

GHGA
THE
GERMAN
HUMAN
GENOME -
PHENOME
ARCHIVE

GHGA

*A national data infrastructure
for genomic data*

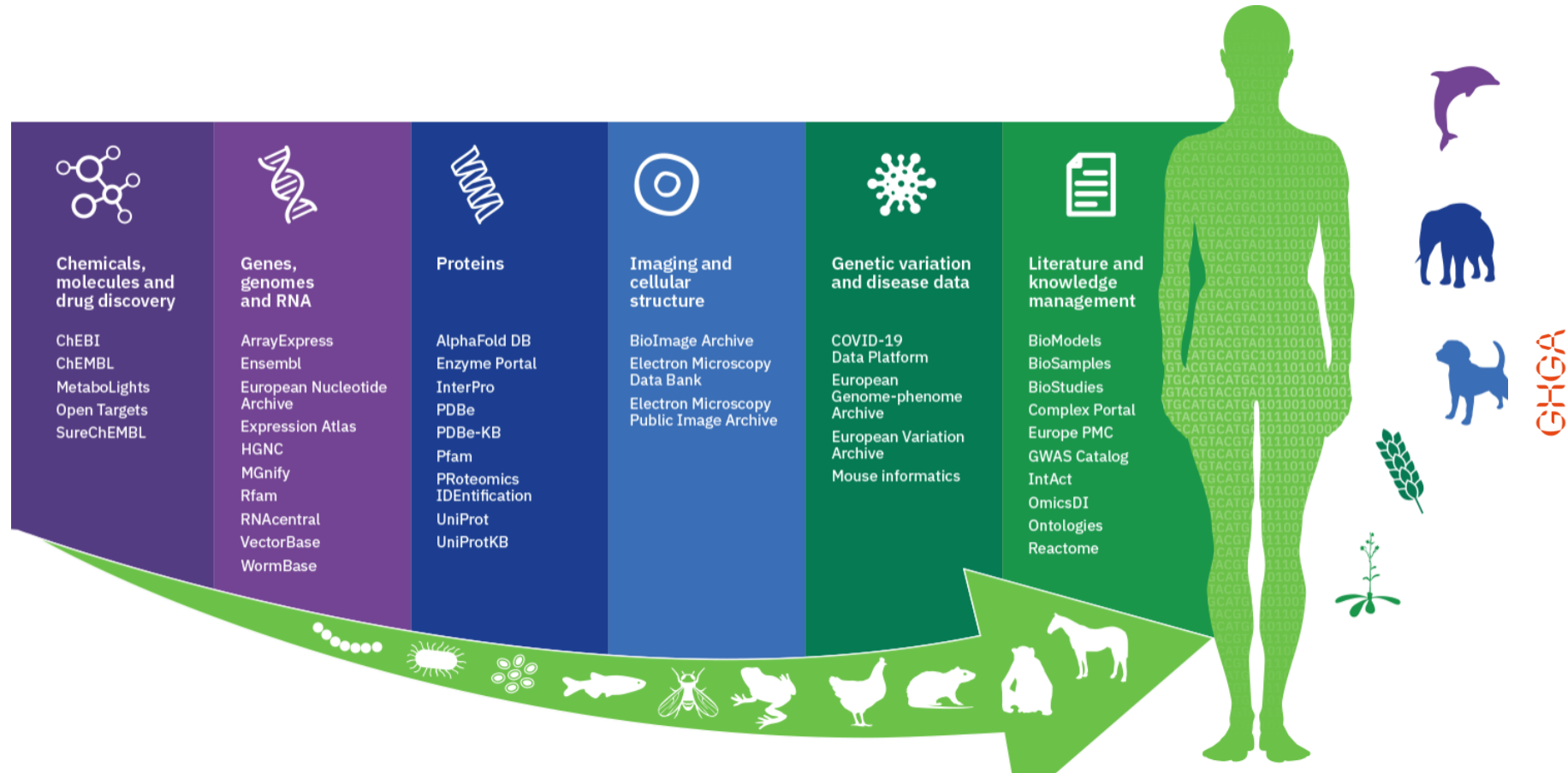
Oliver Kohlbacher
University of Tübingen and
University Hospital Tübingen

