GHGA

THE GERMAN HUMAN GENOME-PHENOME ARCHIVE

The German resource for efficient genome research. We care about human omics data.

GHGA

THE GERMAN HUMAN GENOME-PHENOME ARCHIVE

Content

GHGA – The German Human Genome-Phenome Archive

We care about human omics data. We make it usable. We protect it.

To leverage the full potential of human omics data, the German Human Genome-Phenome Archive (GHGA) is establishing a common infrastructure to archive, share, and analyse such data in a secure manner. To meet this challenge, GHGA brings together the research and clinical fields. Combining broad expertise, we are building a safe data portal as well as an ethico-legal framework to allow safe sharing of human omics data.

This brochure gives an overview about the world of human omics research and its potential for health care: improved disease prevention, diagnosis, and treatment. Highlighting the importance of data sharing for research, while keeping patients' interests at heart, the following pages showcase GHGA's mission and solutions to achieve this balance. From software development to data analysis, from FAIR principles to multilayered approaches to data security, GHGA will be the German resource for efficient genome research!

Providing insights into how sharing human omics data can revolutionise research and health care, this brochure demonstrates that GHGA is key to making this revolution happen in Germany.

The GHGA Editorial Board,

Jörn Walter
GHGA Co-Spokesperson for Outreach,
Saarland University

Juliane Winkelmann
GHGA Co-Spokesperson for Outreach,
Helmholtz Munich

Oliver Kohlbacher
GHGA Board of Directors,
University of Tübingen

Eva Winkler
GHGA Board of Directors,
NCT Heidelberg

Jan Korbel
GHGA Board of Directors,
EMBL Heidelberg

Oliver Stegle
GHGA Board of Directors,
DKFZ Heidelberg

GHGA

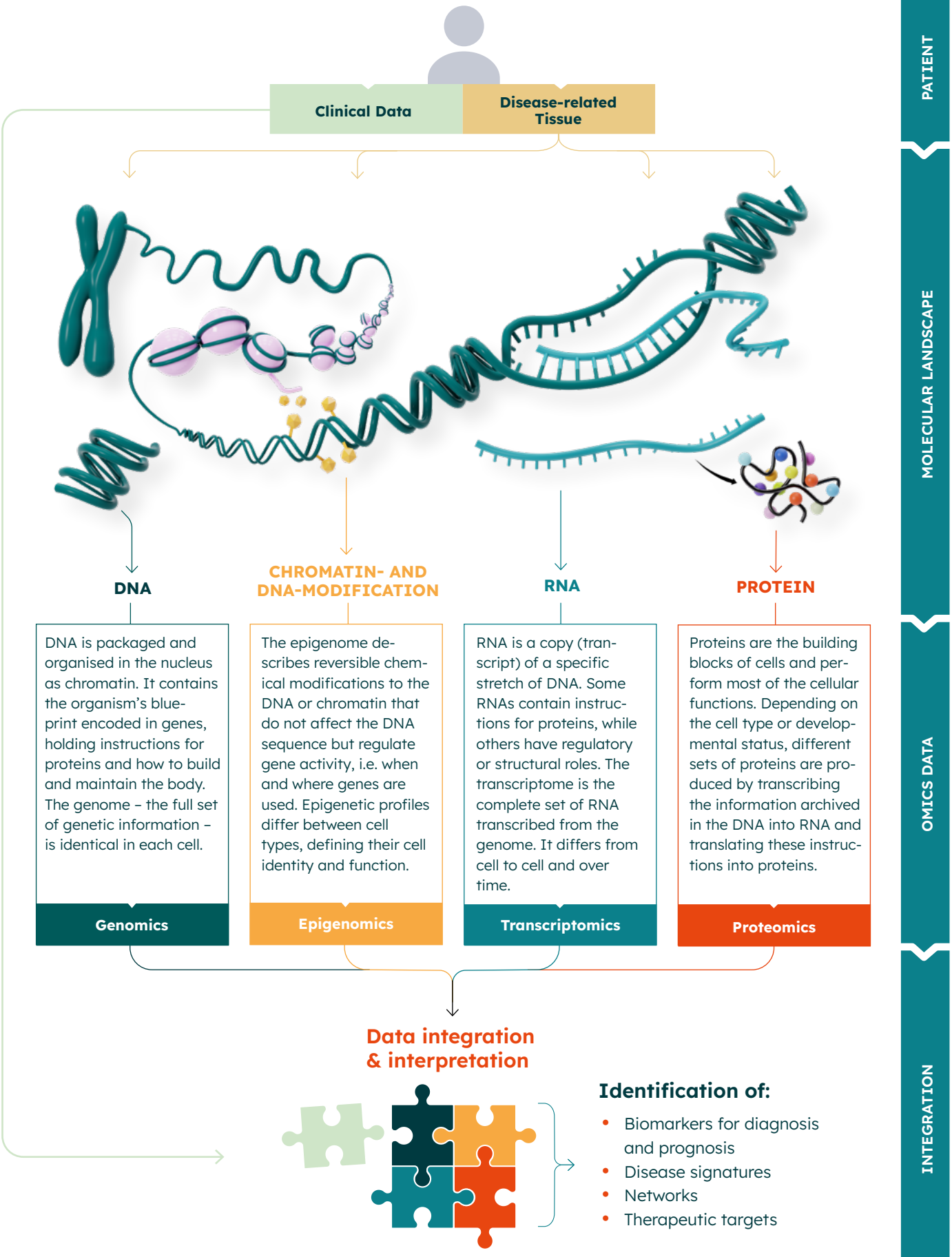
What are Omics data?

