

GHGA: Data Protection Framework

Structure of the Document

A. Introduction	2
I. Background: Potential of Omics Data for Progress in Biomedical Research	2
II. Nature of the Data Protection Framework	2
III. GHGA: Aims and mission	3
IV. Glossary	4
B. Data in GHGA	8
I. Research data	8
II. Non-personal Metadata	9
III. Administrative Data	10
1. Personal Administrative Data	10
2. Other administrative data	10
C. Involved actors, their roles in data processing and related data protection responsibilities	11
D. Legal basis of data processing	14
I. Research Data and Personal Metadata submitted by external data submitters	14
II. Research Data and Personal Metadata submitted by an institution operating a GHGA Data Hub	14
III. Access to Research Data and Personal Metadata by the GHGA Data Hubs	15
IV. Processing of personal Administrative Data by the GHGA Data Hubs	15
E. Data transfer	16
I. Data transfer by GHGA Central	16
II. Transfer by the Data Controllers	16
F. Description of data processing	17
I. Concepts for access roles and access rights	17
1. Research Data	17
2. Personal Administrative Data	18
II. Creation and collection of data	18
III. Storage and deletion	19
G. Data subjects' rights	19
I. Related to Research Data and Personal Metadata	19
II. Related to personal Administrative Data	20
H. Annexes: Data protection documentation	20

A. Introduction

I. Background: Potential of Omics Data for Progress in Biomedical Research

Large-scale human omics data has the potential to transform the understanding of disease, leading to improved diagnostics and the development of effective prevention programmes and personalised treatments. Combining genomic data with other molecular data types — such as transcriptomes, microbiomes, proteomics —, the resulting uniquely rich datasets enable integrative analyses to be carried out at unprecedented depth and scale and facilitate new insights into molecular disease processes, thus having implications for basic research and the grounding of personalised healthcare.

Findings from science and research are essential for the interpretation of omics data, for the development of transferable diagnosis and treatment options, and for the continuous evaluation and optimisation of treatment. These activities benefit and depend on access to large integrated datasets. Besides data collected for scientific research purposes, the use of healthcare-driven genomic data being repositioned for research therefore forms an essential part of the long-term setup of a learning healthcare system. Altogether, scientific insights can provide an important contribution to more targeted care and thus ultimately improved treatment.

In order to compile and uniformly analyse human omics data in Germany for the benefit of patients and the public, there is an essential need to create a data infrastructure that corresponds to the needs of data processing for such scientific research whilst also complying with ethical and legal standards, whereby data protection is top priority. As part of the National Research Data Infrastructure (NFDI) Initiative in Germany, the GHGA Consortium is launching an infrastructure to support the sharing of omics data for research while ensuring the data protection needs of the data subjects.

II. Nature of the Data Protection Framework

The Data Protection Framework describes the data processed in the GHGA Data Infrastructure, the main data processing steps, the involved actors, their role in data processing, and their data protection obligations, as well as the legal basis for the processing of data. It clarifies whether international data transfers take place within the data processing in the realm of the GHGA Data Infrastructure. Last but not least, it provides an overview of data subjects' rights.

The Data Protection Framework provides the basis for the data processing registry and other documentations that need to be established by processors, sub-processors of the GHGA Data Infrastructure as well as Data Controllers. The Data Protection Framework will be the basis of the operation of the GHGA Data Infrastructure and will be made available online on the GHGA website (www.ghga.de), where also regular updates will be published. In addition, it will be part of the legal agreements necessary for the operation of the GHGA Data Infrastructure.

III. GHGA: Aims and mission

GHGA, the German Human Genome-Phenome Archive is an academic research consortium aiming to provide national research data infrastructure providing a secure and trustworthy long-term archive of human omics data. To accomplish this aim, GHGA streamlines data deposition by direct data and metadata transmission from major national omics centres and provides subsequent infrastructure to facilitate access and reuse of deposited data for research. Thus, GHGA's role is two-fold: on the one hand, it provides a secure, trustworthy, and convenient repository for omics data, enabling access to large community reference data collections. On the other hand, it is based on the principle of inclusive scientific research through informing affected patients about the use of their data and enabling their participation by including their opinions in policies and research practices. Parallel to this service, it engages national communities to encourage responsible data sharing of omics data, including training of community members and next generation scientists on efficient and responsible use and management of omics data in scientific research.

