

GHGA
THE
GERMAN
HUMAN
GENOME -
PHENOME
ARCHIVE

The German Human Genome-Phenome Archive (GHGA)

*A national research data infrastructure
for human omics data*

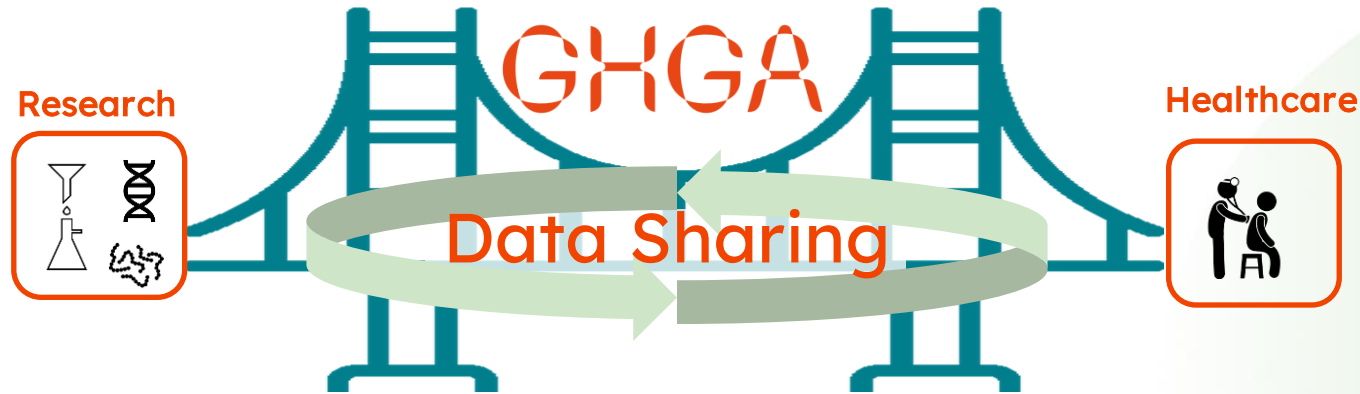
MIRACUM-DIFUTURE Symposium 2023
October 10, 2023

Oliver Kohlbacher
GHGA Co-Director

In cooperation with

nfdi

The GHGA Vision: Enabling Genomic Medicine



GHGA

Main Aims:

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

- From Researchers for Research-

Funding: NFDI - The National Research Data Infrastructure Initiative

- ❖ FAIR access to research data for all and across all domains of science.
- ❖ Funded with 900 M€ for 10 years
- ❖ 26+1 consortia selected for funding in 3 cohorts
- ❖ Wide variety of topics, including lifes sciences (NFDI4Health, NFDI4Microbiota, NFDI4Bioimage, NFDI4Immuno) with strongest interaction
- ❖ BASE4NFDI-Consortium starting in March 2023 ⇒ Development of common Base-Services

Bundesministerium für Bildung und Forschung

< Nationale Forschungsdateninfrastruktur

22.04.2021 | FORSCHUNG

Nationale Forschungsdateninfrastruktur

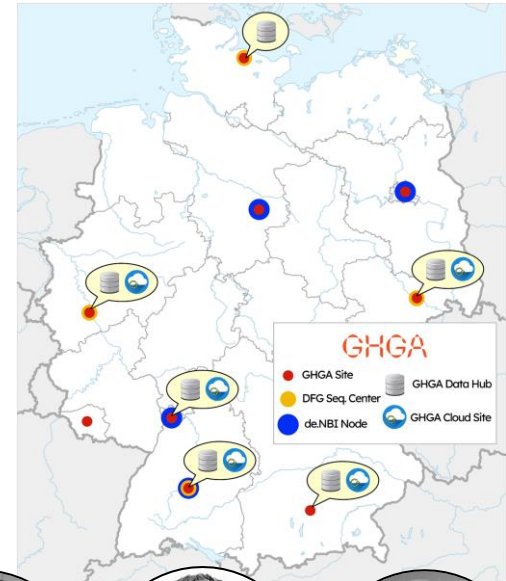
Forschungsdaten werden oft nur lokal und vorübergehend gespeichert. Dadurch gehen Datensätze verloren. Mit der Nationalen Forschungsdateninfrastruktur möchte das BMBF das ändern: Sie soll ein Wissensspeicher für die ganze Forschungslandschaft sein.