Oliver Stegle
DKFZ, Heidelberg and
EMBL, Heidelberg

In cooperation with

nfdi

Bio Data Resources to Power AI



Data Growth in Healthcare – Powering AI

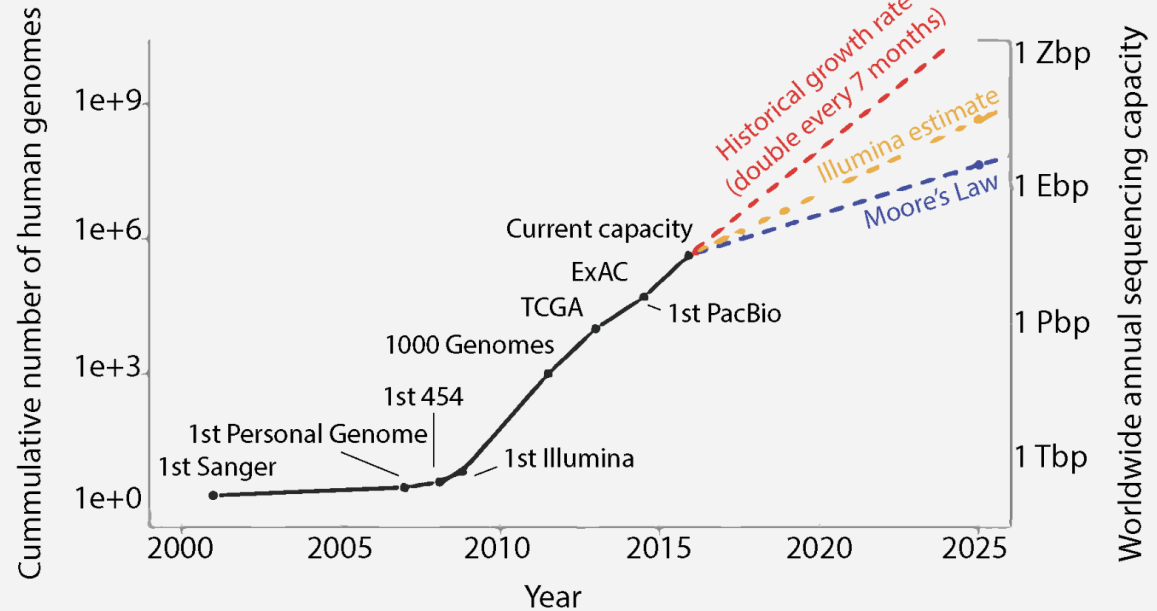
Environmental data

Imaging

Phenotype data

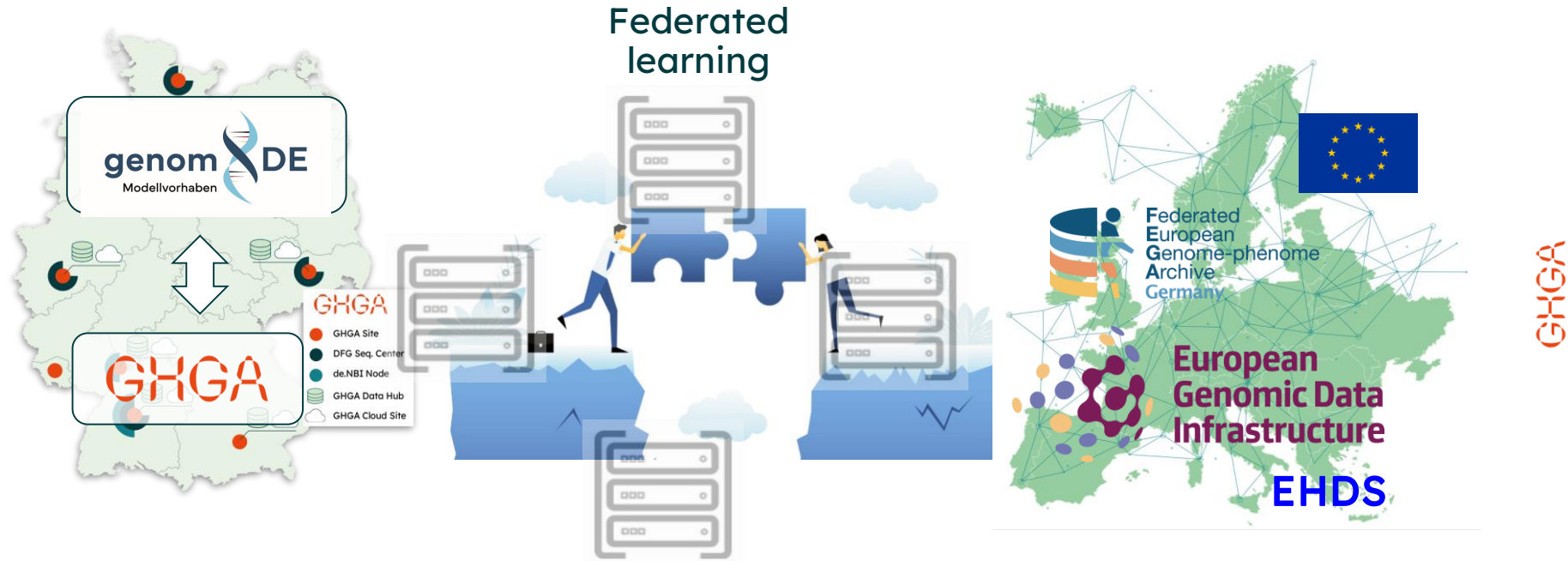
Single-cell profiling

Genome sequencing



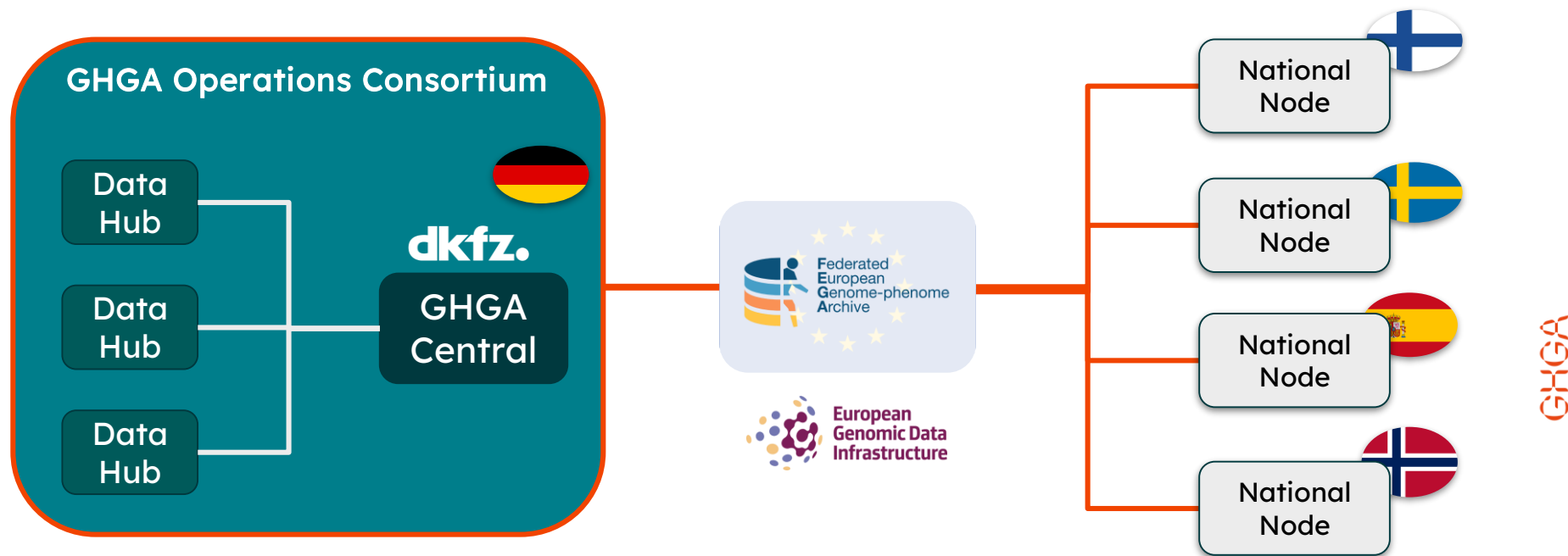
* adapted from Stephens et al. PLOS Bio 2015

Closing the Gap - Connection to Europe



GHGA connects *national* and *European* health data spaces.

Federation Concept of GHGA



Dual Federation

- **Within Germany**
 - Federal Data Protection Office (BfDI) demands federated storage of genome data
 - Sufficiently large infrastructure is not available in one place
- Data distributed across infrastructure from different partners within Europe
 - Genomic data sharing is being reorganized within Europe in a federation (FEGA)
 - GHGA is the German national node within FEGA

GHGA Service Roadmap



GHGA Metadata Catalog

- First operational phase
- A collection of existing national resources
- Standardised data set library with (EGA compatible) GHGA Metadata Model



GHGA Archive

- Harmonised Metadata
- Streamlined data deposition
- National FEGA node
- Unified ethico-legal framework



GHGA SPE

- Restricted SPE (rSPE) for internal use
- Community SPE (cSPE) for external use



GHGA Atlas

- Standardised data analysis
- Data visualisation
- Statistics and aggregation
- Integration of multiple omics modalities and connecting omics data to phenotypic data

81 studies containing data from 3,200 individuals (WGS, WES, transcriptomes)

UP to 500k genomes (or related omes) with intention to deposit

GHGA 2.0
(2025-2030)

The German Human Genome-Phenome Archive Data Portal

The GHGA Data Portal is a secure national infrastructure for human omics data available under controlled access. Access to the archived data can be requested from the data controllers who are responsible for evaluating the data access request.

The datasets within are annotated following the [GHGA Metadata Model](#) [↗](#), which is compatible with the metadata model of the EGA.

For further documentation, please visit the [GHGA User Documentation](#) [↗](#).