GHGA is a research consortium consisting of five public universities, seven extramural research entities as well as four university hospitals in Germany, which together form the GHGA Consortium. The GHGA Consortium receives funding by the German Research Foundation as part of National Research Data Infrastructure Germany (NFDI) program. Within the GHGA Consortium, the GHGA Operations Consortium is formed by the following institutions: Eberhard-Karls-Universität Tübingen, Technische Universität München, Technische Universität Dresden, Universität zu Köln, the German Cancer Research Center and the Universitätsklinikum Schleswig-Holstein Kiel. All institutions are organised under public law and fall under the respective federal state level regulations and the oversight by the competent state supervisory authority.

The GHGA Consortium is established with the aim to provide human omics data generated in Germany for scientific research analysis as comprehensively and nationwide as possible. GHGA maximises the scientific, health-related, medical and societal benefits of human omics data (such as data from genomes, transcriptomes, epigenomes, or metagenomes) generated in Germany by providing an infrastructure to compile and store this data as well as by opening up the possibility of its further processing *for health-related scientific research purposes in the public interest*. Through its research and community activities, GHGA shall contribute to leveraging the scientific potential of data infrastructures and community reference datasets, opening up novel scientific avenues based on omics research in the interest of both individualised and population-based benefits for patients and the public.

Data processing that takes place by relying on GHGA Data Infrastructure shall thus fulfil the following purposes:

- integration, storage and provision of human omics data;
- development of research methods and analysis of human omics data;

- creation of the scientific basis for genomic medicine in Germany on the technological level and by adopting appropriate policies;
- establishment of an infrastructure for Germany through which the offers of cooperation and coordination by the European Union in the area of scientific research can be perceived;
- design of this infrastructure for Germany in a way that it can function as a gateway towards global connectivity in the area of human omics research, increasing the scientific value and FAIRness of omics data.

When dealing with human omics data, the institutions involved in the GHGA Consortium are aware of the sensitive character of the data that is processed in their infrastructure, the scope of application of data protection law regarding their data processing activities as well as their corresponding data protection obligations, and have established the following Data Protection Framework. This Data Protection Framework represents the current operating processes of the GHGA Data Infrastructure and GHGA Operations Consortium. As the Project matures there are likely to be additional or differing data processing activities undertaken, these changes will be included in this Framework which should be understood to be a living document.

IV. Glossary

European Genome-phenome Archive

The European Genome-Phenome Archive (EGA) provides archiving and sharing support for personally identifiable genetic and phenotypic data. It is operated jointly by the European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI), an intergovernmental organisation, and the Centre for Genomic Regulation in Barcelona (CRG). The EGA is developing a federated model through which national nodes will take on the archiving of genomic-phenomic data for their country; this federated network of institutions will be referred to as the federated EGA (fEGA). EGA will take over the coordinating function as operator of EGA-Central, GHGA is the national node for Germany.

GHGA Central

Deutsches Krebsforschungszentrum (DKFZ) shall be the legal entity responsible for GHGA Data Infrastructure. The responsibilities are outlined in the GHGA Data Hub Cooperation Contract and are to be fulfilled in accordance with the regulations outlined in that contract. These responsibilities will include taking on the role of a data processor for the Research Data and Personal Metadata submitted to the GHGA Data Infrastructure. DKFZ is referred to as GHGA Central when acting in this role. DKFZ will also operate a GHGA Data Hub.

GHGA Consortium

The organisations that have signed the GHGA Cooperation Contract in December 2020 and receive (some of) their funding as part of the Nationale Forschungsdateninfrastruktur (NFDI) from the German Research Foundation (DFG) and are working as part of the GHGA Project.

GHGA Data Hub

A GHGA Consortium member that stores and processes Research Data and Personal Metadata for the purposes of archiving and secondary analysis for scientific research purposes on behalf of GHGA Central. The responsibilities are set out in the GHGA Data Hub Cooperation Contract.

GHGA Data Infrastructure

GHGA Central, together with the GHGA Operations Consortium, will operate an IT infrastructure to enable FAIR (findable, accessible, interoperable and reusable) data sharing of human omics data for scientific research purposes as defined by the GHGA Data Hub Cooperation Contract, the Joint Controller Contract and the Central-to-Data Hub Bilateral Contracts.

GHGA Helpdesk

The GHGA Helpdesk will be operated by staff from each institution that operates a GHGA Data Hub. The GHGA Helpdesk will be the main point of contact for service users. The helpdesk software is operated by the DKFZ as GHGA Central.