PATIENT

MOLECULAR LANDSCAPE

OMICS DATA

INTEGRATION



Omics data: A driver for research and health care

Thanks to advances in sequencing technology, more and more human omics data are being generated in research and medicine. Becoming standard for molecular diagnostics, omics data will revolutionise health care by improving the prediction and prevention, as well as diagnosis and treatment of certain diseases. However, managing and sharing this data is a challenge.

Technological advancement makes genome sequencing affordable

In 2001, the Human Genome Project finished sequencing the first human genome. Since then, rapid technological developments (such as high-throughput sequencing/next generation sequencing (NGS)) have dramatically decreased sequencing costs to around 300 Euro per genome. Advances in bioinformatic analysis methods have solidified the role of sequencing as a standard research and diagnostic technique.

Affordable technology accumulates data

Not only is the number of genomes sequenced per year increasing, but so is the density of information gained as sequencing methods advance. Multi-omics approaches, in which different omics entities are analysed from the same individual or sample, are also on the rise. The rapid growth of available data is a major challenge, but also an unprecedented opportunity for research as it enables a layered understanding of complex interactions.

Revolutionising health care: omics data improves disease prevention, diagnosis and treatment

Omics approaches are becoming an increasingly important tool in health care. They allow the diagnosis of (rare) diseases whose underlying genetic modifications can often only be identified by sequencing. Medically indicated screening for disease risk genes can also assist in disease prevention, allowing closer monitoring of individuals at risk. Individually tailored therapeutic approaches based on genome analysis account for biological variation between patients and have tremendous potential in helping patients not responding to standardised therapy. So from cancer to rare diseases, omics data can greatly contribute to monitoring and improving health conditions, directly impacting patient care.

Genomics is increasingly part of standard care

At present, molecular analysis of genomic data for diagnosis or personalised therapies is largely not part of standard care in Germany. However, this is about to change. In 2021, a new law laid the legal basis for a model project that aims to integrate genomic medicine into routine clinical care starting in 2024.

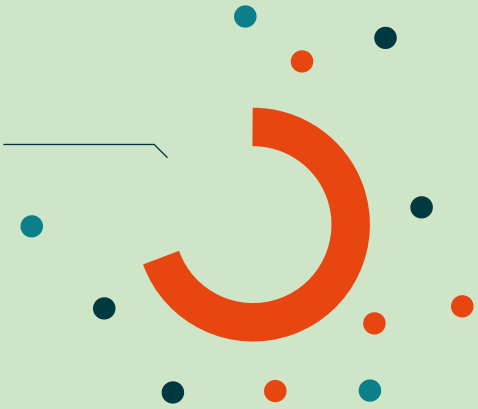
> **Be it in research or clinical care, data is powerful. Only when combining data from larger cohorts will researchers be able to produce robust results and make new scientific discoveries.**

genomDE is a national initiative developing the necessary concepts and infrastructure to advance the use of genomics in routine healthcare. GHGA has been contributing to the development of these concepts and has been preparing for its data hubs to serve as genome data center for health care in the future.

How big is Big Data?

70 %

of cancer cases will be diagnosed with genomics by 2027



40

exabyte of human genome data produced by 2025

=

4 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0

bytes

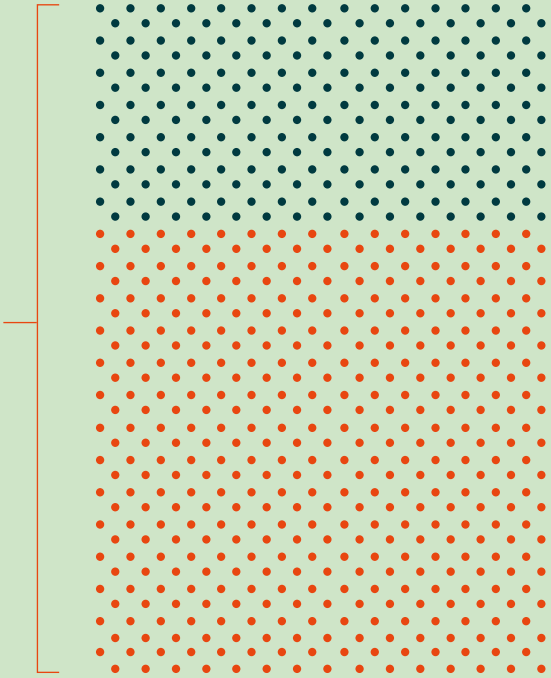
=

1 billion movies

66 years of music

60

million genomes sequenced by 2025 in the G20



20

million from rare diesase

40

million from cancer

Sharing is caring: Improved patient care through data sharing

Data often lie dormant, stored locally with the collecting party, once the initial research has been concluded. It is not available for secondary research as it is not findable or usable. Sharing omics data in a protected and safe manner will unlock the data's full potential: enable scientific discoveries and the development of diagnostic tools and therapies. The following examples showcase the impact of data sharing on research and health care.