Die NFDI soll Standards im Datenmanagement setzen und als digitaler, regional verteilter und vernetzter Wissensspeicher Forschungsdaten nachhaltig sichern und nutzbar machen.

The GHGA Consortium

- Funded since 10/2020 as one of the initial nine first-round NFDI consortia
- Leading Institutions in genomic medicine
 - 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
- Funding from end of 2020 to 2025



O. Kohlbacher
(Univ. Tübingen)



J. Korbel
(EMBL)



O. Stegle
(DKFZ & EMBL)



E. Winkler
(Univ. Heidelberg)

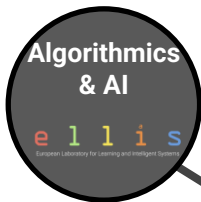
GHGA Board of Directors

GHGA - Interfaces and Related Initiatives

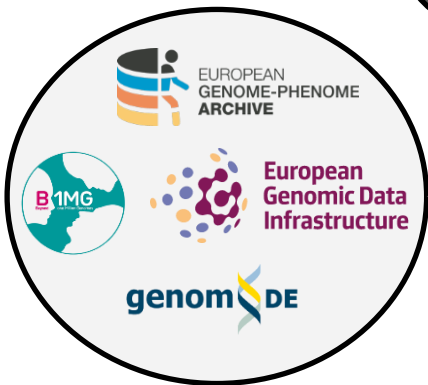


Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.

Technical
Harmonization



UACs
(Use & Access)



Data Exchange

APIs

GHGA

Tech Standards

ELSI Standards

HPC Infrastr.

