

GHGA
THE
GERMAN
HUMAN
GENOME -
PHENOME
ARCHIVE

*The German Human
Genome-Phenome Archive
(GHGA)*

eosc | FAIR-IMPACT **National Roadshow Series**

ONLINE WORKSHOP
Germany
24th January 2024

Partner **nfdi**
Nationale Forschungsdaten Infrastruktur

FAIR-IMPACT events
National Roadshows

Dr. Anandhi Iyappan,
Metadata Standards
Officer, EMBL

24-01-24

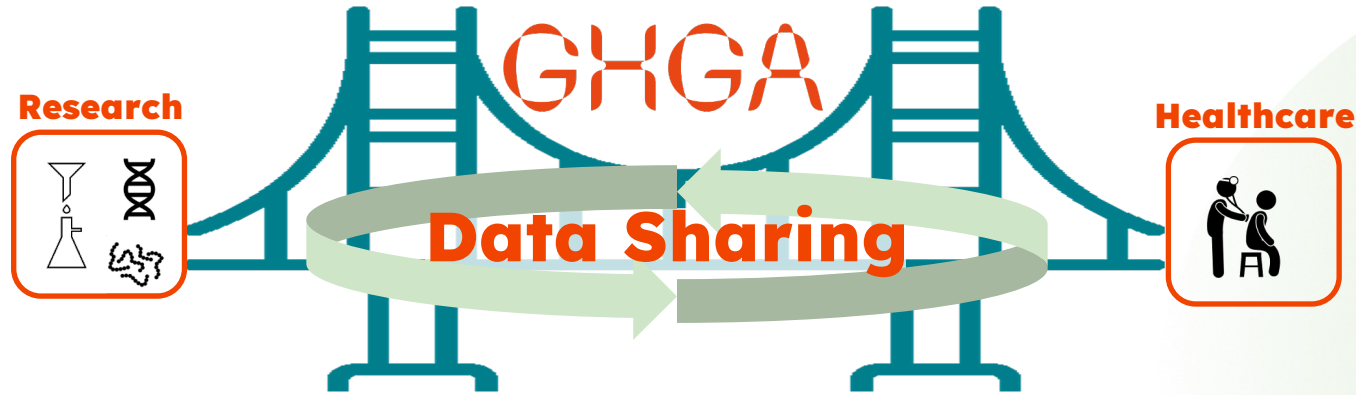
HELMHOLTZ

dkfz. GERMAN
CANCER RESEARCH CENTER
IN THE HELMHOLTZ ASSOCIATION
Research for a Life without Cancer

In cooperation with

nfdi

The GHGA Vision: Enabling Genomic Medicine



- Secondary use of clinical omics data in research: biological discovery & replication of findings to show validity
- Rapid (exponential) growth of available data is a major challenge and opportunity

⇒ Establishment of a nationally coordinated, interdisciplinary infrastructure integrating genome research and healthcare

Data Types in GHGA

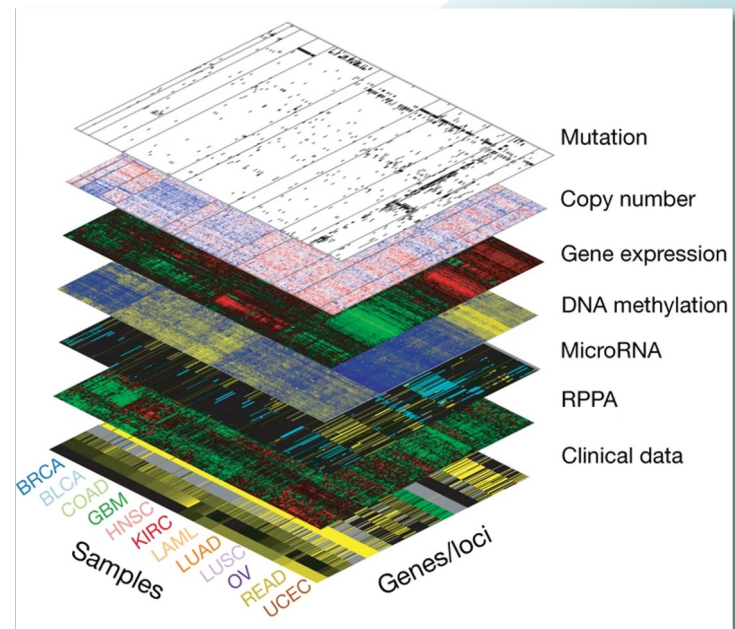
- HUMAN Genome, Exome, Epigenome, Transcriptome (including bulk and single cell data)
- ⇒ **Connection to clinical metadata (“Phenome”) is key!**