Browse data

Statistics



Total datasets:

71



Experiments: 1588

478 ILLUMINA HISEQ 2000
27 ILLUMINA HISEQ 2500
30 ILLUMINA HISEQ 4000
44 ILLUMINA HISEQ X
988 ILLUMINA NOVASEQ 6000
18 MGISEQ-2000RS
2 NEXTSEQ 550
1 UNSPECIFIED



Individuals: 16444

5747 FEMALE
6299 MALE
4398 UNKNOWN



Files: 38924

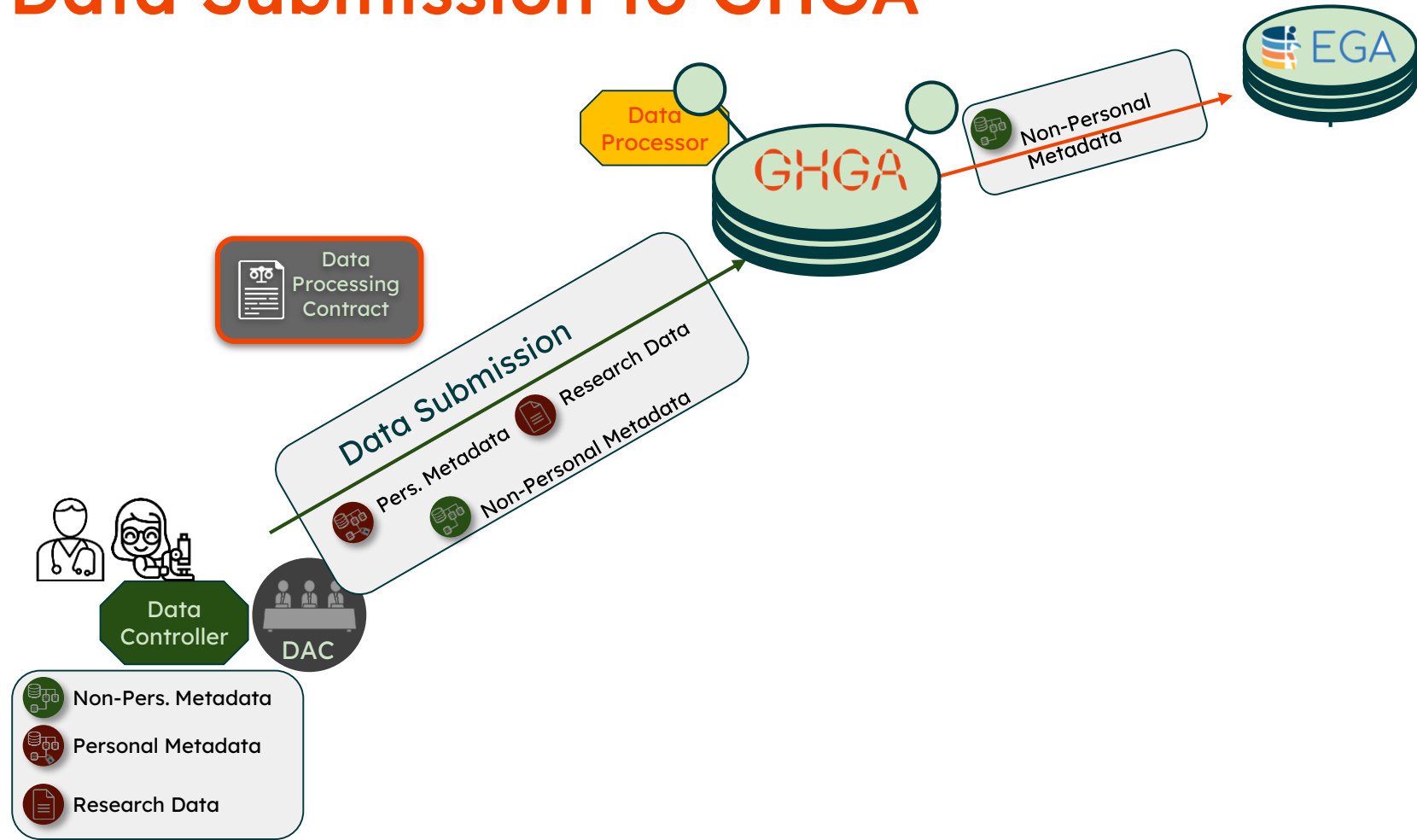
11 TXT
3444 BAM
2579 OTHER
2 VCF
9120 FASTQ
23768 IDAT

