Lookup Service.

- prefrontal cortex
- prefrontal association cortex
- prefrontal bone
- prefrontal gyrus morphology trait
- Prefrontal hypometabolism in FDG PET

[Looking for a particular ontology?](#)



The Ontology Lookup Service (OLS) repository aims to provide a central point to the latest information on the website of the OLS and to maintain and update the OLS EBI.

### Jump to

- prefrontal cortex **UBERON** **UBERON:0000451**
- Prefrontal cortex **FMA** **FMA:224850**
- Prefrontal Cortex **NCIT** **NCIT:C154778**
- prefrontal cortex **EFO** **UBERON:0000451**
- prefrontal cortex **BTO** **BTO:0002807**

Search OLS for **prefr**

### Tools

The SPOT team also provides tools like Zooma and Oxo. Oxo provides mappings between different ontologies. We are happy to assist in building ontologies in OLS and in building spreadsheets.

### Contact Us

For feedback, enquiries or suggestion about OLS or to request a new ontology please contact [ols-support@ebi.ac.uk](mailto:ols-support@ebi.ac.uk). For bugs or problems with the code or API please report on [GitHub issue](#). For announcements relating to OLS, such as new releases and new features sign up to the [OLS announce mailing list](#).

### Data Content

Updated 15 Mär 2019 04:34

- 226 ontologies
- 5,418,186 terms
- 21,689 properties
- 480,657 individuals

[Tweets by OLS](#)



# Ontology Lookup Service

Home **Ontologies** Documentation About

OLS > Uber-anatomy ontology **UBERON** > **UBERON:0000451**

## prefrontal cortex

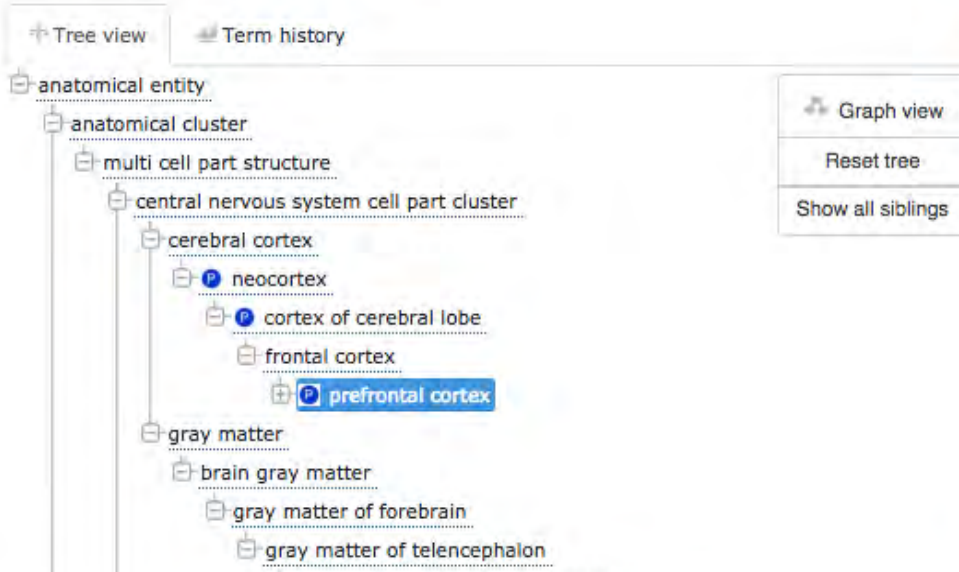
Search UBERON

[http://purl.obolibrary.org/obo/UBERON\\_0000451](http://purl.obolibrary.org/obo/UBERON_0000451)