Using population data for research

The **German National Cohort** (NAKO Gesundheitsstudie) is a long-term population-based study organised and conducted by a network of German research institutions. The aim is to shed light on the causes of common diseases such as cancer, diabetes, cardiovascular diseases, and infectious diseases, and to uncover the role of e.g. environmental factors, nutrition, lifestyle, and our genes. The comprehensive data collection of more than 200,000 participants is now expanding to omics data. This data will help to elucidate the causes and risk factors of these widespread diseases as well as identify opportunities for early detection and prevention. The omics data will be stored within GHGA, making this large population-based data set also available for secondary research use.

AI assisted health care requires data

Artificial intelligence (AI) is already widely used in health care through an increasing number of applications: robot assisted surgery, apps to help identify skin cancers, MRI and X-ray image analysis, wearables for diabetes patients, etc. Large data sets are needed to train the AI algorithms behind the applications. The continuous expansion of AI assisted health care has the potential to save billions of Euro per year, worldwide. GHGA will provide secure access to large, homogenised data sets to enable and promote AI-based big data analysis for research and clinical care purposes.

Sharing data for precision cancer medicine

Comprehensive genomic and transcriptomic analyses can enable personalised medicine and improve patient care. The ongoing multicenter observational NCT/DKFZ/DKTK **MASTER (Molecularly Aided Stratification for Tumour Eradication Research)** study demonstrates that molecular analysis provides diagnostic and therapeutic benefits for cancer patients. Using a standardised precision oncology workflow, the patient's tumour and control genome are analysed and the results are discussed in a multidisciplinary molecular tumour board. Targeted therapy recommendations were given in more than 85% of cases, of which 32% were managed accordingly. With a significantly increased progression-free survival compared to previous treatments in more than a third of these patients, the MASTER programme is a clinical use case for genomic medicine, where data sharing is key to accumulate large genomic and clinical data sets. The MASTER data set was the first available within GHGA.

> **Recognising the potential of data sharing, GHGA provides a strategy for secure archiving and effective access to human omics data to enable innovative research and programmes for health care.**

At GHGA, we work together to allow omics data to take centre stage and effectively inform research and health care.

Data sharing during crisis: the corona pandemic

Genome research had a major impact on the Corona pandemic. From decoding the SARS-CoV-2 genome and developing a vaccine in record-breaking time, to monitoring emerging variants to inform public health policy and predicting individual risk factors by sequencing human host genomes – research shaped the course of the pandemic.

Data sharing enables and accelerates research discovery. In crisis situations, the need to share research data quickly, yet securely, is most pressing. Establishing infrastructure to facilitate secure and FAIR data access for legitimate research questions will provide Germany with a solid foundation for the future. During the corona pandemic, GHGA was part of initiatives such as DeCOI and CoGDat, and established a portal to collect and share SARS-CoV-2 raw sequencing data, enabling genetic surveillance of virus variants across Germany.

Establishing a reference genome collection for rare diseases

Rare diseases are often caused by variants in the genetic code. Only when large genome reference data sets are created and shared across clinics can patients be successfully diagnosed. The rarer the disease, the larger the data set needs to be in order to find the genetic change. A swift and accurate diagnosis and treatment of the disease is immensely beneficial for the individual patients – but it also eases the burden on the healthcare system. Using sequencing techniques early in the diagnostic process can more than triple the diagnostic success rate, at one-third of the cost per diagnosis. Working closely together with the rare disease community and initiatives such as Solve-RD, GHGA aims to establish a reference genome collection and provide workflows tailored to their community needs.

GHGA: The national infrastructure for secure data sharing

Creating a safe home for human omics data, GHGA will, for the first time, make large data sets from all over Germany accessible in one infrastructure. As a national initiative, we can offer researchers a unique ethico-legal framework specific for Germany.

Igniting scientific discoveries

Decentralised collection and storage of human omics data makes it difficult for researchers to find data they need or to safely share data they generate. GHGA will help German institutions to bring these data together and make them easier to find – enabling researchers to create new cohorts to validate their findings and provide opportunities for new and novel big data approaches. This will ignite further biological discoveries and help to translate research findings into the clinical routine, benefiting patients and adding further value to genomic medicine.

From infrastructure development to the ethical-legal framework for data sharing, our interdisciplinary team works together to find optimal solutions.

Unleashing the full potential of data

Human omics data contains a depth of information which can help answer a variety of research questions – far beyond the initial purpose it was collected for. Making such data available to other researchers and clinicians allows new research hypotheses to be asked without the need to collect new data, thus saving resources such as time and funds.

Enabling sustainable research

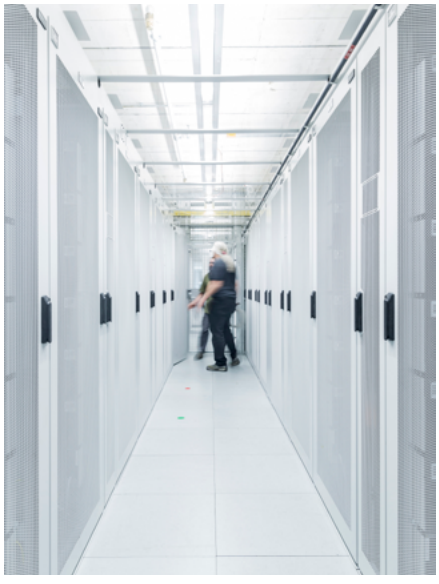
Establishing GHGA as a long-term archive guarantees the sustainable use of data via shared standards and infrastructure used for all human omics data. Storing data in a single infrastructure will reduce the need for duplicate copies of data, allowing efficient use of storage space, saving costs and energy, and protecting the environment.

