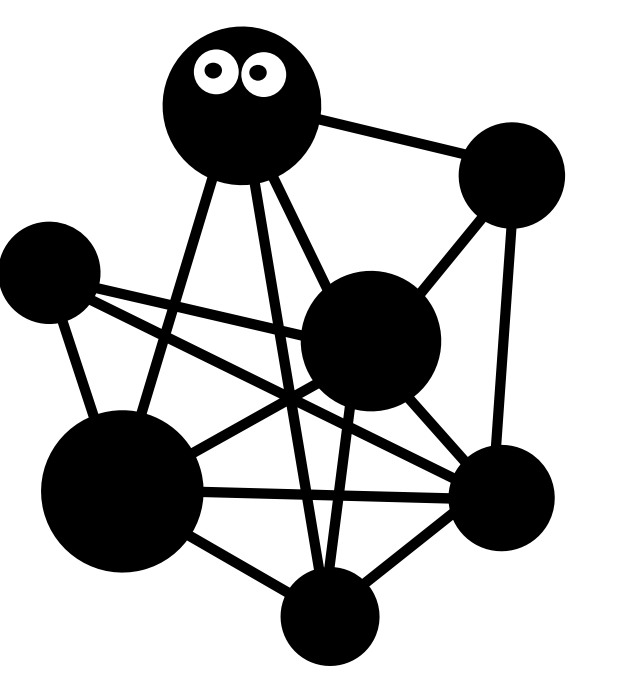




# MeDaX - our vision for bioMedical Data eXploration

Judith A.H. Wodke<sup>1</sup>, Ron Henkel<sup>1</sup>, Dagmar Waltemath<sup>1</sup>  
<sup>1</sup> Medical Informatics Laboratory, University Medicine Greifswald



## Motivation

An immense amount of (bio)medical data is collected in clinical everyday life, providing an enormous potential for research and evidence-based medicine. However, systematic sharing and re-usage of especially clinical data is prevented by several reasons. Most reasons can be categorized as either:

- i) data complexity and heterogeneity (**scientific**),
- ii) lack of appropriate tools for storage and comparison (**technical**), and
- iii) data security and protection of personal information (**juridical**).

## ETL concept

extr(ack) responsibly



**Individual data extraction** from diverse data sources [A-D]  
**DQA for source data:** Adaptation/expansion of the MIRACUM DQA tool [17]