The anterior part of the frontal lobes of the brain, lying in front of the motor and premotor areas. This brain region has been implicated in planning complex cognitive behaviors, personality expression, decision making and moderating correct social behavior. The basic activity of this brain region is considered to be orchestration of thoughts and actions in accordance with internal goals. The most typical psychological term for functions carried out by the pre-frontal cortex area is executive function. Executive function relates to abilities to differentiate among conflicting thoughts, determine good and bad, better and best, same and different, future consequences of current activities, working toward a defined goal, prediction of outcomes, expectation based on actions, and social 'control' (the ability to suppress urges that, if not suppressed, could lead to socially-unacceptable outcomes). Many authors have indicated an integral link between a person's personality and the functions of the prefrontal cortex. - definition adapted from Wikipedia [ NIFSTD:nlx\_anat\_090801 ]

**Synonyms:** prefrontal association cortex

## Mapping Service EBI-EXO



### Term info

database cross reference

- o FMA:224850
- o <http://braininfo.rprc.washington.edu/centraldirectory.aspx?ID=1072>
- o MA:0000906
- o DHBA:10172
- o GAID:676
- o MESH:A08.186.211.730.885.213.270.700
- o EMAPA:35356
- o [http://en.wikipedia.org/wiki/Prefrontal\\_cortex](http://en.wikipedia.org/wiki/Prefrontal_cortex)
- o BTO:0002807
- o EFO:0001384
- o [http://uri.neuinfo.org/nif/nifstd/nlx\\_anat\\_090801](http://uri.neuinfo.org/nif/nifstd/nlx_anat_090801)
- o BAMS:FrA

Subsets

# Customization of OLS + OXO to Semantic Lookup Platform

## ❖ OLS Ontologies

- ✓ NCBI Taxonomy
- ✓ UBERON
- ✓ HPO
- ✓ GO
- ✓ Cell Ontology
- ✓ ChEBI
- ✓ Disease Ontology
- ✓ Cellular Microscopy  
Phenotype Ontology  
(CMPO)
- ✓ NCIT

## ❖ Added Resources and mappings

- ✓ HGNC, EntrezGene, MGI, UniProt
- ✓ dbSNP
- ✓ miRBase
- ✓ MeSH
- ✓ ICD
- ✓ ATC

(Gelbe Liste)