GHGA Operations Consortium

GHGA Central and the GHGA Data Hubs who are responsible for storing and processing Research Data and Personal Metadata.

GHGA Project

The Project refers to the overall GHGA structure. This includes GHGA Central, the GHGA Consortium, and GHGA Partners who are working together to make genomic-phenomic data available.

Administrative Data

Data which are generated through the operation of GHGA Data Infrastructure. This may include personal data which is directly identifying, such as names and email addresses which are used to communicate with, and support, Service Users. It may also include personal data and business data which are used internally by staff working on behalf of GHGA Central or GHGA Data Hubs. Personal Administrative Data is jointly controlled by the GHGA Operations Consortium members according to the Joint Controller Contract.

Metadata

Information that describes or annotates Research Data to aid understanding or to describe the relationship between data items. It may be personal or non-personal.

Non-personal Metadata

Information that describes or annotates Research Data to facilitate its interpretation or to describe the relationship between data elements. For example, the name of the instrument used to generate the data or information defining a group of Data Subjects. Non-personal Metadata will be available for public online search within the GHGA Data Infrastructure.

Personal Metadata

Information that describes or annotates Research Data to facilitate its interpretation or to describe the relationship between data elements. For example, demographic data or information on the ancestry of the Data Subjects of the Research Data that allow conclusions to be drawn about individuals and thus fall within the scope of the GDPR, Art. 4 No. 1 GDPR. Personal Metadata are made available to the Data Requester together with the Research Data only under controlled access after release by the Data Submitter.

Omics Data

The Research Data collected as part of omics-based research. This research focuses on collecting information regarding the entire set of certain molecules in a sample. Within the context of GHGA, Omics Data linked to genetic information of an individual are of particular interest since in many cases omics data would fall under the definition of personal data in Art. 4 Nr. 1 GDPR. Types of Omics Data considered in GHGA are e.g.: genomics – the entirety of the hereditary information in a sample's DNA; transcriptomics – the entirety of the RNA transcribed from DNA; epigenomics – information on epigenetic modifications of the genetic materials.

Research Data

Omics or other forms of genetic (Art. 4 Nr. 13 GDPR) and health data (Art. 4 Nr. 15 GDPR) that are used for scientific research purposes. This is considered to be special category personal data under Art. 9 (1) in conjunction with Art. 4 Nr. 1 GDPR.

Data Controller

Institution that is the controller (as defined by the GDPR) of Research Data and Personal Metadata and that contracted GHGA for processing the Research Data and Personal Metadata on their behalf. The term is also used for the corresponding party if the processing of the Research Data and Personal Metadata is only intended, i.e. before such a contract is established.

Data Access Committee

An abstract body, potentially a panel of people, deciding on data access requests and authorised by the Data Controller.

Data Requester

User or institution that requests access to Research Data and Personal Metadata via the GHGA Data Infrastructure.

Data Steward

Data Stewards are the staff members who (besides other duties) operate the GHGA Helpdesk. They are responsible for assisting users of the GHGA Data Infrastructure with data transfer to the GHGA Data Infrastructure, access requests and responding to queries. Data Stewards will be employed at each of the GHGA Data Hubs.

Data Submitter

Users who are depositing data with GHGA Central (and includes the Data Controller, if not the same person(s), as defined in the Data Processing Contract).

Service User

Person who is contacting the GHGA Helpdesk for any form of user support.

Central-to-Data-Hub Bilateral Contract

Agreement between GHGA Central and a GHGA Data Hub. Based on the GHGA Data Hub Cooperation Contract it regulates the relationship between GHGA Central and the Data Hub and the corresponding rights and responsibilities in full detail. In particular it defines the processor to sub-processor relationship between GHGA Central and the individual Data Hub. It also enables adjustments with respect to local infrastructures and federal data protection law where required.

Data Processing Contract

Bilateral agreement signed by GHGA Central and a Data Controller who wishes to deposit data in the archive. The agreement regulates the rights and duties of the controller and GHGA Central in processing the deposited data.

Federated EGA Collaboration Agreement

Agreement between GHGA Central and EGA Central. Defines the rights and responsibilities of EGA Central and GHGA as the German node. Describes the governance structure of the Federated EGA network. Includes details around IP, liability, disputes, and off-boarding. No personal Research Data or Personal Metadata is exchanged between GHGA Central and EGA Central (see §2.4.2).