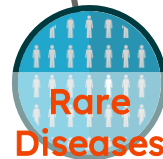
Portals



NGS-CN



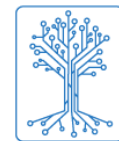
Cancer
Disease-related
communities



Rare
Diseases



deNBI
GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE

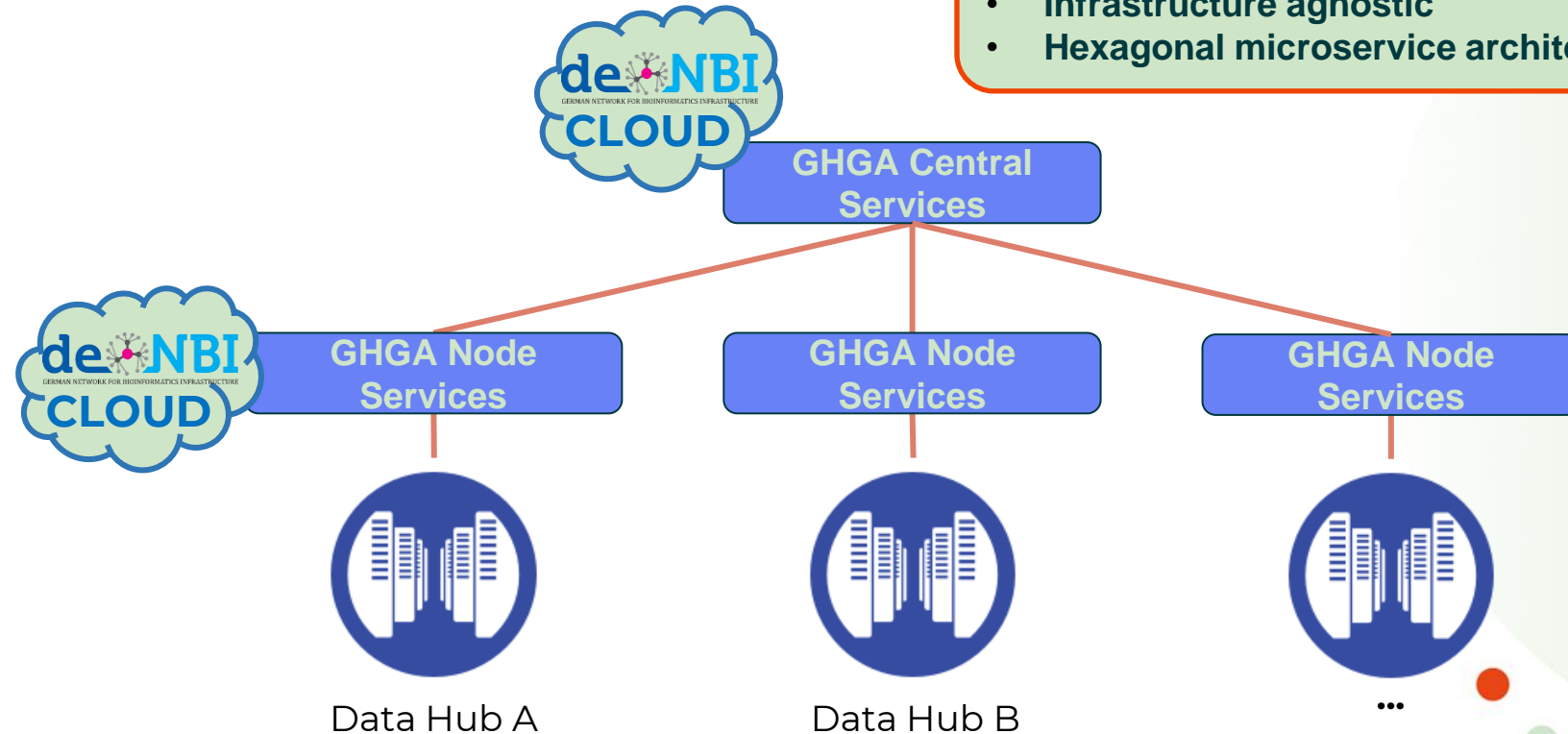


GAIA-X

Technical
Infrastructure

GHGA Architecture

- **Federated** (GHGA Central & Data Hubs)
- **Cloud-native**
- **Infrastructure agnostic**
- **Hexagonal microservice architecture**



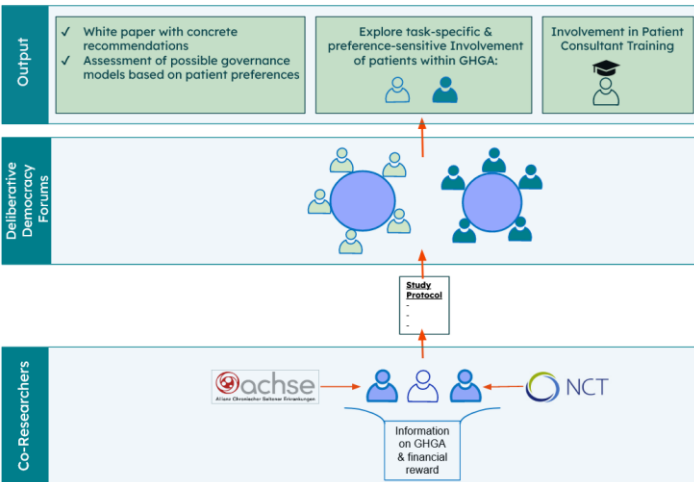
Project Status

Legal and Ethical Questions, Patient Engagement

Patient Engagement

PaGODa Study: "Patientenperspektiven auf Governance-Strukturen von Datenbanken"

Apondo, Winkler et al.



Whitepaper on Consent Modules

July 13, 2022

Project milestone Open Access

Consent Modules for Data Sharing via the German Human Genome-Phenome Archive (GHGA)

Bruns, Andreas; Benet-Pages, Anna; Eufinger, Jan; Graessner, Holm; Kohlbacher, Oliver; Molnár-Gábor, Fruzsina; Parker, Simon; Schickhardt, Christoph; Stegle, Oliver; Winkler, Eva

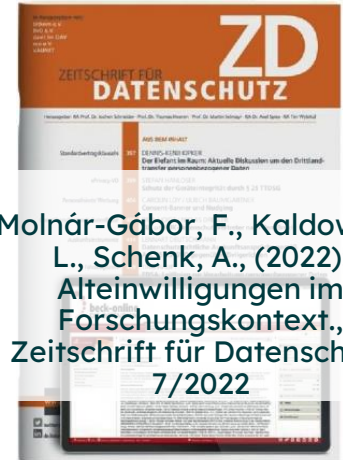
The German Human Genome-Phenome Archive (GHGA) is a research consortium which is currently establishing a federated data infrastructure allowing the secure storage of and controlled access to omics and related health data consented for scientific research use. This whitepaper contains guidance for Data Submitters on how to update their consent forms such that data may be shared via GHGA for secondary research use.