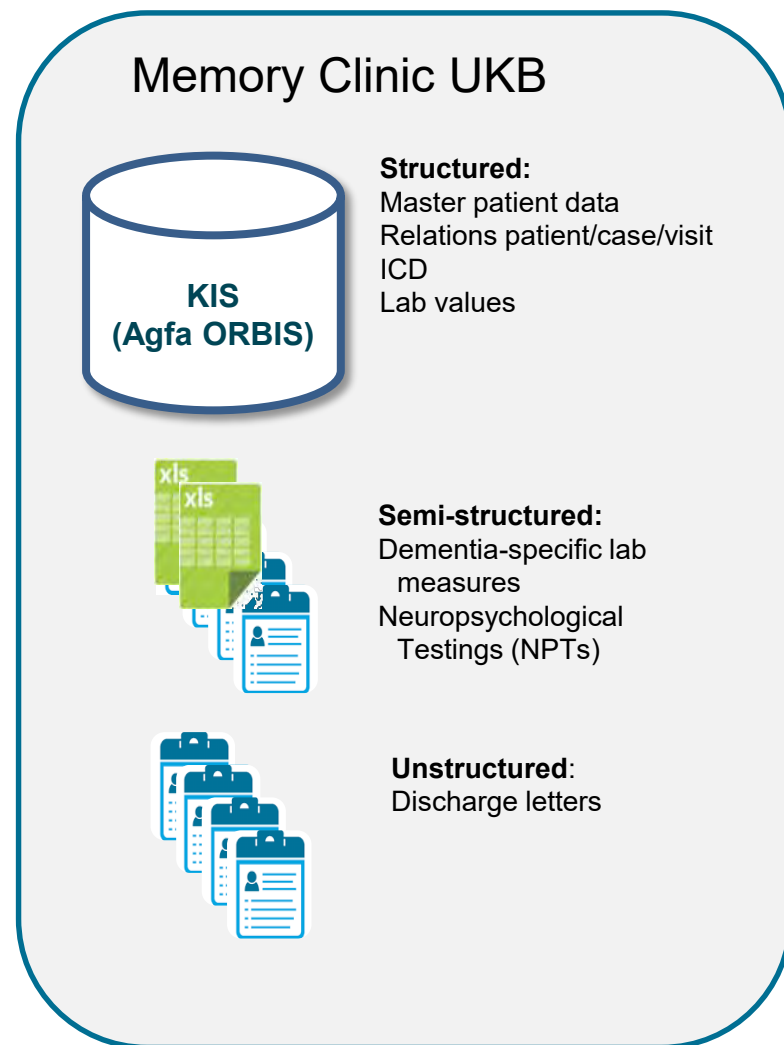
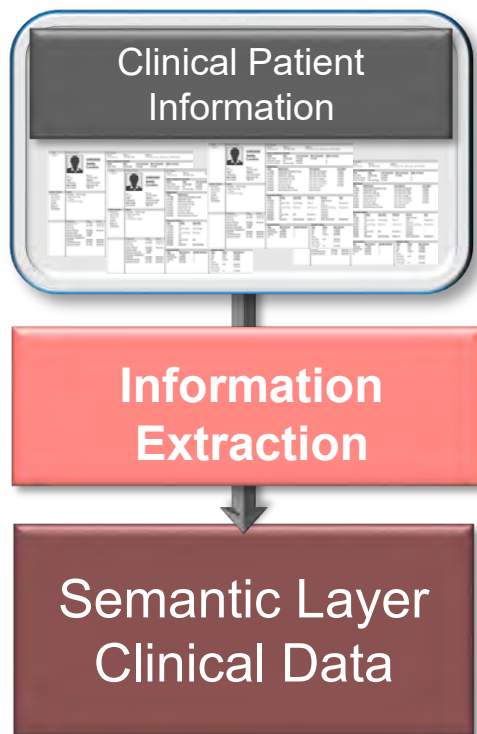
(LOINC)