GHGA Cooperation Contract

Regulates the organisation of the GHGA Consortium to carry out the GHGA Project. Does not regulate the exchange of Personal Data.

GHGA Data Hub Cooperation Contract

This contract is agreed by GHGA Central and GHGA Operations Consortium members that operate, or wish to operate, a GHGA Data Hub. The contract sets out the structure of the Operations Consortium, including a definition of the roles and responsibilities of members as well as the data governance framework. The appendices define a number of common standards that will be utilised by GHGA, including the Joint Controller Agreement for Personal Administrative Data.

Joint Controller Agreement for Personal Administrative Data

The document governs the joint controllership of the Administrative Data by the institutions which are part of the GHGA Operations Consortium. Appendix to GHGA Data Hub Cooperation Contract.

Metadata Processing Contract

This agreement covers the deposition and sharing of metadata in the GHGA Metadata Catalog during the GHGA Catalog phase of the project. During GHGA Catalog, only Non-personal Metadata describing Research Data will be processed and shared; the corresponding Research Data and Personal Metadata remain with the Data Controller and are not stored by the GHGA Operations Consortium.

B. Data in GHGA

I. Research data

1. *Omics data submitted by a Data Submitter:* DNA and/or RNA sequencing data, or other omics data produced for a study will be submitted to the GHGA Data Infrastructure. GHGA Central will accept both the raw data like the sequencing reads, and also processed results from these raw data with relevant data processing metadata (i.e. software and version used in the processing, cf. Section II. below). Submitted omics data should not contain any directly identifiable information, and should be pseudonymised before it is shared via the GHGA Data Infrastructure.

2. *Processed genomics data produced by GHGA:* If instructed to do so by the Data Controller, GHGA Central will perform standardised data processing on the omics data submitted on their before. This processing will produce general quality control measures (e.g. on file integrity), quality metrics (e.g. sample validity, sequencing quality, etc.) and processed results (e.g. standardised sequencing alignment), and will be linked to the raw data they come from.

3. *Information about individuals involved in the study stored on the GHGA Data Infrastructure:* Some personally identifiable information, such as detailed phenotypic and demographic information about individuals will be stored within the GHGA Data Infrastructure in a pseudonymised form as Personal Metadata. These metadata will be stored separately from

the public Non-personal Metadata, and will only be shared under controlled access conditions (as all other data aside from the public Non-personal Metadata). The rights of the individuals affected can be enforced via the Data Controller for this data.

4. Identifiers: GHGA Central will store unique identifiers for each study, experiment, sample, file, and individual metadata submitted to the GHGA Data Infrastructure. Data Submitters will be able to share a pseudonymised, unique identifier for each item they submit. GHGA Central will also generate its own unique, standardised identifiers, both in the form of a long, universally unique UUID, and also a simpler accession number (i.e. GHGA-SA-000001). If the datasets are also deposited in other relevant data repositories, GHGA Central will make links to these identifiers with cross-references.

Accordingly, *Research Data* processed in the GHGA Data Infrastructure includes personal data, as defined by Art. 4 Nr. 1 GDPR, and sensitive data, according to Art. 4 Nr. 13 (genetic data) and Art. 4 Nr. 15 (health data) in conjunction with Art. 9 (1) GDPR. Furthermore, it includes both personal and sensitive data processed for scientific research purposes as a primary data processing purpose, and data collected for healthcare purposes and deposited in the GHGA Data Infrastructure for further processing for scientific research purposes. All identifying or identifiable Research Data will be received by GHGA Central in pseudonymised form, as defined by Art. 4 Nr. 5 GDPR although items such as a whole genome sequence will remain unique to an individual.

II. Non-personal Metadata

Non-personal Metadata processed within the GHGA Data Infrastructure is a non-personal form of metadata that describes a given research dataset and is shared publicly via the GHGA Data Portal. It is used by researchers searching for data that is relevant for their needs. This will include the following information (subject to further extension with non-personal data):

- 1. Dataset:* Defines the collection of runs/analysis data files that will require a controlled access mechanism. This is associated with a Data Access Policy.
- 2. Analysis:* References the files that contain genomic-phenomic analyses (typically BAMs or other similar formats). This is associated with Sample and Study.
- 3. Run:* Links together Sample, Experiment, and Files. Any file generated from a sequencing run is linked to an Experiment via a Run object.
- 4. Experiment:* Contains information about the sequencing methods, protocols, and machines used. Links together Sample and Study.
- 5. Sample:* This will contain high-level information about the sequencing sample (Data Subject).
- 6. Study:* Contains information about the sequencing study.
- 7. Data Access Committee:* Contains the information about the DAC, including the contact person.
- 8. Data Access Policy:* Sets out the policies regarding data access, defined using the DUO framework. This is associated with a DAC.

