



# Terms of Use GHGA Data Infrastructure

Authors: The GHGA Consortium\*

Version: 1.0 – April 10<sup>th</sup>, 2024

\*GHGA – The German Human Genome-Phenome Archive – is building a secure national omics data infrastructure, enabling the use of human genome data for research purposes while preventing data misuse ([www.ghga.de](http://www.ghga.de)). GHGA is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Grant Number 441914366 (NFDI 1/1). Questions and feedback should be directed to: [contact@ghga.de](mailto:contact@ghga.de)

## Contents

<i>§1 - Definitions</i> .....	1
<i>§2 - Introduction</i> .....	2
<i>§3 - Purpose</i> .....	2
<i>§4 - Duration</i> .....	2
<i>§5 - Services</i> .....	2
[1] Secure storage of, and archival services for, omics data .....	2
[2] Maintenance of a metadata catalogue .....	4
[3] Support for Data Controllers to share omics data to approved researchers .....	4
[4] Interoperability with the Federated European Genome-phenome Archive (EGA) .....	5
[5] GHGA Helpdesk Service .....	5
<i>§6 - Complaints</i> .....	6
<i>Annex 1 - Information to be provided before Submission</i> .....	7

## §1 - Definitions

**GHGA** – The GHGA Project (GHGA) is a consortium of institutions across Germany that is working together in order to develop infrastructure to securely store and support disseminate omics research data. GHGA serves as the national node for Germany as part of the federated European Genome-phenome Archive (EGA).

**GHGA Central** - Deutsches Krebsforschungszentrum (DKFZ) is the leading organisation for GHGA and is responsible for signing legal agreements with external parties on behalf of GHGA.

**GHGA Data Hub** - An organisation that is part of GHGA that provides secure data storage, computational infrastructure and user support. The following GHGA partners are operating GHGA Data Hubs: Deutsches Krebsforschungszentrum (DKFZ) in addition to its role as GHGA Central, Charité – Universitätsmedizin Berlin (Charité), Eberhard-Karls-Universität Tübingen (EKUT), Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft (MDC), Technische Universität München (TUM), Technische Universität Dresden (TUD), Universität zu Köln (UzK), Universitätsklinikum Schleswig-Holstein (UKSH).

**GHGA Operations Consortium** - The GHGA Operations Consortium is made up of the organisations in GHGA Central and the GHGA Data Hubs that have been approved and contractually onboarded. Collectively they work to support Data Controllers who wish to share human omics data, and researchers wishing to access data. Not all institutions which contribute to GHGA are members of the GHGA Operations Consortium.

**GHGA Data Stewards** - GHGA Data Stewards provide support to users of the GHGA Data Infrastructure. This includes responding to queries sent to the GHGA Helpdesk, checking Non-personal Metadata, and releasing data upon instruction from Data Controllers. GHGA Data Stewards are based at the GHGA Data Hub

**GHGA Data Infrastructure** - The infrastructure operated by the GHGA Operations Consortium where omics data and personal metadata is stored on behalf of Data Controllers.

**Data Controller** - The person or organisation that is legally responsible for defining the essential means of processing for the omics data deposited in the GHGA Data Infrastructure.

**Data Requester** - A person who makes a request to access the omics data stored in the GHGA Data Infrastructure.

**Data Submitter** - The person or organisation that deposits data in the GHGA Data Infrastructure. In many cases, but not exclusively, they will also be the Data Controller.

**Service User** - Someone who interacts with the GHGA Operations Consortium for any purpose.

**Administrative Data** - Data which are generated through the operation of GHGA Data Infrastructure. For example, the names and contact details of Data Submitters or people who contact GHGA.

**Metadata** - Information that describes Research Data. Non-personal Metadata is included in the GHGA Metadata Catalog to help people find suitable Research Data. Personal Metadata may include detailed health or demographic information that helps Data Requester understand the Research Data they have been approved to access.

**Research Data** - Omics or other forms of genetic and health data that are archived for scientific research purposes.

## §2 - Introduction

The German Human Genome Phenome Archive (GHGA) is a consortium that has created a national data infrastructure to securely store, and facilitate the sharing of, human omics data in Germany. The GHGA Data Infrastructure is run cooperatively by the GHGA Operations Consortium. In order to support the research community, a number of services are offered by the GHGA Operations Consortium, and these are described in §5 - Services.