Projects that are planning to use GHGA

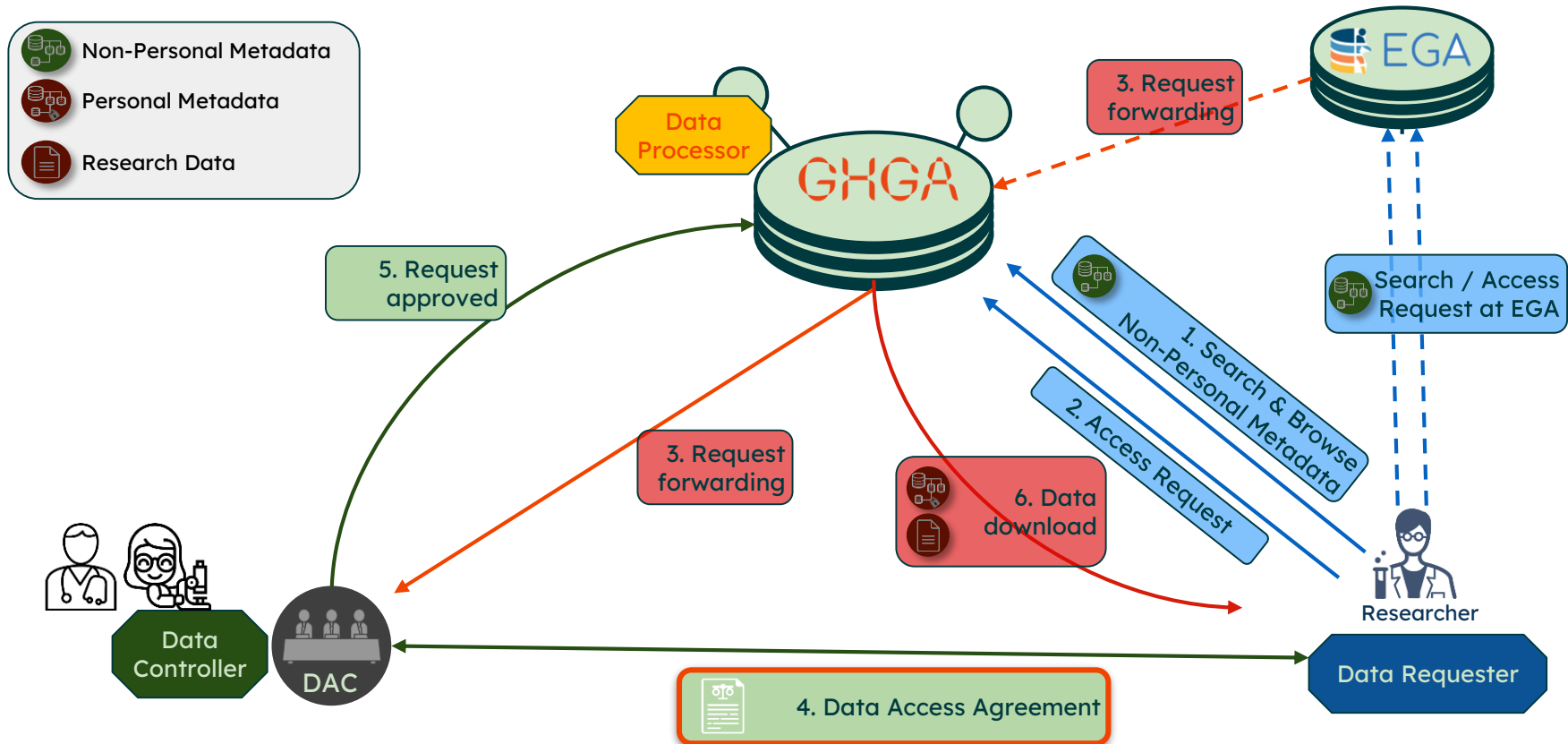
- In general: GHGA will be open for data submissions from all German research institutions
- We have inquiries from projects aiming for a total of 500,000 genomes in the next 5 years