Easy deployment of customized version

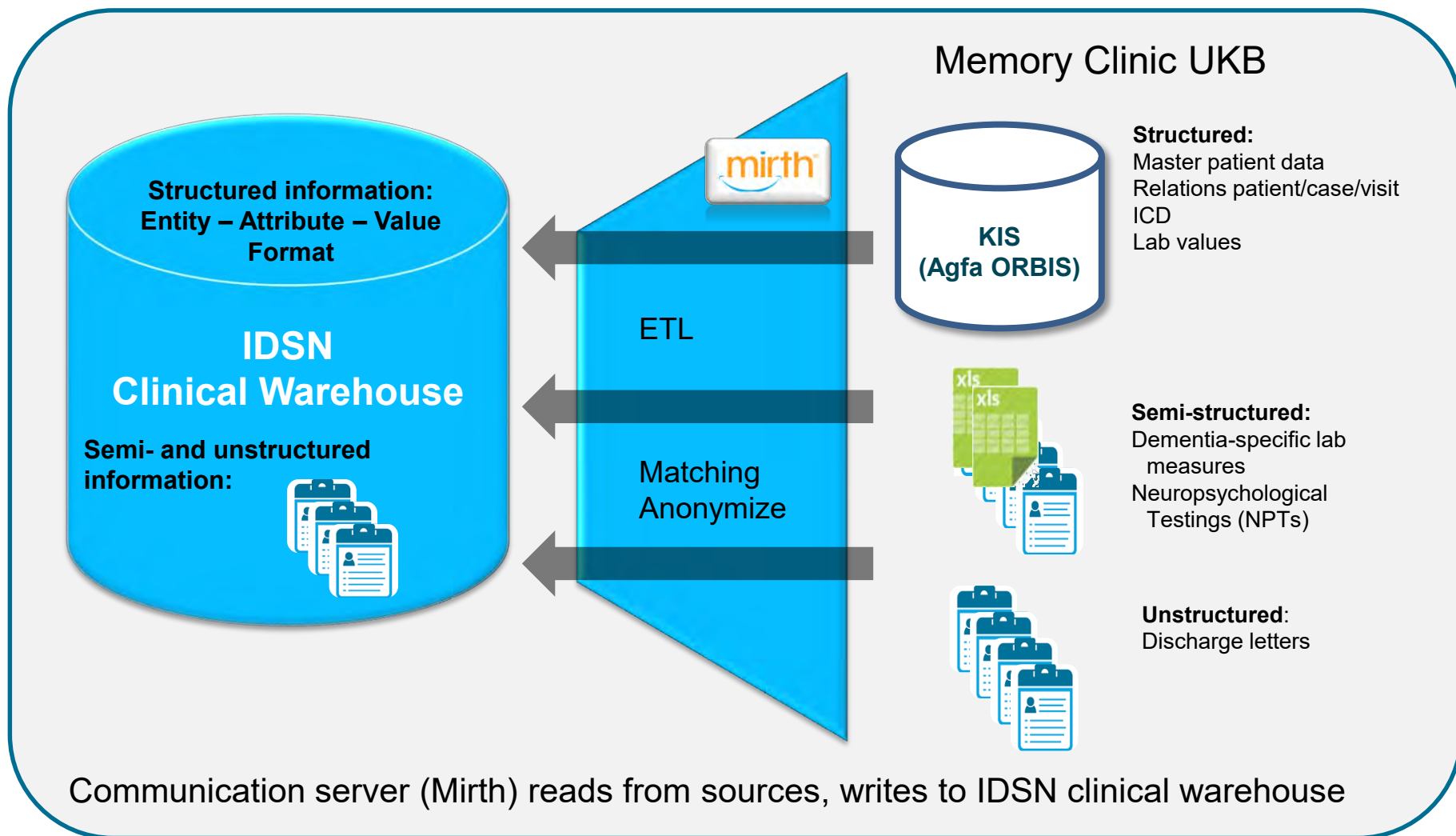
Public version currently established at



# Extraction of clinical routine data

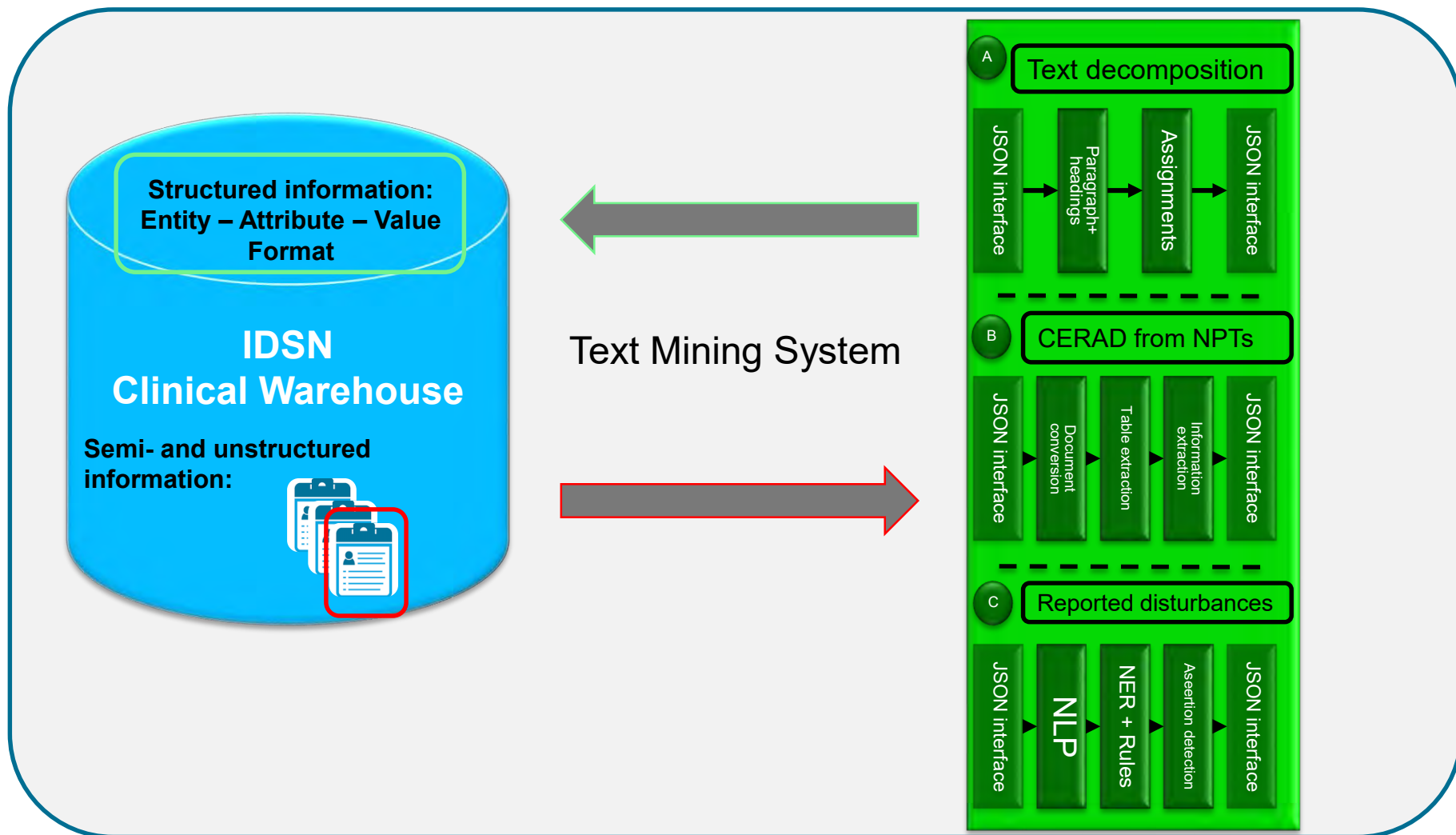