Funding: The GHGA consortium is funded by the German Research Foundation (DFG) within the framework of the National Research Data Infrastructure (NFDI).



<https://zenodo.org/record/6828131#.Yzvmbi0Rp-V>

Legal Foundations



Molnár-Gábor, F., Kaldowski, L., Schenk, A., (2022): Alteinwilligungen im Forschungskontext., Zeitschrift für Datenschutz 7/2022

Outreach and Training

Also on Spotify, YouTube & iTunes
ghga.de/codedeslebens



Podcast



Stay up to date:
GHGA Newsletter
See www.ghga.de or



Lecture Series

Lecture Series (Virtual) Advances in Data-Driven Biomedicine



Ivo Gut
CNAG-CRG



Helen Parkinson
EMBL-EBI



Joachim Schulze
DZNE



Christine Mundlos
Achse e.V.



Paola Quattroni
HDI



Adrian Thorogood
University of Luxembourg



Barbara Prainsack
University of Vienna



Annette Peters
Helmholtz Munich



Becca Wilson
Newcastle University



William Rayner
Helmholtz Munich

Workshops

The German Human Genome-Phenome Archive
@GHGA_DE

Today we had our @GHGA_DE workshop at #ghf2022 .
Presentations by @jan_eufinger , Fruzsina Molnar-Gabor @UniHeidelberg , Julien Gagneur @gagneurlab
and Stephan Ossowski @uktuebingen .

6:36 nachm. · 17. März 2022 · TweetDeck

Public Engagement

The German Human Genome-Phenome Archive
@GHGA_DE

We really enjoyed our day at #scienceinthecity and can only recommend going to this great exhibit, which showcased a new project each day. @DKFZ

You can read our recap of the day here.
ghga.de/news/detail/gh...

Training

Webinar Series

Webinar
30.06.2022
14:00 (CEST)

Beginners Guide
to DNA and Sequencing

Seminar by
Nicholas Smith
GHGA / TU Munich

More info: www.GHGA.de

Webinar
07.10.2022
13:00 (CEST)