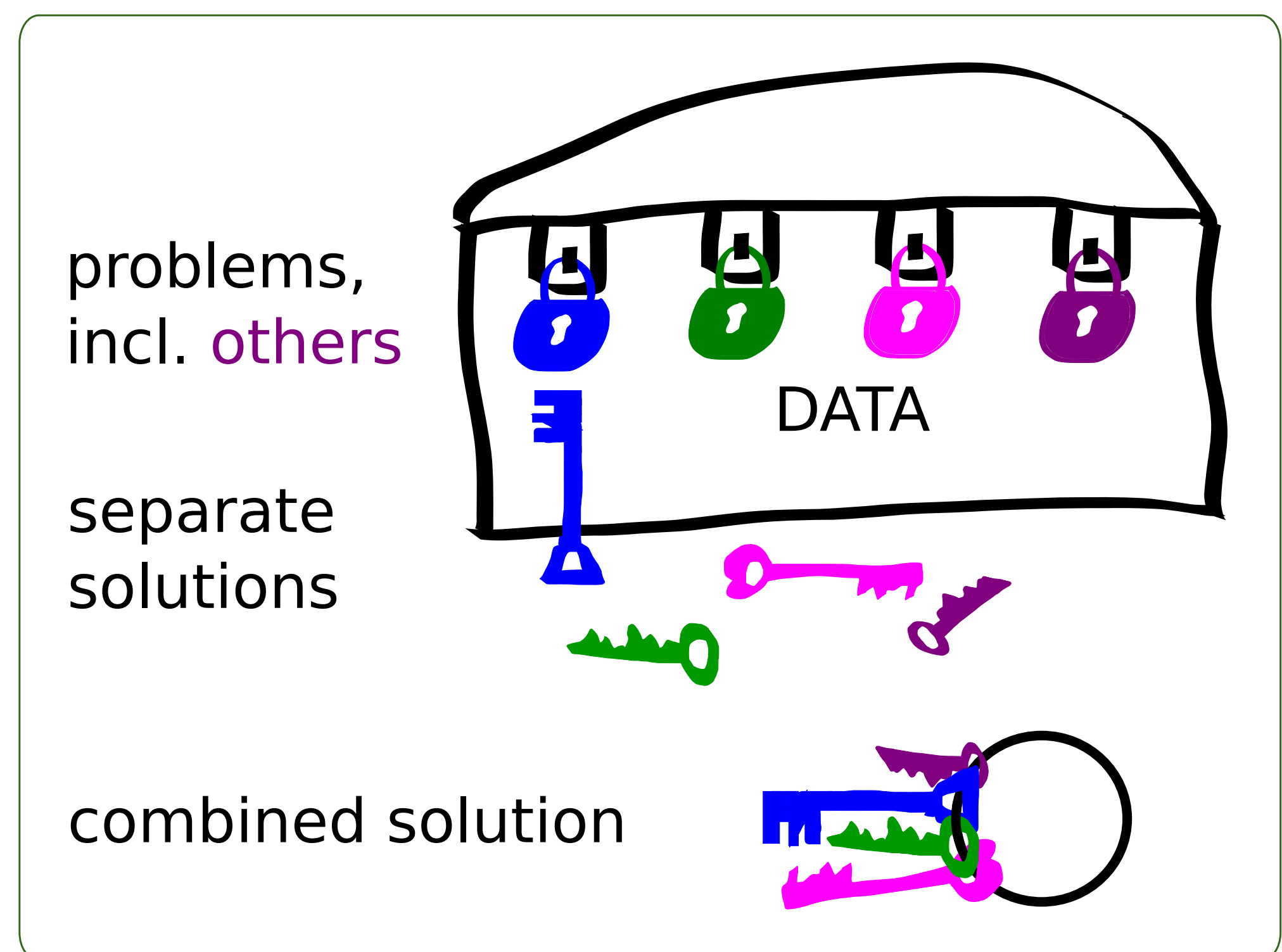
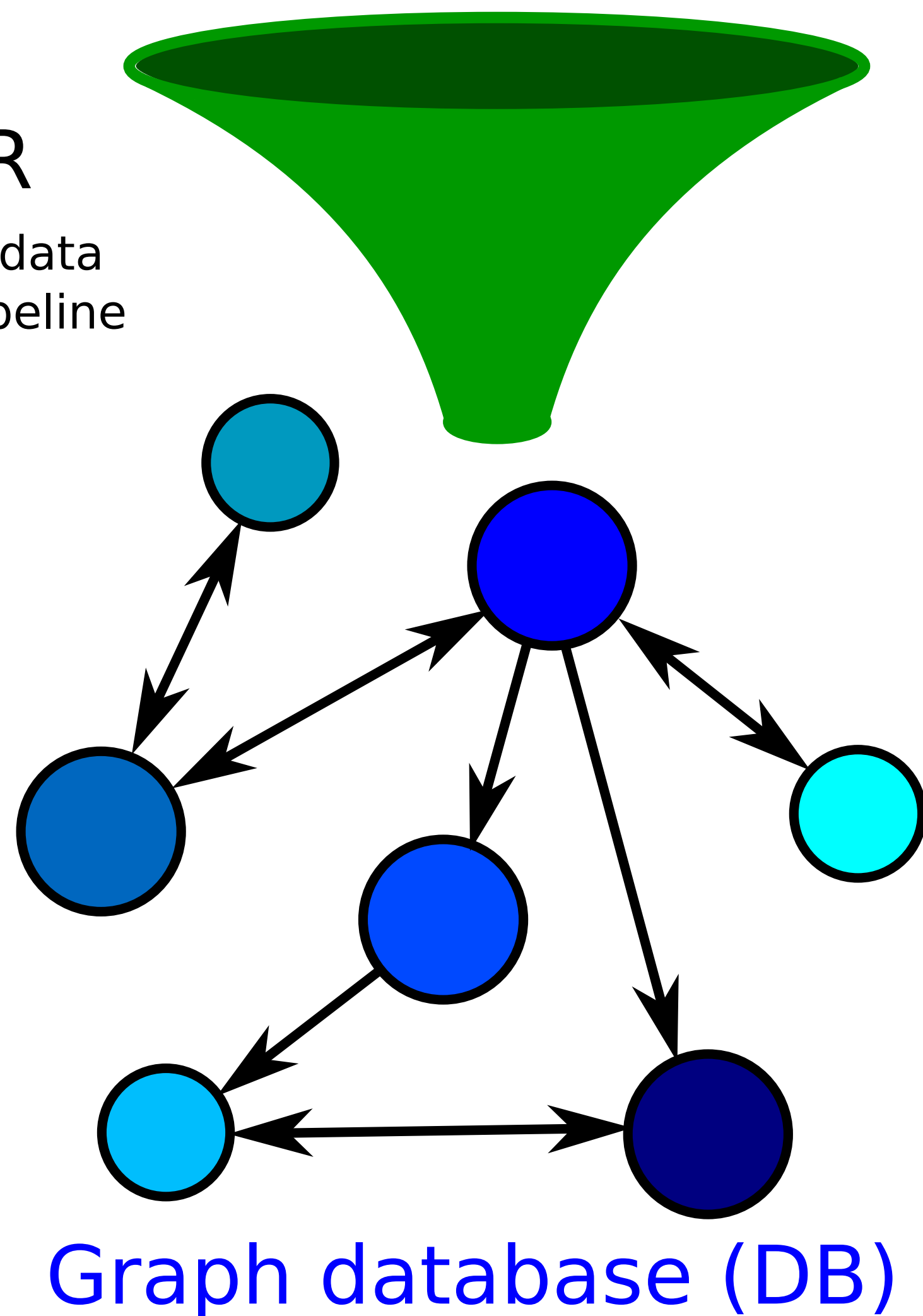
transform carefully