Data Controllers will be requested to confirm having conducted a data protection risk assessment that the metadata submitted as Non-personal Metadata does not fall within the scope of application of data protection law, as per Art. 4 Nr. 1 GDPR in conjunction with recital 26 GDPR.

III. Administrative Data

1. Personal Administrative Data

The following personal data is collected exclusively from GHGA users in the context of GHGA user administration (Administrative Data). Such data will not be collected or received by GHGA on patients and probands related to the submitted Research Data and Personal Metadata.

The following data is stored for each user:

1. Data Submitter:

- a. Name
- b. Email address
- c. Institutional affiliation
- d. Phone number (if Data Submitter provides it)
- e. Elixir-AAI ID to confirm identity

2. Data Requester:

- a. Name
- b. Email address
- c. Institutional affiliation
- d. Phone number (if Data Requester provides it)
- e. Elixir-AAI ID to confirm identity

3. Service User - user who is contacting GHGA for any other support

- a. Name
- b. Email address
- c. Institutional affiliation
- d. Phone number (if Service User provides it)

2. Other administrative data

Other administrative data processed by GHGA will be non-personal data generated by GHGA Central and the GHGA Data Hubs staff and comprises the following:

Non-personal Data	Producer
a. Standard Operating Procedures	GHGA Central and GHGA Data Hubs

b. Internal and external policy documents	GHGA Central and GHGA Data Hubs
c. Statistics relating to the GHGA's performance for example, the number of depositors, the number of datasets	GHGA Central and GHGA Data Hubs
d. Press releases	GHGA Central and GHGA Data Hubs

C. Involved actors, their roles in data processing and related data protection responsibilities

The main setup of the GHGA Infrastructure is based on the distribution of data protection roles in a way that secures the data controller position of Data Controllers over their data, with GHGA Central taking up the role of a data processor and the GHGA Data Hubs becoming sub-processors. This means that the decision on sharing Research Data and Personal Metadata via the GHGA Infrastructure and providing access to it, remains with the respective Data Controller; GHGA Central as a processor acts only upon instruction by the Data Controllers. In cases whereby a GHGA Data Hub shares data via the GHGA Infrastructure over which they exercise data protection control, the data protection roles remain assigned as sketched above and based on the Data Submitter function of the respective GHGA Data Hub.

	Role(s)	Data protection roles	Contracts
DKFZ (GHGA Central) foundation under public law	<ul style="list-style-type: none"> - Leader of the GHGA Consortium - Member of the GHGA Consortium - Research Data and Personal Metadata provision for Data Requesters on instruction of the Data Controller - User services - Interaction with EGA Central related to the 	<ul style="list-style-type: none"> - data processor for Research Data and Personal Metadata submitted by Data Controllers - joint controller for personal administrative data together with the other data hubs 	<ul style="list-style-type: none"> - GHGA Cooperation Contract - GHGA Data Hub Cooperation Agreement, incl. contract for the processing of personal Administrative Data based on Art. 26 GDPR (joint controller) - Central-to-Data Hub Bilateral Contract with the GHGA Data Hubs (processor-sub-processor)

	<p>sharing of personal Administrative Data</p> <ul style="list-style-type: none"> - Negotiate contracts on behalf of the GHGA Consortium with relevance for personal data processing as defined by the Consortium contract - Act as a central point of contact for national and international cooperation partners related to the processing of personal data 		<ul style="list-style-type: none"> - Data Processing Contract as a processor based on Art. 28 GDPR (controller-processor) - Contract with EGA Central (EMBL-EBI) and CRG
GHGA Data Hubs	<ul style="list-style-type: none"> - Member of the GHGA Consortium - Data processing: Research Data and Personal Metadata archival/encryption/file storage on behalf of the Data Controller, as instructed by GHGA Central - Research Data and Personal Metadata provision for Data Requesters as instructed by GHGA Central - 	<ul style="list-style-type: none"> - data sub-processor to GHGA Central for Research Data and Personal Metadata submitted by Data Controllers - joint controller for personal Administrative Data together with the other Data Hubs 	<ul style="list-style-type: none"> - GHGA Cooperation Contract - GHGA Data Hub Cooperation Agreement, incl. contract for the processing of personal Administrative Data based on Art. 26 GDPR (joint controller) - Central-to-Data Hub Bilateral Contract with GHGA Central (processor-sub-processor) [- Authorisation by the Data Controller in the Data Processing Contract based on Art. 28 GDPR (controller-processor)
	<p>Research data providing entity (Data Submitter to the GHGA Data Infrastructure)</p>	<ul style="list-style-type: none"> - Data transmission into the GHGA Data Infrastructure - (sole) Data Controller for Research Data and Personal Metadata submitted by the Hub itself 	<ul style="list-style-type: none"> - Data Processing Contract with GHGA Central based on Art. 28 GDPR (controller-processor)