Bridging research and clinical care

Sharing omics data across research and clinical care will establish a loop where research informs health care decisions – from diagnosis and therapy to policy – and clinical data feeds back into the research progress. This way, patients will benefit from research based on richer data sets.

Making German research internationally visible

As a national node of the federated EGA (European Genome-phenome Archive) and the GDI (European Genome Data Infrastructure), GHGA will follow national data protection regulations and at the same time be closely linked to international data infrastructures – enabling German researchers to shape future international standards for data exchange and take on leading roles in international research consortia.



Easy data sharing, but with controlled access

Researchers, clinicians and institutions who have collected omics data will be able to use GHGA's streamlined data deposition services and infrastructure to safely share data with scientists they have approved. This will make the process of data sharing more efficient, and enable researchers to begin their research more quickly.

Valuing patients' rights and views

To address ethico-legal challenges in the context of omics data sharing, GHGA develops guidelines and consent modules to inform patients as well as to simplify the consent process for clinicians and researchers. Offering transparent insights into GHGA governance, an open dialogue with patients will allow active participation of all stakeholders.

Protecting sensitive omics data

Understanding the vulnerability of human omics data, GHGA takes multi-layered steps to protect sensitive data. We are building a secure software infrastructure aligned with an ethico-legal framework that not only guarantees GDPR-compliance, but also a development in exchange with and focused on patients' needs.

Triggering a mindshift regarding genome research

Transparent communication around topics relevant to GHGA – such as data sharing and data protection – will further fuel the open science movement among researchers and clinicians. In addition, actively engaging the public will highlight the benefit of genome medicine for each and everyone's health.

A strong network across Germany and beyond

GHGA operates as an interconnected network across Germany. Serving as a German node in European initiatives, GHGA shapes future international standards for data sharing and contributes to a sustainable data infrastructure worldwide.

Federated network for central access

The GHGA Data Portal will be the single point of contact for the up- and download and analysis of omics data. Behind the scenes, this 'central' face of GHGA is serviced by data hubs operating as a federated network. The data hubs are located at GHGA partner institutions which are leading institutions in genomic medicine and major omics data producers (such as NGS-CN).

We are building upon existing infrastructure, partnering with local, regional, and national high performance computing centres, as well as centres that operate existing cloud infrastructures (e.g., de.NBI/ELIXIR-DE) and have the capacity and experience to sustainably operate a robust and scalable infrastructure.