Research Consent

Introduction to consent
by
Andreas Bruns

GHGA consent tool
by
Simon Parlier

More info: www.GHGA.de

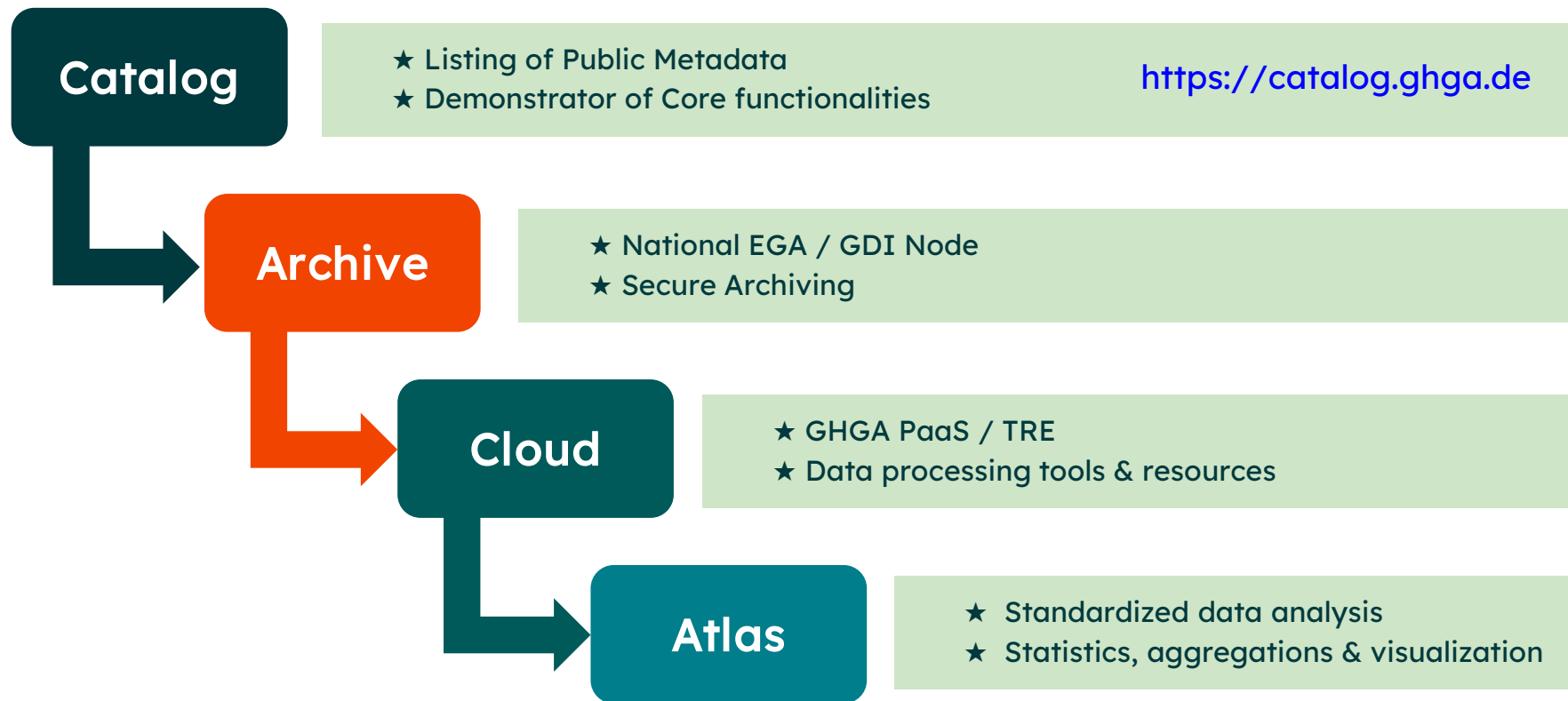
Courses / Collaborations

EMBL Bio-IT GHGA

FAIR in
(biological) practice

30.05 - 02.06.2022

Project Phases and Features



[Search](#)[DE](#)[HOME](#)[ABOUT US](#)[IMPACT](#)[NEWS & EVENTS](#)[RESOURCES](#)

The German Human Genome-Phenome Archive

We are building a **secure national omics data infrastructure**, enabling the use of human genome data for research purposes while preventing data misuse.

[Our Mission](#)

GHGA Metadata Catalog

A public frontend for the discovery of human omics study data from German research institutions.

Consent Tools

GHGA has developed different tools to help clinicians, researchers and institutions wanting to submit omics data to

GHGA Lecture Series

"Advances in Data-Driven Biomedicine" diving into fascinating world of data-driven medicine and their

[Who we are →](#) [How we work →](#) [Jobs →](#)

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

The screenshot shows the GHGA Data Portal interface. At the top, there is a navigation bar with the GHGA logo and 'DATA PORTAL' text. The navigation menu includes 'Home', 'Browse Data', 'Access Data', 'Submit Data', 'Metadata Model', and 'FAQ'. A 'Log in' button is located in the top right corner. Below the navigation bar, there are two buttons: 'Visit EGA Website' and 'Request Access'. The main content area displays the title 'Dataset for head and neck cancer RNA' and the Dataset ID 'EGAD00001008861'. It also shows the Study Type 'Cancer_genomics' and the Dataset Type 'Transcriptome Profiling By High-Throughput Sequencing'. A 'Description' section follows, stating: 'Rare cancer sequencing data of 92 runs in tumor/control pairs, which were uploaded to umbrella studies. The sequencing was always paired'. A date indicator shows '2023-05-11 Update date'. Below this, there are four filter buttons: 'Study', 'Project', 'Publication', and 'DAP/DAC'. The 'Study' filter is selected, leading to a detailed view for 'Study' EGAS00001004815. This view includes the title 'Comprehensive Genomic and Transcriptomic Analysis of Rare Cancers for Guiding of Therapy (H021)', a 'Description' paragraph, and the 'Type: Cancer_genomics'. At the bottom of the page, there are three expandable sections: 'Experiment Summary (12 experiments)', 'Sample Summary (61 samples)', and 'File Summary (184 files: 1.9 TB)'.