	Scientific research with the Research Data and Personal Metadata	Data Requester	- Data Access Agreement with the Data Submitter as Data Controller
Data Controllers (clinics, research institutions, healthcare providers) and the Data Submitters who operate on their behalf.	Research Data and Personal Metadata providing (data submission into the GHGA Data Infrastructure)	- Data transmission into the GHGA Data Infrastructure - Data Controller for research data	- Data Processing Contract with GHGA Central based on Art. 28 GDPR (controller-processor)
Data Requesters (public and private entities, their employed and affiliated researchers)	Scientific research with the Research Data and Personal Metadata	Data Requester	- Data Access Agreement

The **GHGA Data Hub Cooperation Agreement** further details the GHGA Consortium Contract and sets out agreements between GHGA Operation Consortium members, including agreements on services provided through the consortium, main modalities of data sharing and distribution, the governance structure, and risk management. It defines joint controllership for personal Administrative Data.

Joint Controller Agreement: as part of the GHGA Data Hub Cooperation Agreement, the GHGA Data Hubs and GHGA Central have entered a contract of joint controllership for the processing of personal Administrative Data in accordance with Art. 26 GDPR.

Central-to-Data Hub Bilateral Contracts are signed between a GHGA Data Hub and GHGA Central through which an institution wishing to operate a Data Hub commits to acting in relation to the role described in the cooperation contract as a sub-processor for Research Data and Non-personal Metadata to GHGA Central.

Data Processing Contract: GHGA Central as a data processor for Research Data and Non-personal Metadata will enter into data processing agreements with Data Controllers in accordance with Art. 28 GDPR. In this contract, the GHGA Data Hubs will be defined as sub-processors of research data (not submitted by themselves as Data Controllers) and will thus secure authorisation of their engagement by the Data Controller.

Data Access Agreement: DAAs are agreed between Data Requesters and Data Controllers. The GHGA Consortium, with particular guidance by its Ethical and Legal Workstream, will produce a template version that Data Controllers can adapt and use.

D. Legal basis of data processing

I. Research Data and Personal Metadata submitted by external data submitters

The submission of Research Data and Personal Metadata by external data submitters and its provision to those requesting access is based on different legal bases depending on the applicable state and federal legal requirements for them as data controllers. GHGA Central as a processor is bound by the instructions for data processing by the Data Controllers. When implementing and executing these instructions, the GHGA Data Hubs as sub-processors are bound by the instructions of the GHGA Central as a processor.

In any case, it remains the obligation of the respective Data Controller to secure and be able to prove that they have a sufficient and robust legal basis for engaging GHGA Central as a data processor, together with the GHGA Data Hubs as sub-processors, to process personal data via the GHGA Data Infrastructure.

II. Research Data and Personal Metadata submitted by an institution operating a GHGA Data Hub

For Research Data and Personal Metadata submitted by the GHGA Data Hubs, the institutions operating them remain data controllers. The legal basis for submitting data is to be defined based on the applicable data protection law for the relevant institution. This also applies for new Data Hubs joining the GHGA Operations Consortium.

As the most important and viable legal basis, GHGA Data Hubs should verify the explicit consent of the Data Subject according to Art. 9 (2) (a) GDPR and specific provisions thereto in federal state data protection laws as well as federal state hospital laws. Alternatively, drawing on broad consent of the Data Subject in accordance with Rec. 33 GDPR as well as the applicable resolutions of the Conference of the Data Protection Supervisory Authorities can be proved. Finally, research clauses in applicable federal state data protection laws, such as § 36 LDSG Saxony, § 13 (1) LDSG Baden-Wuerttemberg, § 17 DSD North-Rhine Westphalia, as well as the research clauses in federal state hospital laws such as § 38 LKHG Schleswig-Holstein and Art. 27 (4) (1) BayKrG might provide a sufficient legal basis to process data via the GHGA Data Infrastructure. Additionally, provisions on the modalities on how data can be processed for scientific research purposes in these applicable laws need to be respected, even when they do not serve as a legal basis for personal data processing for scientific research purposes, such as Art. 25 BayDSG.