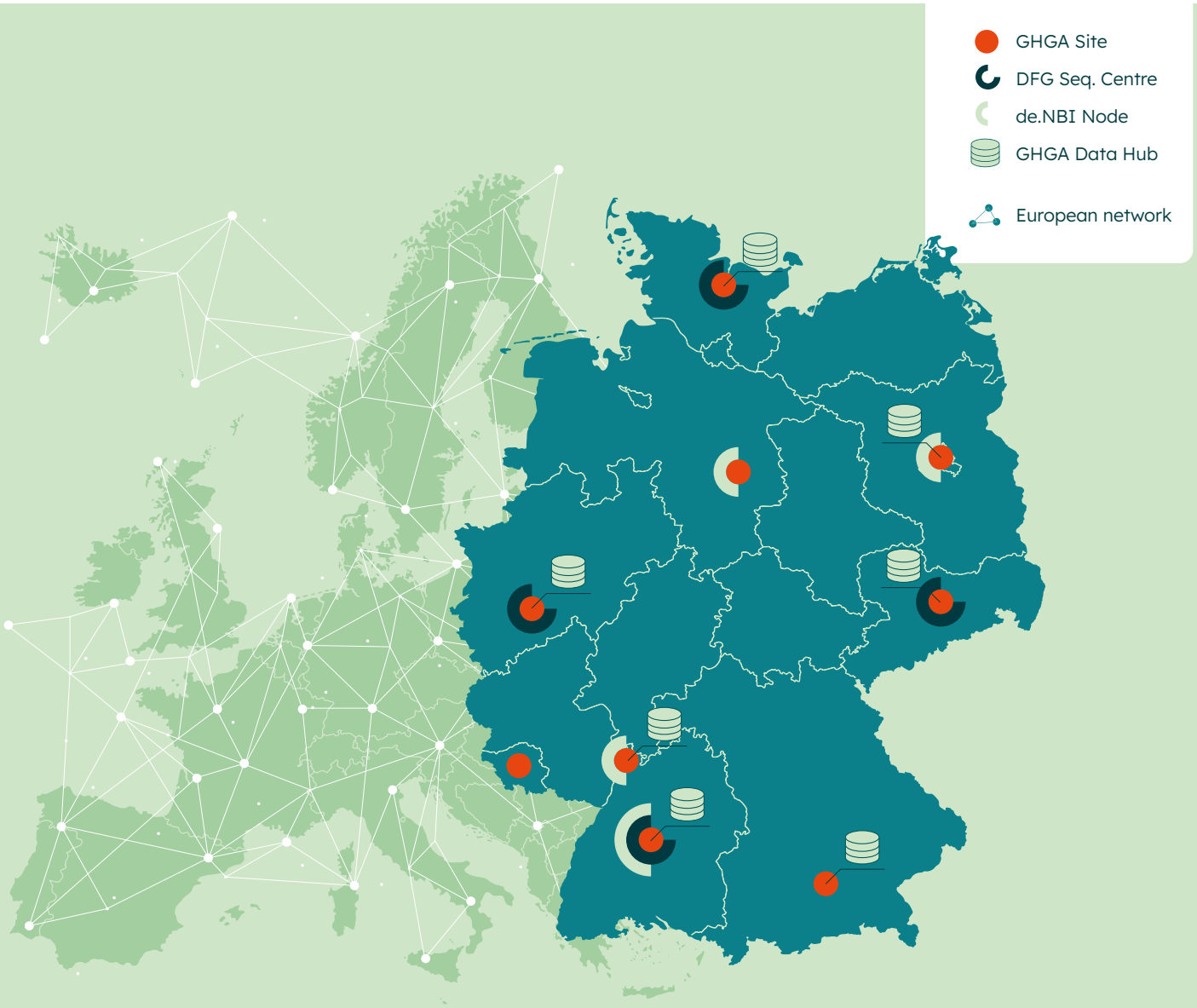
Bundling expertise in one initiative

Within GHGA, an interdisciplinary team of software developers and IT operations specialists, ethics and legal experts, bioinformaticians and computational biologists, along with clinicians, works together on establishing a secure infrastructure for omics data. The team is backed by renowned senior researchers from 21 universities, Helmholtz Centers and research institutions across Germany who contribute their expertise to the project.

GHGA is funded as part of the **National Research Data Infrastructure** (NFDI) via the DFG (Deutsche Forschungsgemeinschaft). Within the NFDI, valuable data from science and research are systematically accessed, cross-linked and made available in a sustainable and qualitative manner.

The **Next Generation Sequencing Competence Network** (NGS-CN) is a DFG-funded initiative comprising four highly specialised centres that provide the infrastructure and expertise for high-throughput sequencing and make omics research widely accessible to the life sciences and medicine.

Learn more about how we work



A strong partner for Europe

The current home of personally identifiable genetic and phenotypic data collected within biomedical research projects across Europe is the **European Genome-phenome Archive** (EGA). In the federated EGA, a network of national nodes will share and commit to joint standards and infrastructures to enable data sharing across Europe. GHGA is the German federated EGA node.

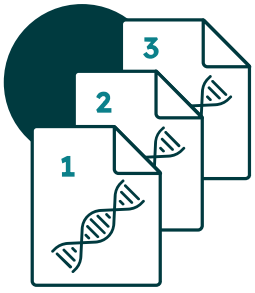
The European **1+ Million Genomes** (1+MG) initiative is aiming to collect data from more than one million human genomes. It will grant secure access to the data and corresponding clinical data across Europe for better research, personalised health care and health policy making. To see the 1+MG vision to fruition, the implementation of a European infrastructure was funded via the **Genomic Data Infrastructure** (GDI) project. Building upon a network of national infrastructures, technical standards will be aligned and a common ethico-legal framework defined – while abiding to national requirements. GHGA forms the German GDI node, connecting the genomic data stored in GHGA to the pan-European GDI infrastructure and integrating our expertise into the European network.

On a global scale, GHGA is engaging with the **Global Alliance for Genomic Health** (GA4GH) to implement and contribute to the international framework of standards and harmonised approaches for effective and responsible genomic and health-related data sharing.

Portfolio

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 procedures
- Unified ethico-legal framework
- Harmonised metadata

GHGA Atlas

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

GHGA Cloud

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

Learn more about the current status of our portfolio

Sharing data – FAIRly

We cannot share data without describing it. Metadata provides information about the properties of a given data set – the 'what', 'when', 'where', 'who', 'how' and 'why'. Without rich metadata describing omics data sets, these cannot be found, used, or interpreted. Metadata includes information on protocols or instruments used to generate the data, as well as information about sample specifications such as the cell or tissue type or disease status, and may also include information on the data donor (e.g. their age or biological sex).

The GHGA Metadata Model was developed to harmonise the way data is described in different settings – thereby making data sharing easier. Utilising established and widely used ontologies and vocabularies helps data producers to describe their submitted data as well as to retrieve data of interest in a FAIR manner.

The first phase of GHGA – the GHGA Metadata Catalog – is a public frontend for the discovery of human omics study data from German research institutions. Allowing the search for non-personal metadata, it aims to create a resource that collects information on human omics data sets available from German institutions for secondary research under controlled access conditions.

Learn more
about the GHGA Metadata Model

What is FAIR?

The FAIR data principles are a guide to enhance the secondary use of data. Data should be findable, accessible, interoperable, and reusable.

Findable

(Meta)data should be easy to find by humans and computational systems through the use of unique identifiers and comprehensive data descriptions.

Accessible

Clearly defined procedures for secure and transparent access must be in place so that the user knows how to access the data.

Interoperable

Data can be exchanged across different workflows or systems and integrated with other similar data by relying on (metadata) standards and ontologies.

Reusable

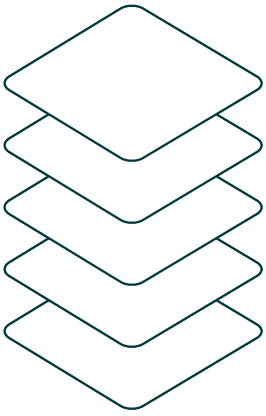
Well described (meta)data enables optimal secondary use.