# Extraction of clinical routine data

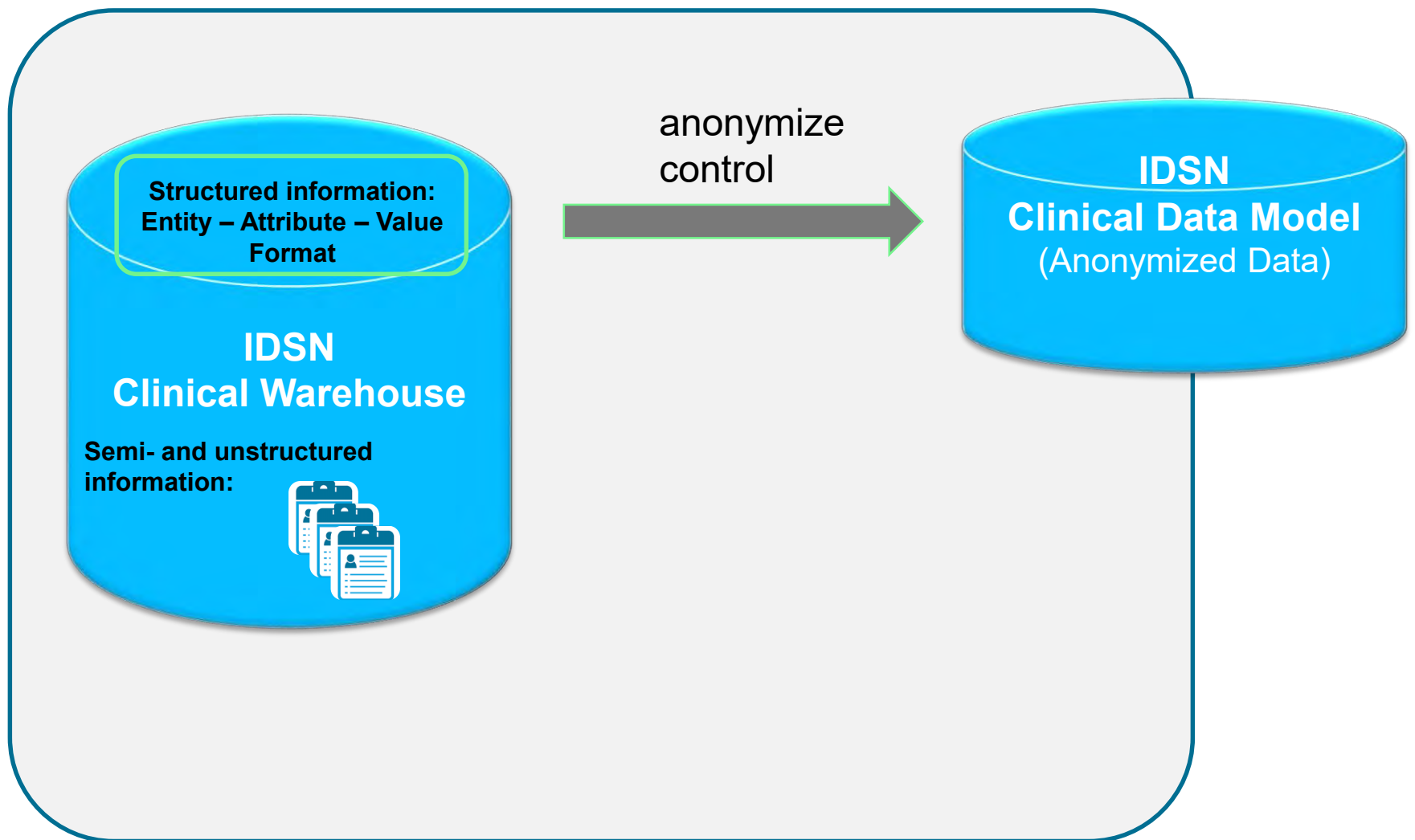




# Extraction of clinical routine data



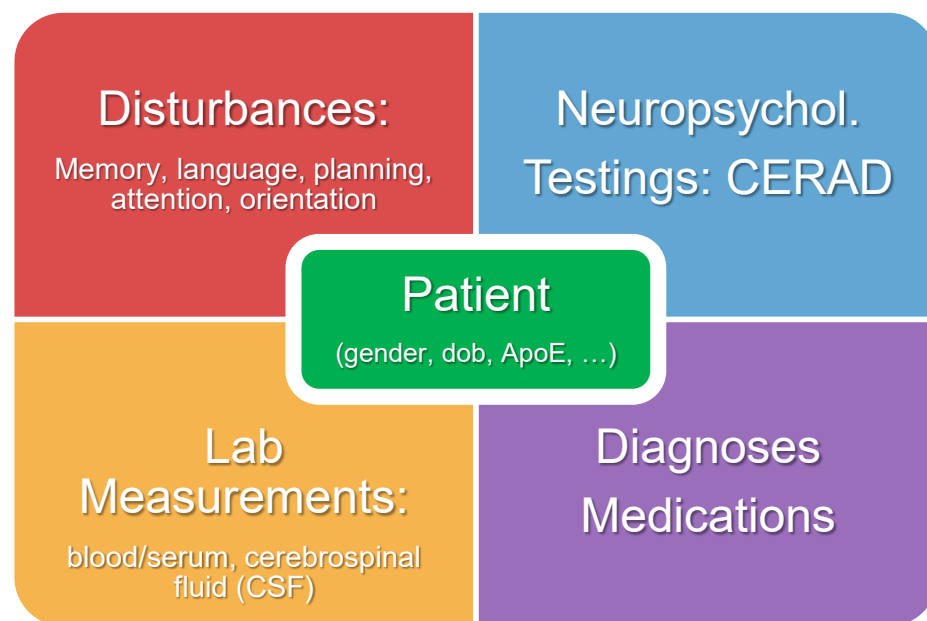
# Transform to clinical data model



# Clinical layer

---

- Common data model for DZNE and UKB semantic layer
- Interoperability: public terminologies(?)
- FHIR: interface, metadata