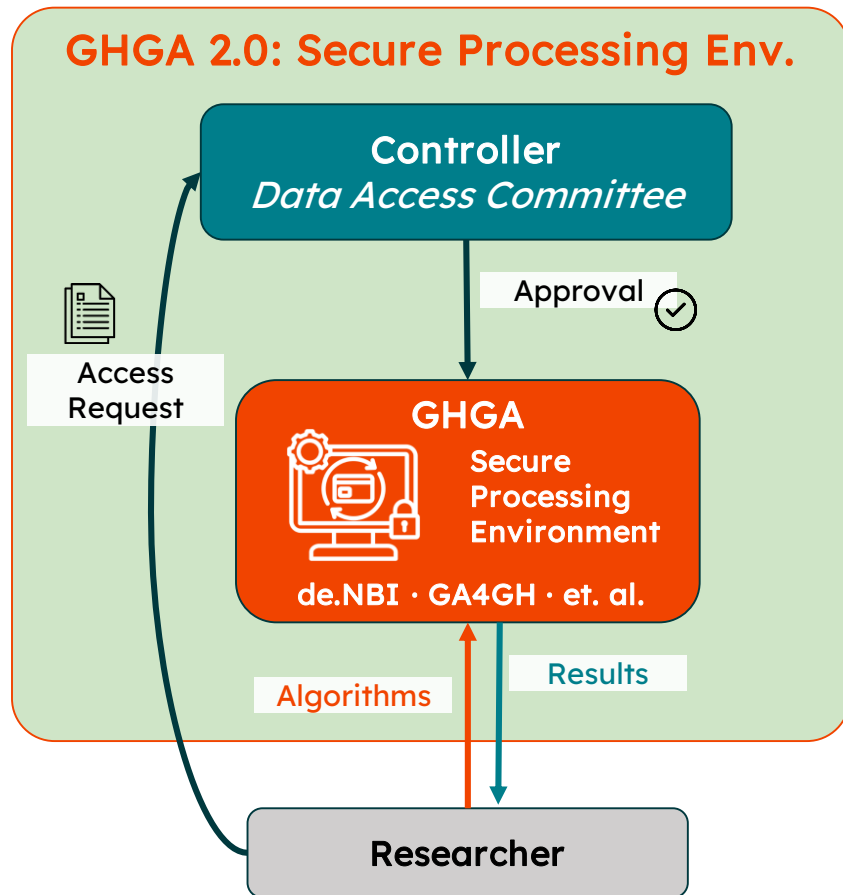
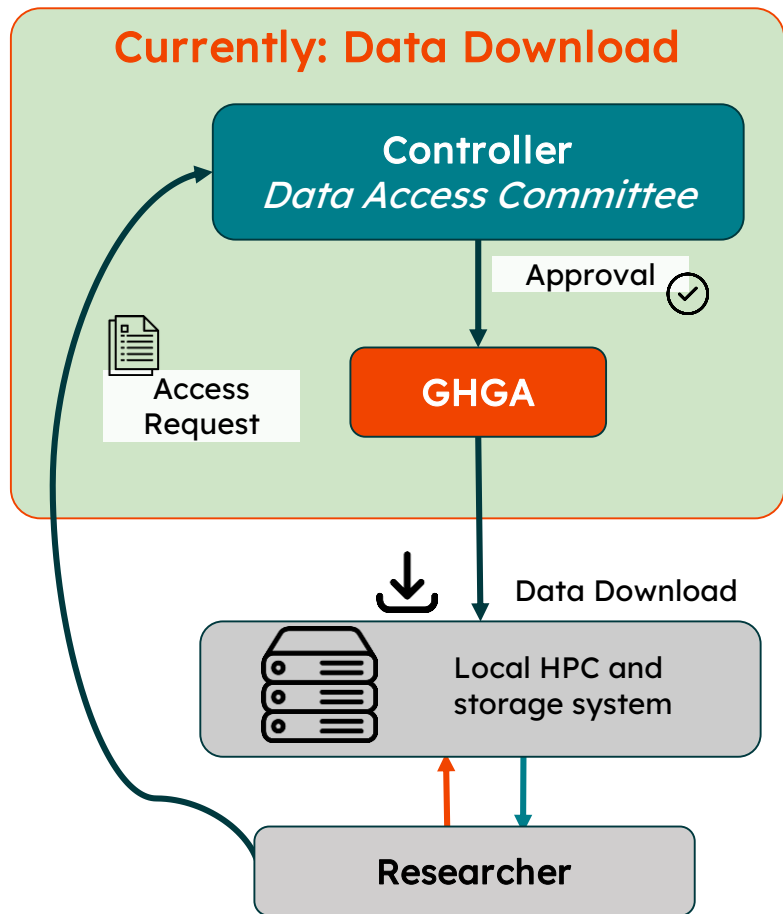
Data Submission to GHGA



Data Access in GHGA



Secure Processing Environments in GHGA



genomDE Model Project

Mission: Bring genome sequencing to routine care in rare disease and oncology

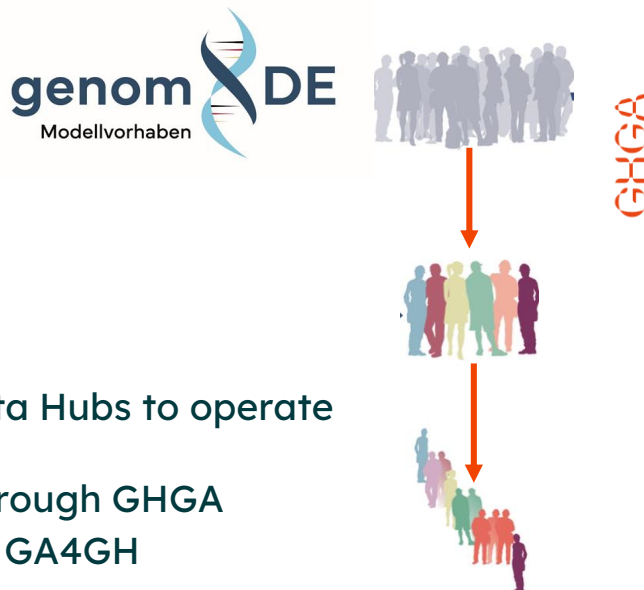
- Based on national law (§64e SGB V)
- Funded for by health insurance and BMG
- Designed to be **research compatible**
- Legal basis for research: **MII BC 1.6d+**
- Estimated 90,000 whole exomes & genomes, corresponding to **~15 PB of raw and processed data**