A multilayered approach to data protection: Keeping patient data safe

GHGA takes a multilayered approach to data security. We build advanced infrastructure as a basis for safe data archiving and sharing. A framework for GDPR-compliant data processing and controlled, yet FAIR, data access adds another layer to ensure that data is protected and at the same time fulfils its potential to advance research.

0,1 %

Like a fingerprint, our DNA sequence is unique. Only 0.1% of genetic information differs from one person to another.



Cybersecurity through advanced infrastructure

GDPR-compliant data processing

Managing consent: a guide for researchers and clinicians

Controlled, yet FAIR, data access

Beyond legal – ethical considerations

Why do omics data need special protection?

The DNA carries information about disease likelihood, sex, ethnicity, or other sensitive information. Hence, special care must be taken when handling human omics data. Data collected for research purposes are typically pseudonymised, however, individual gene sequences could potentially still lead to re-identification. Omics data is usually not publicly available, but archived securely and only made available to approved researchers under specific conditions.

Patients want to support research with clinical data

Omics data is sensitive and requires special protection. Patients know that. Yet they are willing to donate their data to science – hoping to help future patients with new developments and treatment options. A study involving cancer patients found that 97 percent are generally willing to make their clinical data available for biomedical research purposes. The major condition for their consent? Maximum data security. A goal GHGA strives towards.

Safe data storage at the GHGA Data Hubs with a zero trust model.

Cybersecurity through advanced infrastructure

GHGA enables highly sensitive omics data to be stored and analysed in a unified, privacy-compliant framework. Here, the security of the infrastructure and the data is of crucial importance. In our private cloud environment, all data is physically located at our data hubs. This allows us to combine the advantages of modern cloud computing with the data security and control of on-premises IT infrastructures. We thereby ensure that no resources are shared with other users and that all compliance requirements are met. In addition, the zero trust model is implemented at all levels of the application: every user, inside or outside the network, requires strict verification of their identity before they are granted access to authorised data.

GDPR compliant data processing

Sensitive personal data is protected under the General Data Protection Regulation (GDPR). The interpretation of the GDPR is dependent on the country the data is handled in. As a German initiative, GHGA addresses the legal basis for data processing and consent in the national context. To further protect the sensitive data, our experts work on risk assessments, de-identification and anonymisation methods, and a code of conduct for data sharers.

Beyond legal – See how patients shape GHGA on page 24.

Data stewards will assist data submitters with the up- and download of their data as well as with questions around controlled data access.

> **GHGA works on the legal interoperability for data processing within the EU and in international data spaces. As part of EU-initiatives like FAIR Data-Spaces and GA4GH, our experts shape international standards. Data sharing across national borders will empower international collaboration, make German research visible and improve the quality of science, ensuring greater benefits for the people it serves.**

Managing consent: a guide for researchers and clinicians

Typically, informed consent from patients and research participants is required in order to share omics and related health data for research. To provide guidance for clinicians, researchers and institutions wanting to share data via GHGA, we have developed modules that can be integrated into consent forms. These updates inform patients and research participants about the possibility of sharing their omics data with genome archives such as GHGA.

In an effort to make already existing data available for research purposes, we developed an app that supports the assessment of the legal validity of sharing data with any given previous (legacy) consent form.

Learn more
about the GHGA Consent Tools

Controlled data access

Access to data stored in GHGA is restricted. Only non-personal metadata is publicly available within the GHGA Data Portal. Researchers intending to use any of the archived data or view personal metadata must apply for access to the responsible legal person. A data access committee or a comparable instance will review the legitimacy of the request before granting access. This step ensures that only researchers with a valid research purpose gain access to sensitive data – adding another layer of protection.

Researchers or institutions submitting data to GHGA remain the controllers of the data, and it is their decision who is granted access. GHGA serves as a mediator in this process. Dedicated data stewards at our data hubs, trained in technical and ethico-legal aspects of managing omics data, will assist users in submitting data, offer guidance on how to manage data access requests, and enable secure access via encrypted downloads.

Beyond storage: facilitating data analysis and interpretation

GHGA's mission goes beyond data archiving. To promote standardised data analysis of the data sets stored within GHGA, we aim to provide best practice workflows and develop global standards for both data sharing and analysis.

Balancing de-centralised data storage for GHGA

Developing software that is deployable across several data hubs, while flexible enough to adapt to the local resources, requires agile development processes and deployment practices. This makes us think of software development and operation as one unit. Choosing progressive yet robust architecture patterns, we implement a domain-driven microservice architecture. To be independent of a specific IT infrastructure and to enable frictionless continuous deployment, we rely on the container orchestration solution Kubernetes and its associated ecosystem.

Developing standards and best practices

Putting great effort into aligning with national and international software standards, we actively push their development forward by taking part in community efforts of the NFDI, ELIXIR Europe, nf-core and GA4GH. For example, concepts for data encryption (Crypt4GH compatible), federated identity and permission management (GA4GH Visa and Passport compatible), and workflow execution services (WES, GA4GH compatible and implemented through nextflow) have been developed in accordance with GA4GH standards.

Enabling collaborative and reproducible data analysis

Translating raw omics data into meaningful insight requires bioinformatic workflows to analyse the data. Standardised analysis workflows allow research and diagnostics to be more reproducible and efficient while providing best practice examples. GHGA aims to provide robust, accurate and reproducible workflows for various tasks, e.g. for the identification of critical variations within a given genome sequence. Here we are using, improving, and continuously benchmarking existing workflows and aligning with standards defined by communities such as GA4GH, nf-core and NGS-CN.