Common for all Data Types:

1. Highly sensitive ⇒ not anonymizable!
2. Often high data volumes

Regulated Data Access:

- **Access to raw data** is often essential for research
- Access only under “Controlled Access” with clear contractual regulations according **to GDPR and patient consent**



Source: ICGC/TCGA; The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 45, 1113–1120 (2013)

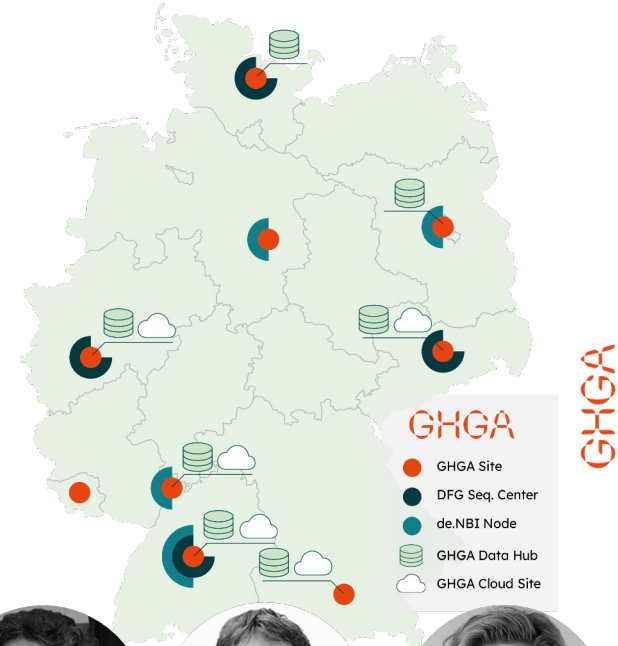
The GHGA Consortium

- Funded since 10/2020 as one of the initial nine first-round NFDI consortia (currently until 09/2025)
- Representing 10 universities, 6 Helmholtz & 5 other research institutions (17 funded partners)
- Network of seven data hubs co-located with major academic sequencing centers (TÜ, K, DD, KI, HD, M, B)
- Connected to national cloud infrastructure

⇒ **de.NBI cloud**



- 15 M€ funding over five years plus own contributions (at least 5 M€) by the partner institutions
- **DKFZ as the coordinating institution**



O. Kohlbacher
(Univ. Tübingen)



J. Korbel
(EMBL)



O. Stegle
(DKFZ & EMBL)



E. Winkler
(Univ. Heidelberg)

GHGA Board of Directors

GHGA Workstreams

Architecture



O. Kohlbacher
O. Stegle
L. Kuchenbecker
K. Breuer

Data Hub: Operations & Data Stewardship



S. Motameny
I. Buchhalter
R. Schregle
P. Menges

Metadata



S. Nahsen
T. Ulas
A. Iyappan
K. Mauer

Workflows



J. Gagneur
S. Ossowski / D. Hübschmann
C. Mertes
F. Heyl

ELSI



E. Winkler
F. Molnar-Gabor
A. Bruns
S. Parker

Training



W. Huber
O. Kohlbacher
J. Philipp
N. Schatlowksi

Public and Community Outreach



J. Walter
J. Winkelmann
U. Träger
N. Gasparoni

Project Management



O. Stegle
O. Kohlbacher
J. Eufinger
N. Schatlowksi

National and International Context

GHGA Phases and Features



GHGA Metadata Catalog

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



GHGA Archive

- National EGA functionality
- Streamlined data deposition
- Central infrastructure for data access committees
- Unified ethico-legal framework
- Harmonised Metadata



GHGA Cloud

- Cloud-based analytics platform (Paas) for large-scale omics data
- Community-specific data portals



GHGA Atlas

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



The German Human Genome-Phenome Archive

Metadata Catalog



The GHGA Metadata Catalog is a discovery platform for human omics data available for research that can be requested from one of the GHGA Data Hubs.

The public catalog is a first step towards our goal to provide comprehensive data archival services for human omics data.

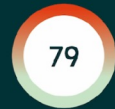
The datasets within are annotated following the [GHGA Metadata Model](#), which is compatible with the metadata model of the EGA.