**Unify data:** Build on existing standards (HL7/FHIR [1], openEHR [2], bio-ontologies [3], COMBINE [4], etc.)  
**FAIRify [5] data:** Add data provenance information  
**DQA & data enrichment:** Property assignment (scores for similarity measures), automated semantic enrichment

high quality  
FAIR data

load FAIR

**Generalized** data integration pipeline



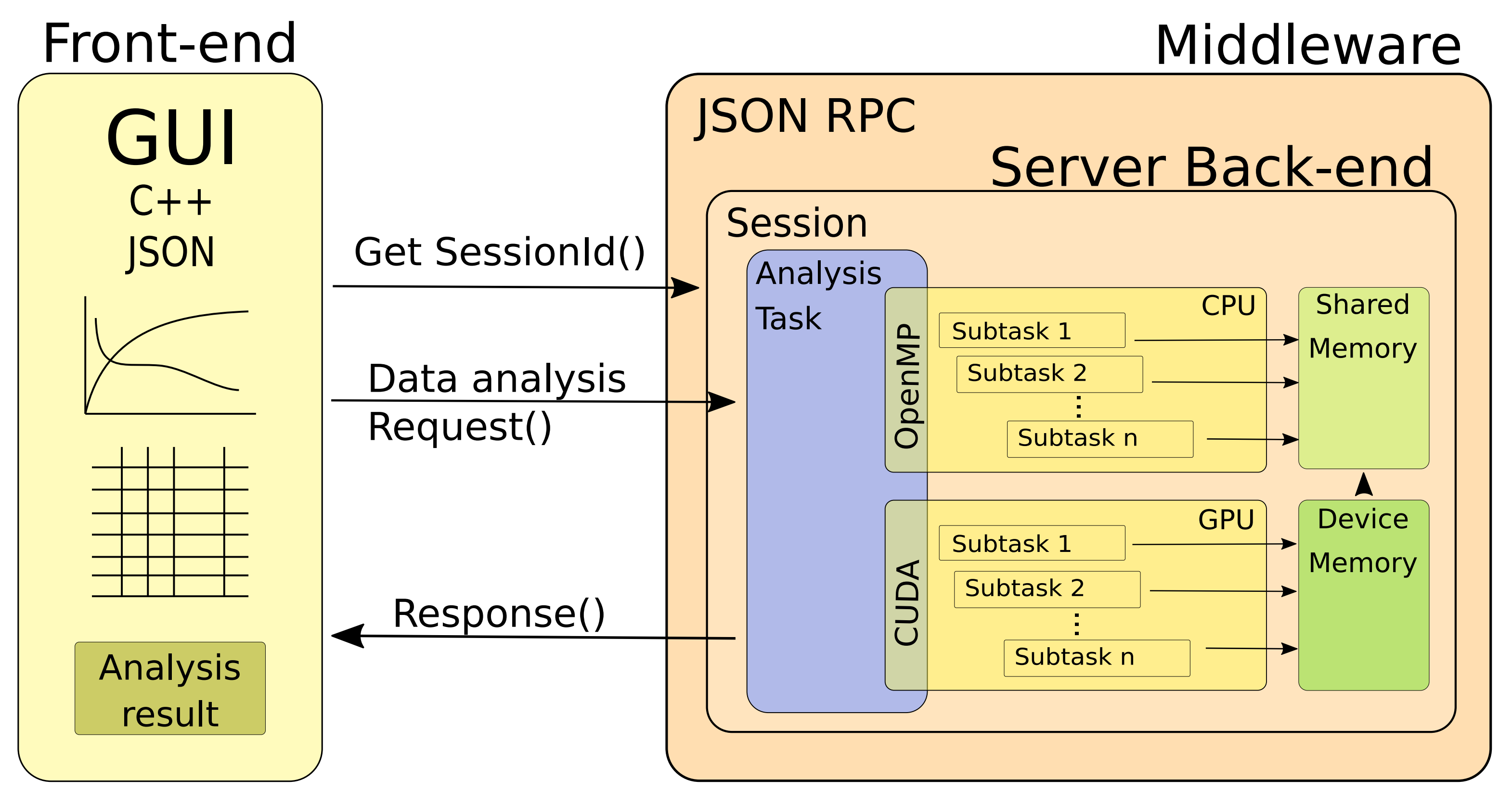
## GOAL: MeDaX information and research platform