Cross-study comparisons and the joint analysis of multiple cohorts can only be done with unified processing of the data sets. GHGA workflows are being developed together with the respective communities and FAIRly published in an open-source manner (e.g. nf-core). They can be run independently and locally by the users. In later stages of GHGA, workflows can be used in a trusted research environment provided by the data hubs avoiding data transfer. No matter where the workflows are executed, the results will be reproducible and comparable.

GHGA does not stand alone – we collaborate with global initiatives to set community standards for sharing of human omics data.

A FAIR platform with metadata management tools.

A secure and federated data storage for large omics files.

An identity and permission management for submitters and access requests.

Compliance with and contribution to community tools and standards.

Highlighting the power of data

GHGA's communication channels are diverse. Reaching different audiences with the same message: data sharing in omics research is safe, if all necessary safety precautions are taken, and important to drive scientific discovery.

Building infrastructure with the research community in mind

We are constantly interacting with omics data producers and users. Only when we know what the communities need can we deliver a platform that serves all of their requirements.

Training future scientists

GHGA aims to educate students and researchers about the challenges and opportunities surrounding omics research and data sharing principles through lectures, webinars, and hands-on training activities. The training portfolio ranges from topics like meta-data and FAIR, to ELSI topics such as consent and data protection issues, as well as analytical themes (e.g. bioinformatic analysis or statistics) – tailored to a diverse set of GHGA communities.

Learn more
about our training opportunities

Making omics research visible

GHGA is passionate about engaging with the public to increase understanding and awareness around omics research and ultimately enabling informed decisions on data sharing. Genome research can be life saving, is interesting and all around us. Making this visible for everyone, we are seeking a dialogue with the public by exploring different approaches, such as local events including science slams and science pop up stores, or our podcast 'Der Code des Lebens'!

Listen to
our podcast

Valuing patients' perspectives

Patient engagement not only ensures accountability, but also improves research outcomes by shifting the focus of research to areas that patients need, ultimately improving patient care. By involving patients in governance, we not only hope to enhance accountability and transparency, but also allow patients to take an active role in decisions and discussions about their data.

We are committed to addressing patients' needs and expectations through active and continuous patient participation.

Dialogue with patients

GHGA strives to understand the expectations and concerns of patients. To incorporate the valuable participation of patient experts, we have involved representatives from the rare disease and cancer patient communities early on and held deliberative forums to elicit their views on transparent and trustworthy governance. Through a continued exchange of ideas and perspectives on issues concerning omics data, we hope to make patients, not just their data, an integral part of GHGA's growth and development.

Making patients' voices heard

By collaborating with patients and their representatives and making their participation in governance a priority, GHGA is taking steps to ensure that the patient voice is not only heard, but that it also has a meaningful impact. Establishing a Patient Advisory Board gives patients a direct voice in decision making processes. Highlighting patients' views in GHGA outreach activities allows matters close to patients' hearts to take the spotlight.

The future of omics in health care: Experts' opinions

The question of how to get the precision treatments into the precise group of people who can be expected to benefit is one of the big challenges of oncology today.

The main drivers of drug resistance in cancer therapy are molecular heterogeneity and molecular evolution of the tumour, forced by Darwin´s natural selection mechanisms. Our current molecular diagnostic standards (majority single gene, or gene panels, collected on very small tissue samples, usually years before analysis) are more comparable to historical road maps that do not show current actual conditions, such as construction sites, new roads, or traffic jams. But, due to the dramatic superiority of Google Maps, classic road maps have long been obsolete. For a next generation Google Maps-like patient navigation, we need to combine deep omics data with big longitudinal real-world data. GHGA has the potential to provide the infrastructure for the development of such advanced swarm intelligence based adaptive algorithms, by uniting so far heterogeneously collected and analysed omics data with associated longitudinal clinical data on a national level.

We need a "Google Maps approach" for cancer therapy decision making! GHGA establishes the infrastructure to combine new omics approaches with longitudinal clinical data towards dynamic swarm intelligence based algorithms for advanced patient navigation.

Prof. Dr. Thorsten Schlomm
Professor of Urology and Chairman of the Department of Urology at Charité - University Hospital of Berlin; Founder of DNA-Med; GHGA Co-Spokesperson Outreach

Prof. Dr. Dr. Eva Winkler
Heisenberg Professor and Head of the Section for Translational Medical Ethics at the University Hospital Heidelberg and Executive Director at the National Center for Tumor Diseases (NCT), Heidelberg; GHGA Board of Directors and Co-Spokesperson for ELSI

“GHGA aims to enable data-driven research in the interest of patients and the public. For us, patient engagement is not a secondary process but an integral part of our mission. We therefore work with patient advisors and patient organisations to ensure that diverse communities and perspectives are taken into account from the early development stages through to the full functionality of GHGA. Meaningful patient involvement in research will help shape a future in which the health care system takes the needs and concerns of those seriously who must rely on it.

Prof. Dr. Juliane Winkelmann
Chair of Human Genetics at the Technical University of Munich and Director of the Institute of Neurogenomics at Helmholtz Munich; GHGA Co-Spokesperson for Outreach and Co-Coordinator of the GHGA Data Hub in Munich

“

We can now diagnose more than half of the patients with rare diseases using exome and genome sequencing. Molecular diagnosis is important to assess the prognosis, avoid unnecessary diagnostic tests and also for further family planning. We can even offer personalised therapies in some cases.