Please also see further documentation under ["How to get data access"](#), and ["Submit data to GHGA"](#)

[GET STARTED](#)

Statistics

Total Datasets:



Platforms: 9

Illumina HiSeq 4000:	30
HiSeq X Ten:	45
Illumina HiSeq 2500:	28
Illumina HiSeq 2000:	15
NextSeq 500:	1
Illumina HiSeq X:	1



Individuals: 2597

Female: 1228 Male: 1369

Unknown: 0



Files: 11102

FASTQ: 8420
BAM: 2541
VCF: 141

Our Partners

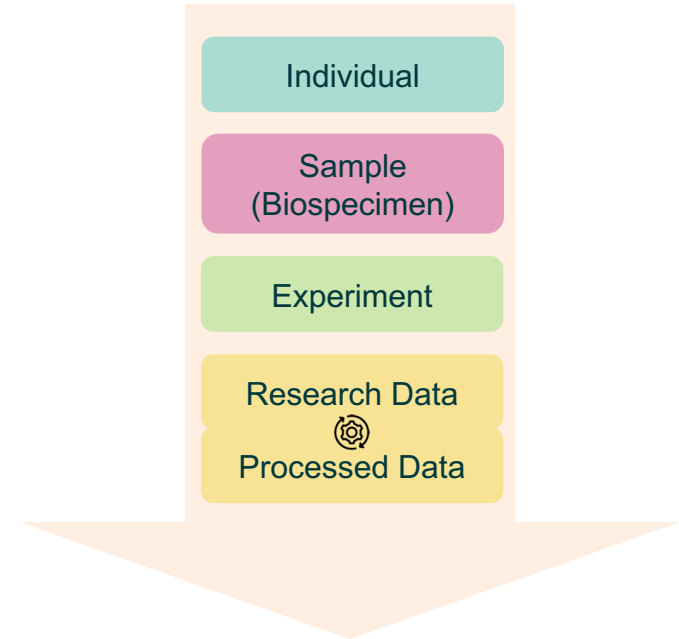
Metadata



Metadata Schema for GHGA Archive

- Based on EGA with implementation of selected ontologies and vocabularies
- Built using LinkML framework
github.com/ghga-de/ghga-metadata-schema
- User documentation to support submission docs.ghga.de/metadata
- Further Extension and alignment with fEGA and GDI
- Standards and Ontologies

Schema Overview



Whitepaper: Metadata Schema for the German Human Genome-Phenome Archive. Zenodo.
[10.5281/zenodo.8341224](https://doi.org/10.5281/zenodo.8341224)

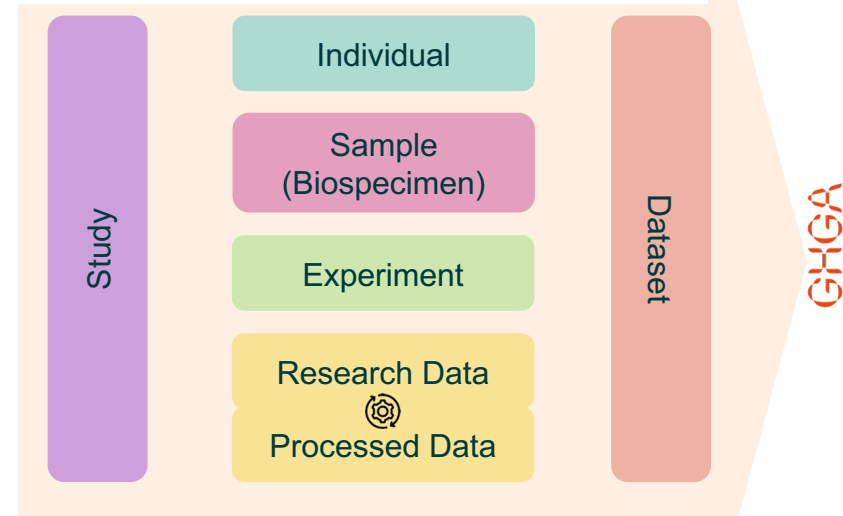
Metadata



Metadata Schema for GHGA Archive

- Based on EGA with implementation of selected ontologies and vocabularies
- Built using LinkML framework
github.com/ghga-de/ghga-metadata-schema
- User documentation to support submission docs.ghga.de/metadata
- Further Extension and alignment with fEGA and GDI
- Standards and Ontologies