GHGA's role

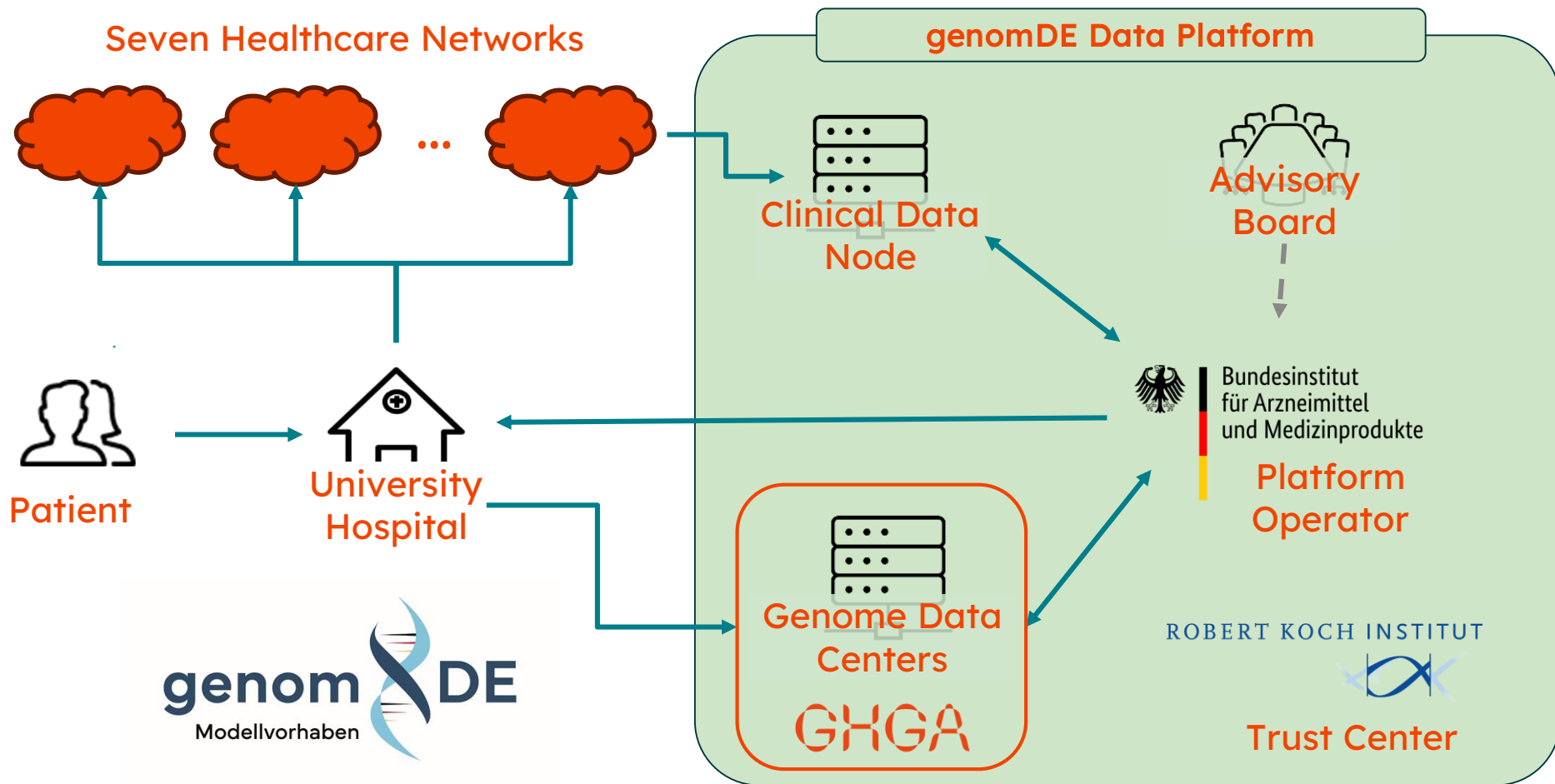
- Conception of the data platform
- GHGA will provide genome data management ⇒ GHGA Data Hubs to operate Genome Data Centers
- Research-consented data (expected 90+%) to be shared through GHGA
- Connect to international activities such as GDI / 1+MG and GA4GH



Fed. Minister of Health Prof. K. Lauterbach launches genomDE Model Project, July 2024