**Main principles** to free and connect biomedical data while fulfilling highest legal, ethical and quality standards:

- i) **Embrace complexity:** Implement a graph DB - appropriate for highly interconnected, heterogeneous data [6-10] - and (semi-)automatic data storage and query pipelines, including semantic enrichment and metadata management.
- ii) **Be efficient:** Re-use and/or improve existing concepts, algorithms, and tools for graph DB, data quality assessment (DQA), data provenance (DP) and similarity measures. Design new ones only where required.
- iii) **Stay safe:** Secure and safe ETL pipeline and database architecture (considered at all project stages): Anonymized data > pseudonymized (consented) data.

## Usability

We will provide an intuitive graphical user interface (GUI) that allows clinicians and scientists to generate new knowledge from interconnecting diverse data, e.g., by obtaining information for patient-specific decision support or modelling data.



## Cooperations are welcome!

The MeDaX project does not produce new biomedical data itself, but instead generates new knowledge from combining already existing data (improving data re-usability). We are highly interested in cooperations with clinicians, scientists, data privacy advocats and everybody else from the interested public as we believe that a sound concept for the MeDaX information and research platform incorporates different stakeholder interests right from the start. In addition, this allows to maximize data quality and benefits for everyone and, thus, ultimately for patients.

## Acknowledgements and funding



all MILAs@UMG for everything :)



## Data sources & references

- [A] MII core data set [11]
- [B] Data Integration Centre @UMG
- [C] population studies [12,13]
- [D] public information portals [15,16]
- [1] D. Bender, K. Sartipi, *Proceedings of the 26th IEEE international symposium on computer-based medical systems* 2013
- [2] D. Kalra et al., *Studies in health technology and informatics* 2005
- [3] M. Salvadores et al., *Semantic web* 2013
- [4] D. Waltemath et al., *J integrative bioinf* 2020
- [5] M. D. Wilkinson et al., *Scientific data* 2016
- [6] C. T. Have, L. J. Jensen, *Bioinformatics* 2013
- [7] S. G. Finlayson et al., *Scientific data* 2014
- [8] I. Balaur et al., *J Comp Biol* 2017
- [9] A. Fabregat et al., *PLoS comp biol* 2018
- [10] D. S. Himmelstein et al., *Elife* 2017
- [11] T. Ganslandt, M. Boeker, M. Löbe et al., *Forum d. Medizin-Dokumentation und Medizin-Informatik* 2018, p.17
- [12] U. John et al., *Sozial-und Präventivmedizin* 2001
- [13] H.J. Grabe et al., *J translational medicine* 2014
- [14] N.F. Noy et al., *NAR* 2009
- [15] nfdi4health, <https://nfdi4health.de/>
- [16] S. Thun et al., *BMC Med Inform Decis Mak* 2020
- [17] L.A. Kapsner et al., *Appl Clin Inform* 2021