Schema Overview



Whitepaper: Metadata Schema for the German Human Genome-Phenome Archive. Zenodo.
[10.5281/zenodo.8341224](https://doi.org/10.5281/zenodo.8341224)

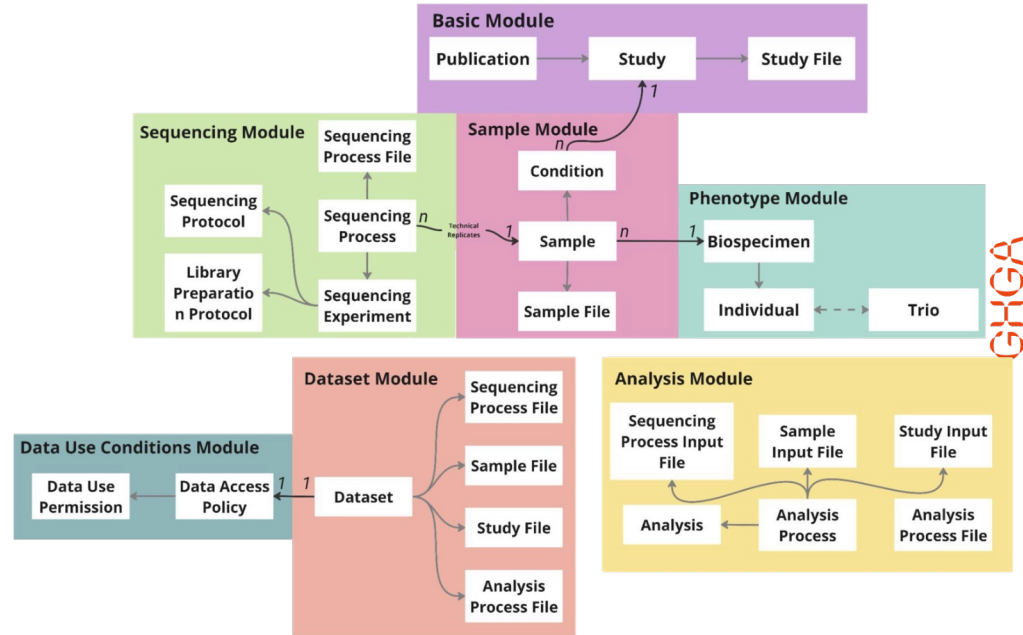
Metadata



Metadata Schema for GHGA Archive

- Based on EGA with implementation of selected ontologies and vocabularies
- Built using LinkML framework
github.com/ghga-de/ghga-metadata-schema
- User documentation to support submission docs.ghga.de/metadata
- Further Extension and alignment with fEGA and GDI
- Standards and Ontologies

Schema Overview



Whitepaper: Metadata Schema for the German Human Genome-Phenome Archive. Zenodo.
[10.5281/zenodo.8341224](https://doi.org/10.5281/zenodo.8341224)

GHGA Co-Spokespersons

- Peer Bork (EMBL)
- Ivo Buchhalter (DKFZ)
- Andreas Dahl (TU Dresden)
- Julien Gagneur (TUM)
- Wolfgang Huber (EMBL)
- Daniel Hübschmann (DKFZ)
- Dirk Jäger (UKHD)
- **Oliver Kohlbacher (EKUT)***
- **Jan Korbel (EMBL)***
- Martin Lablans (DKFZ)
- Ulrich Lang / Stefan Wesner (UzK)
- Peter Lichter (DKFZ)
- Fruzsina Molnár-Gábor (UHEI)
- Susanne Motameny (UzK)
- Sven Nahnsen (EKUT)
- Uwe Ohler (MDC)
- Stephan Ossowski (UKT)
- Annette Peters (HMGU, NAKO)
- Olaf Rieß (UKT)
- Philip Rosenstiel (CAU)
- Thorsten Schlomm (Charité)
- Joachim Schultze (DZNE)
- **Oliver Stegle (DKFZ/EMBL)***
- Jörn Walter (USAAR)
- Thomas Walter (EKUT)
- Juliane Winkelmann (HMGU, TUM)
- **Eva Winkler (NCT/UKHD)***

***: Board of Directors**

GHGA Team



Stay informed

ghga.de @GHGA_DE

bit.ly/GHGANewsletter

Partner Institutions