# Interaction with Medicine Informatics Initiative MII



- From IDSN to SMITH: Extraction pipeline established for IDSN can be used at UKB as starting point for data extraction in SMITH
- From MII to IDSN: Common data model will be compared/adapted to MII common data models
- From IDSN to MII: Additional MI use case for data integration

# Clinical Data Visualisation

## Interactive data exploration:

e.g. „How does age relate to MMSE?“

## Subgroup comparison:

e.g. „How does the ApoE genotype influence the data of first clinical checkup?“

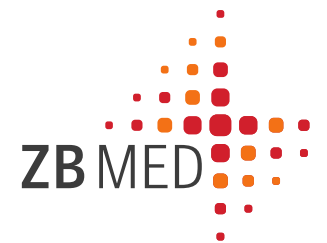
## Longitudinal viewer of single patient data/group data



# Key Aims Translational Phase

---

- Can clinical routine data supplement cohort data?
- Analyse clinical data in interactive analysis for new hypothesis formulations
- Combined analysis of clinical and next generation sequencing data in the disease area frontal lobe degeneration and spinal ataxia
- Data publication as FAIR data



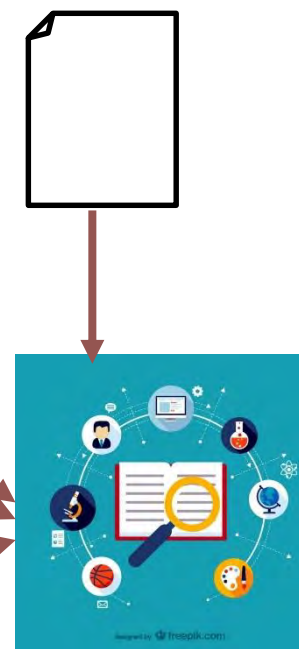
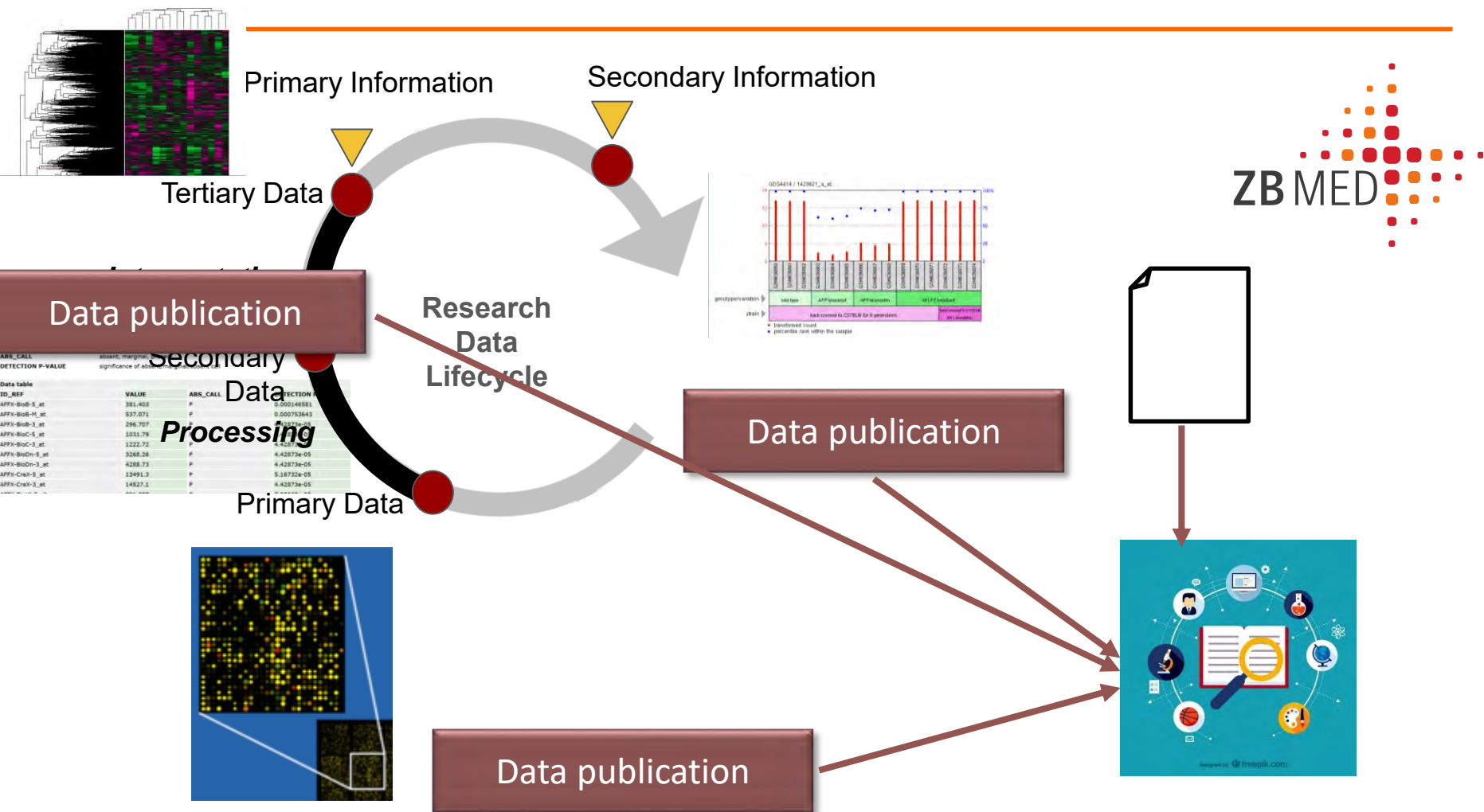
# FAIR DATA Principles