These Terms of Use are a living document as the services offered by the GHGA Operations Consortium will evolve over time. They cover the services offered as part of the first phase of the GHGA Data Infrastructure, which is focused on archiving and helping Data Controllers to share omics data. Future phases, through which the GHGA Operations Consortium will offer additional services, are planned.

## §3 - Purpose

These Terms of Use will define the services offered by the GHGA Operations Consortium and the performance levels of those services that Service Users can expect. It will also define the conditions of the GHGA Operations Consortium from Service Users when using the GHGA Data Infrastructure.

## §4 - Duration

These Terms of Use, in their current version, are in place for as long as the Service User is using the services offered.

They may be regularly updated by the GHGA Operations Consortium. The most recent version is available via the GHGA website<sup>1</sup>.

All services offered are offered to Service Users indefinitely or for as long as the GHGA Operations Consortium is able to support them.

## §5 - Services

*[1] Secure storage of, and archival services for, omics data*

---

<sup>1</sup> [www.ghga.de/terms-of-use](http://www.ghga.de/terms-of-use)

**Services offered:**

- Institutions and individuals can deposit data in the GHGA Data Infrastructure as Data Controllers or Data Submitters. The suitability of the data for deposit is assessed by the GHGA Operations Consortium according to criteria outlined in **Annex 1; the request to deposit data can be refused**. If approved, the GHGA Operations Consortium offers permanent archival services for Research Data and Personal Metadata.
- Data Controllers remain in control of their data, with a Data Processing Contract setting out how GHGA Central, and the GHGA Data Hubs as sub-processors, may process submissions. A Data Controller may terminate their Data Processing Contract at any time and remove their data from the GHGA Data Infrastructure. The most recent version of the Data Processing Contract can be found on the GHGA website<sup>2</sup>.
- The data shall be stored in the GHGA Data Infrastructure without charge.
- The GHGA Operations Consortium uses Life Science Login AAI and/or other identity management technologies to ensure that only authorised persons can deposit and access data.
- The GHGA Operations Consortium will help to maintain the deposited data to ensure its integrity and availability, for example restoring backed up versions of the data in the event of corruption.
- Data Controllers will receive Persistent Identifiers for the Research Data they submit that can be used in publications.

**Users' Obligations:**

- Data Controllers shall deposit Research Data and Personal Metadata with the intention of making it available for scientific research use according to the applicable law for data sharing.
- Data Controllers must notify GHGA Central of any changes that may affect the ability of the GHGA Operations Consortium to communicate with them during the period of time that their data is archived within the GHGA Data Infrastructure. The Data Controller is required to notify GHGA Central if persons who can act on their behalf change.
- Data Submitters must submit data in a suitable format in accordance with current omics data standards, and via methods approved by GHGA. Information on standards suitable for GHGA data deposition are described in GHGA's user documentation<sup>3</sup>.

---

<sup>2</sup> <https://www.ghga.de/data-portal/data-protection>

<sup>3</sup> <https://docs.ghga.de/>

- The Data Controller should assess the data quality prior to submission and will remedy errors found in the data both ante and post deposit. Support via the GHGA Helpdesk is limited to data submissions that satisfy these minimum requirements. The role of GHGA Data Stewards in this process is described in *GHGA Validation and Quality Check Service* below.
- Data Controllers must have a legal basis for sharing data via GHGA. The legal basis must permit the transferring of data to GHGA and should permit sharing of the data with future Data Requesters.

## *[2] Maintenance of a metadata catalogue*

### **Services offered:**

- GHGA uses and maintains a [metadata model](#) that supports the FAIR principle of ‘findability’ for the deposited data. GHGA will continue to develop the metadata model over time to ensure it meets future standards.
- The GHGA Operations Consortium operates a searchable, online, catalog of public Non-personal Metadata for the data that has been deposited. According to the GHGA Data Protection Framework<sup>4</sup>, Non-personal Metadata cannot include items that would be considered personal data according to the GDPR and applicable data protection law.
- The GHGA Operations Consortium may, when instructed by the Data Controller, update the Non-personal Metadata when necessary. For example, if the underlying ontology has been updated and included terms have changed.

### **Users’ Obligations:**

- Data Submitters are to provide the required Non-personal Metadata information, via the GHGA Metadata Model, when depositing data. The information supplied must fulfil all required fields and be accurate.

## *[3] Support for Data Controllers to share omics data to approved researchers*

### **Services offered:**

- Support for Data Controllers who wish to make their research data available safely by providing encrypted storage and access to the data to approved researchers with programmatic tools.
- Support for Data Requesters in contacting the responsible DACs.
- A secure download service linked to an AAI service that is bespoke to specific Data Requesters.