But how can we help those patients, whom molecular diagnosis can’t help yet? Using knowledge generating patient care with systemic multi-omics analyses in the context of research projects can help develop new diagnoses for patients. To enable and simplify data sharing between researchers and facilitate such analyses, we now have GHGA.

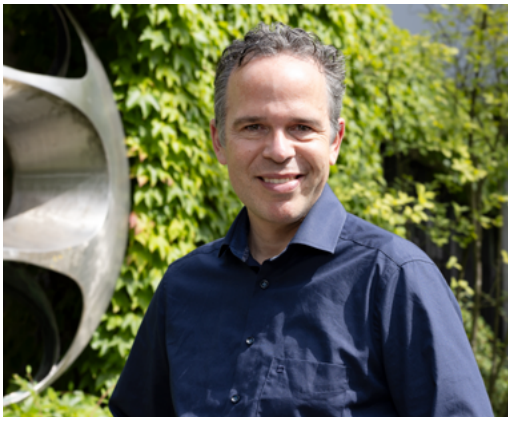
“ Ten years from now, we are hopefully not talking about a research infrastructure anymore, but about a genomic data infrastructure that serves both research and clinical care in a seamlessly integrated way!

GHGA is anchored in research – funded by the BMBF and the federal states and as part of the NFDI. The interesting and exciting question is how do these research efforts interact, integrate and connect to clinical care?

Thinking about the future, we see increasing convergence between research and health care. But how can we create an even more active exchange? How do we get more research ideas into the healthcare system and how do we get more ideas about fostering translation into the research domain?

We are convinced that GHGA can make an important contribution to achieve these goals and become a piece in the puzzle to connect these two worlds, patient treatment and research, even more closely.

Prof. Dr. Oliver Stegle
Head of the Computational Genomics and Systems Genetics Division at the German Cancer Research Center (DKFZ) and leader of the Statistical Genomics and Systems Genetics group at European Molecular Biology Laboratory (EMBL) in Heidelberg; GHGA Board of Directors



Prof. Dr. Oliver Kohlbacher
Chair for Applied Bioinformatics at the Department of Computer Science at Eberhard Karls University Tübingen, Director of the Institute for Bioinformatics and Medical Informatics, and Director of the Institute for Translational Bioinformatics at the University Hospital Tübingen; GHGA Board of Directors

Prof. Dr. Jan Korbel
Head of Data Science at European Molecular Biology Laboratory (EMBL) in Heidelberg and leader of the Bridging Division Mechanisms of Genomic Variation and Data Science group at the German Cancer Research Center (DKFZ); GHGA Board of Directors

Imprint

Editor-in-chief: Prof. Dr. Oliver Stegle
Coordination: Dr. Ulrike Träger

GHGA Office (W620) within the
German Cancer Research Center (Deutsches Krebsforschungszentrum (DKFZ))
Im Neuenheimer Feld 280
D-69120 Heidelberg, Germany
Phone: +49 6221 423601

Contributors:
Dr. Ulrike Träger, DKFZ; Dr. Nina Gasparoni, Saarland University; Dr. Nicole Schatlowski, University of Tübingen; Dr. Jan Eufinger, DKFZ and the GHGA Team

Editorial Board:
Prof. Dr. Jörn Walter, Saarland University; Prof. Dr. Dr. Eva Winkler, National Center for Tumor Diseases (NCT); Prof. Dr. Juliane Winkelmann, Helmholtz Munich; Prof. Dr. Jan Korbel, European Molecular Biology Laboratory (EMBL); Prof. Dr. Oliver Kohlbacher, University of Tübingen

Contact:
contact@ghga.de
www.ghga.de
@ghga_de
linkedin.com/company/the-german-human-genome-phenome-archive

Date: Sep 2023

Photo credits:
DIE KAVALLERIE, U.Träger/DKFZ, M. Stark/DKFZ, J. Jäger/Universität Tübingen, A. Eckert/TUM, S. Ingham/EMBL, Charité

Design and Layout:
DIE KAVALLERIE

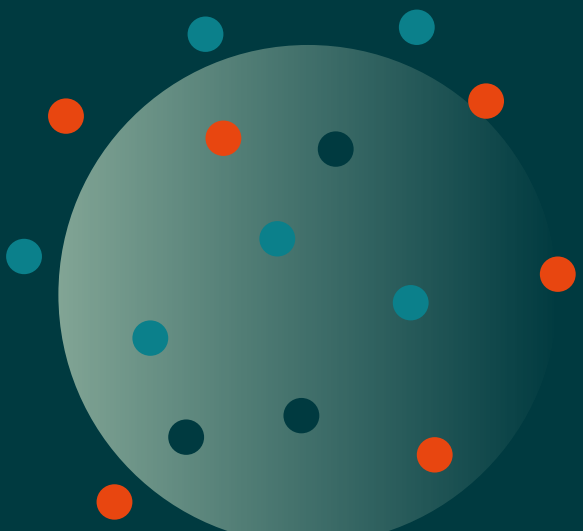
Funding:
The GHGA project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the National Research Data Infrastructure Germany (NFDI, Grant Number 441914366 (NFDI 1/1)).

In cooperation with

nfdi

DFG Deutsche Forschungsgemeinschaft

GHGA



ghga.de