GHGA | The German Human Genome Archive | DATA PORTAL

Home Browse Data Access Data Submit Data Metadata Model FAQ CZ

Total Datasets: 65

Search datasets

Study ID

- EGAS00001004813 62
- EGAS00001005339 2
- EGAS00001006680 1

Project

- MASTER_remastered 62
- MNP2 1
- [] 2

Study Type

- Cancer_genomics 63
- Other 2

Dataset Type

- Exome Sequencing 22
- Transcriptome Profiling By High-Throughput Sequencing 21
- Whole Genome Sequencing 21
- Whole Genome Sequencing, Exome Sequencing, Transcriptome Profiling By High-Throughput Sequencing 1

Dataset ID	Title
EGAD00001008861	Dataset for head and neck cancer RNA
EGAD00001008863	Dataset for hepatopancreaticobiliary malignancy RNA
EGAD00001008859	Dataset for soft tissue tumor RNA
EGAD00001008862	Dataset for neuroendocrine adrenal tumor RNA
EGAD00001008893	Dataset for neuroendocrine adrenal tumor WHOLE GENOME
EGAD00001008858	Dataset for melanoma RNA
EGAD00001008856	Dataset for gynecologic cancer RNA
EGAD00001008847	Dataset for other cancer RNA
EGAD00001008860	Dataset for hematopoietic malignancy RNA
EGAD00001008855	Dataset for colorectal cancer RNA

< 1 2 ... 6 7 >

10

GHGA | The German Human Genome Archive

GHGA DATA PORTAL

Home Browse Data Access Data Submit Data Metadata Model FAQ

Access Requests Management

Dataset:

User:

Requested date from: 01.01.2000

Requested date until: 31.12.2199

Status: Pending

Dataset	User	Starts	Ends	Requested	Status
EGAD00001008861	Christoph Zwerschke	2023-06-25	2024-06-24	2023-06-25	pending
EGAD00001008862	Philipp Walz	2023-06-23	2024-06-22	2023-06-23	pending

GHGA DATA PORTAL


How to Get Data Access

How to Submit Data

About the GHGA Metadata Model

FAQ

©2023 GHGA. All Rights Reserved. [Imprint](#)



Welcome, Christoph Zwerschke!

We will communicate with you via this email address: c.zwerschke@dkfz-heidelberg.de

You can change this email address in your [LS.Login.profile](#).

Your user session is active.

You have [access to one dataset](#).

Logout



How to Get Data Access



How to Submit Data



About the GHGA
Metadata Model



FAQ



©2023 GHGA. All Rights Reserved.
[Imprint](#)

Perspective

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



Perspective

- **Archive**

- **GHGA Archive will be in production in Q1/2024**
- **Data hubs are already starting to store genomic raw data – but not yet within the legal framework of GHGA**
- **Operational status of archive will then require just a change of metadata – no additional data transfer**

- **Cloud**

- **Secure processing environments (SPE) suitable for person-related data are available through the de.NBI Cloud**
- **Data transfer processes are being established to re-encrypt and stage data**

- **genomDE**

- **Several GHGA Data Hubs are preparing to act as Genome Data Centers within genomDE**
- **Local legal constructs (contract data processing) are being prepared to also process non-consented data**

Thanks to the whole GHGA team!



Partner Institutions

HMGU
CHARITÉ
 UNIVERSITÄT BONN
 UNIVERSITÄT DES SAARLANDES
 EMBL
 dkfz.
 TECHNISCHE UNIVERSITÄT DRESDEN
 German Biobank Node
 DZNE
 NCT
 TUM
 Universitätsklinikum Tübingen
 MDC
 LMU
 ikmb

nfdi Nationale Forschungsdaten Infrastruktur
 Funded by **DFG** Deutsche Forschungsgemeinschaft German Research Foundation
 Gefördert durch: **Bundesministerium für Gesundheit**
 SPONSORED BY THE **Federal Ministry of Education and Research**

Stay informed
ghga.de @GHGA_DE
bit.ly/GHGANewsletter