---

<sup>4</sup> <https://www.ghga.de/data-portal/data-protection>

**Users' Obligations:**

- Data Controllers, and their nominated DACs, shall respond to requests to access Research Data and Personal Metadata promptly and shall communicate with Data Requesters in a timely manner.
- DACs should carefully consider applications to access data and should take care that approved projects are within the definitions of applicable law and ethical standards.
- DACs and Data Controllers are to negotiate any necessary legal agreements such as Data Access Agreements with Data Requesters without undue delay.
- Data Controllers are to inform GHGA Central promptly once access requests are approved, so that access can be granted to the approved Data Requester.
- Data Controllers are to inform GHGA Central promptly if access requests are declined.
- Data Controllers are obliged to inform GHGA Central promptly on any other changes regarding the data they control.
- Data Requesters must utilise the Research Data and Personal Metadata they receive safely and within any agreed terms.

***[4] Interoperability with the Federated European Genome-phenome Archive (EGA)*****Services offered:**

- GHGA will serve as the national federated EGA node for Germany. GHGA Central will in the future transfer Non-personal Metadata for data stored in the GHGA Data Infrastructure to the EGA so that data generated in Germany is findable to users of the EGA. GHGA Central may also share the Non-personal Metadata with other metadata archives and collections under appropriate contracts.

**Users' Obligations:**

- Data Submitters are to submit Non-personal Metadata in accordance with the GHGA Metadata Model which has been designed to be interoperable with the EGA Metadata Model.

***[5] GHGA Helpdesk Service*****Services offered:**

- Operation of a Helpdesk service that Service Users can contact for support. The Helpdesk will be available during standard working hours. Service Users contacting the Helpdesk can expect an initial response to their request within 3 working days.

**Users' Obligations:**

- Service Users shall treat Helpdesk staff with respect and engage with them in a professional manner.
- Service Users shall respond when requested by Helpdesk staff within a reasonable time.

*[6] GHGA Validation and Quality Check Service***Services offered:**

- Evaluation of submitted Non-personal Metadata for completeness and consistency, and usage of appropriate ontology terms whenever possible.
- Validation of file uploads to ensure that the submission process was successful.

**Users' Obligations:**

- Data Submitters are to run quality checks on both the Metadata and Research Data files before submission to GHGA. Data Submitters should also assess the anonymity of the Non-personal Metadata they wish to submit.
- When GHGA detects issues with the submitted data, Data Submitters will work together with GHGA Data Stewards and bioinformaticians to correct them.

## §6 - Complaints

All parties of this agreement shall work together in good faith and with respect.

In the event that a Data Controller, Data Submitter, Data Requester, or Service User wishes to make a complaint about the service they have received, they should contact the GHGA Helpdesk<sup>5</sup>.

In the event that the GHGA Helpdesk is unable to resolve a complaint satisfactorily, the matter will be escalated to the GHGA Operations Consortium Board for a formal verdict. The GHGA Operations Consortium shall review all of the evidence provided by the GHGA Helpdesk, the complainant and any other involved parties.

If the complainant is unhappy with the formal verdict of the GHGA Operations Consortium Board they may request an Internal Review. An Internal Review will be a limited scope investigation into whether the complaints procedure has been followed as required, whether all evidence was considered, and whether the formal verdict was reasonable. The complainant may offer additional evidence that was legitimately unavailable to them when the GHGA Operations Consortium Board made their formal verdict. The Internal Review shall be led by a third-party based at the DKFZ and not affiliated with GHGA. The result of the Internal Review shall be considered final.

---

<sup>5</sup> [contact@ghga.de](mailto:contact@ghga.de)

## Annex 1 – Information to be provided before Submission

As part of the Data Submission process, potential Data Submitters will be asked to provide information regarding the following:

- A short description of the data.
- The availability of the data to approved secondary users. Usage conditions should be described, ideally in line with the Data Use Ontology ([DUO](#)) or similar consent management standards.
- The type of data to be deposited (e.g., genomic, phenomic, proteomic, etc).
- The legal identity and location of the Data Controller.
- The approximate total storage size of the data to be deposited (e.g., in GB, TB, or PB).
- Any existing information explaining data access processes or policies (optional).

From these responses the GHGA Operations Consortium Board will assess whether the submission can proceed towards the signing of a Data Processing Contract between GHGA Central and the Data Controller.