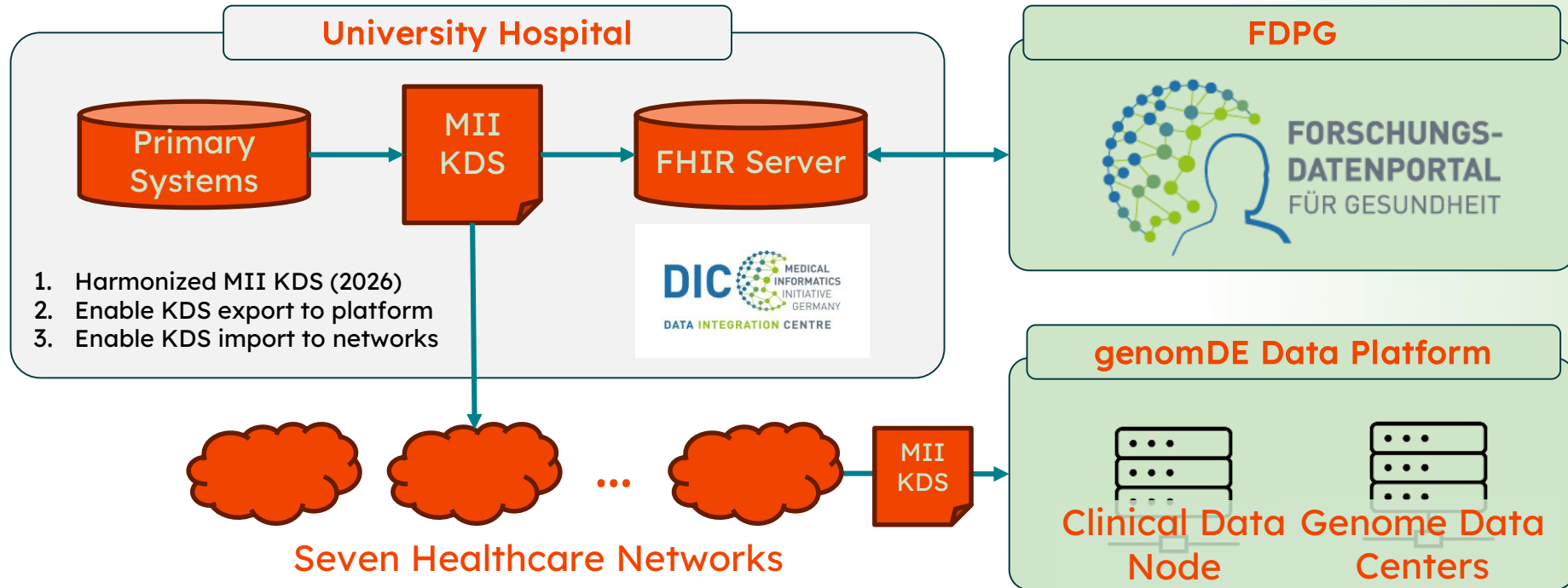
Integration with genomDE



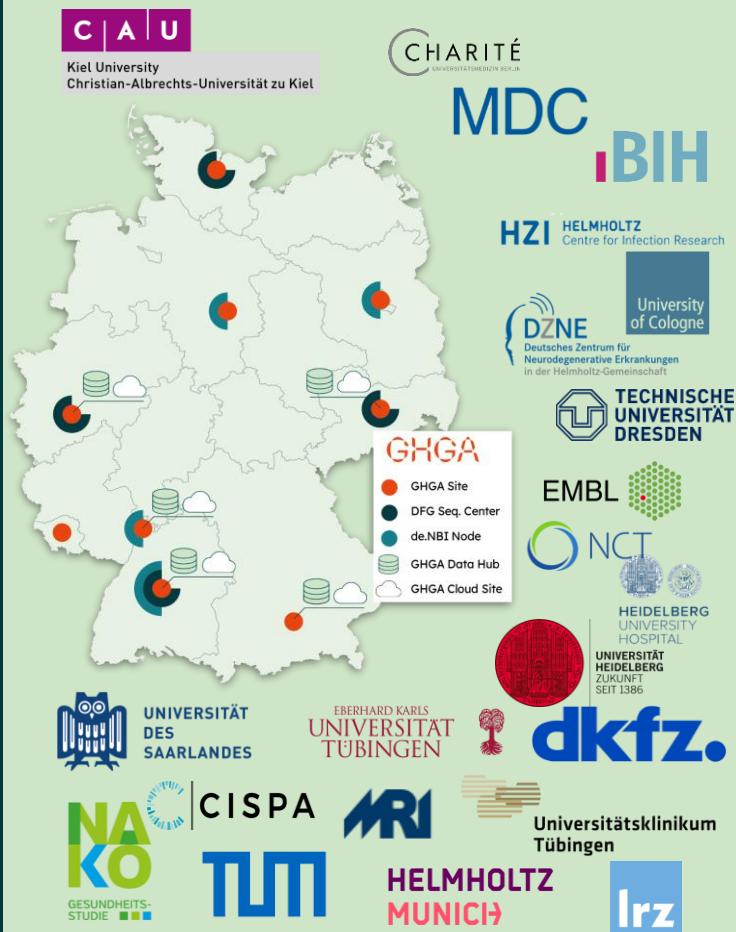
Integrating genomDE with MII/NUM

Goals

- Integrate the data from genomDE MP with MII KDS
- Make genomDE data findable through FDPG



GHGA Partners



GHGA Team



Stay informed

ghga.de

bit.ly/GHGANewsletter

In cooperation with

nfdi

Funded by

DFG Deutsche Forschungsgemeinschaft