---

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**e-usable

The FAIR Guiding Principles for scientific data management and stewardship.  
Wilkinson et al., [Sci Data](https://doi.org/10.1038/sdata.2016.18). 2016. doi: 10.1038/sdata.2016.18

# Data Publication as FAIR Data





# Take home messages

---

- ❖ Semantic Data Layers
  - are core concepts for condensed and semantically interoperable data allows for easy updating and extensions
  - are building blocks for combined data analysis
  - can/should be published as FAIR Data publications
- ❖ Semantic Lookup Platforms are necessary as single access point for
  - Semantic annotation and integration
  - Query expansion and data analysis
- ❖ Text mining is a building block for making clinical information accessible
- ❖ research data management/data FAIRification is necessary for better research

# Acknowledgement IDSN colleagues

---

Sebastian Schaaf (Fraunhofer SCAI)

Sumit Madan (Fraunhofer SCAI)

Lisa Langnickel (Fraunhofer SCAI)

Dr. Maksims Fiosins (DZNE)

Prof. Dr. Stefan Bonn (UKE, DZNE)

Dr. Eugenio Fava (DZNE)

Dr. David Hecker (DZNE)

Dr. Mischa Uebachs (UKB)

Philipp Koeppen (UKB)

Vyara Tonkava (UKB)

Matthias Jungck (UKB)

Kilian Krockauer (UKB)

Prof. Dr. Thomas Klockgether (UKB, DZNE)

Prof. Dr. Peter Heutink (DZNE)

Dr. Javier Simón-Sánchez (DZNE)

**Funding program:**

**I:SEM Integrative Datensemantik in der Systemmedizin**



Federal Ministry  
of Education  
and Research

Grant No.: 031L0029 [A-C].