As previously, it remains the obligation of the respective Data Controller to secure and be able to prove that they have a sufficient and robust legal basis for engaging GHGA Central as a data processor, together with the GHGA Data Hubs as sub-processors, to process personal data via the GHGA Data Infrastructure.

III. Access to Research Data and Personal Metadata by the GHGA Data Hubs

The legal basis for access to data processed within the GHGA Data Infrastructure for scientific research purposes by any Data Requester is to be defined based on the applicable data protection law for the Data Requester. If the Data Requester is based at an institution operating a GHGA Data Hub, then the relevant data protection law of the federal state and, in certain cases, the hospital law of the federal state shall apply. Hereby, GHGA Data Hubs as Data Requesters should particularly verify, whether legal basis provisions related to data processing for scientific research purposes are applicable for their accesses; such as § 13 (1) LDSG-BW, § 36 LDSG Sachsen; § 17 DSD NRW; § 13 LDSG Schleswig-Holstein. The duty of the controller to define the legal basis for the transmission remains unaffected.

IV. Processing of personal Administrative Data by the GHGA Data Hubs

The processing of personal Administrative Data generated within the GHGA Data Infrastructure (see Section B.III.1) is carried out in accordance with Art. 6 (1) (b) as well as Art. 6 (1) (f) GDPR for the purpose of user administration, i.e. in particular:

- to manage authorisations for the transmission as well as the retrieval of Research Data and Personal Metadata,
- to document the authorship of the transmitted Research Data and Personal Metadata,
- to communicate with Service Users via the GHGA Helpdesk e.g. to provide support or guidance on using the GHGA service,
- the processing of personal Administrative Data regarding staff working on behalf of GHGA.

The processing of personal Administrative Data is described in the Data Processing Contract and the Data Access Agreement agreed between the Data Requester and the Data Controller respectively (cf. F.II below). For other forms of personal Administrative Data, GHGA Central and the GHGA Data Hubs has a legitimate interest in processing the data in order provide a service to users of GHGA. Finally, the processing of personal Administrative Data belonging to staff working on behalf of GHGA is described in their employment contracts with their respective institutions. The data security measures of Section G below also apply to these data.

The processing of GHGA personal Administrative Data is also carried out for purposes of authentication and authorisation on the user portal as well as to facilitate user navigation on the GHGA web portal.

The GHGA Data Hubs and GHGA Central act as joint controllers of personal Administrative Data, as described in the Joint Controller Agreement between the parties, and in accordance with Art. 26 GDPR.

E. Data transfer

I. Data transfer by GHGA Central

Non-personal metadata and non-personal Administrative Data (cf. Section B) are planned to be transferred to the institutions operating the EGA. As part of the preparation for and the process of transferring data, GHGA Central carries out a risk analysis to ensure that the data transferred falls outside the scope of the GDPR, because it does not constitute personal data within the meaning of Art. 4 No. 1 GDPR and a potential identification of the data subjects pursuant to recital 26 s. 3 and s. 4 GDPR is not reasonable, as the transferred data does not allow the establishment of a personal reference according to this standard.

Furthermore, the technical and organisational measures taken by the GHGA Central and the GHGA Data Hubs (cf. Section G and H) ensure that the risk of establishing the personal link is excluded or, if applicable, remains below the standard of reasonableness according to recital 26 s. 3 and s. 4 GDPR.

The basic modalities for this are defined by the Federated EGA Contract agreed by GHGA Central on behalf of the GHGA Operations Consortium.

II. Transfer by the Data Controllers

Data Controllers will be responsible for approving access to the Research Data and Personal Metadata that they control via the GHGA Data Infrastructure. GHGA Central will only make the Research Data and Personal Metadata available to approved researchers under instruction from the Data Controller.

Researchers may request access data from any country, and it will be the responsibility of the Data Controller to ascertain if an adequate legal basis as well as a suitable admissibility ground exists for the transfer of the Research Data and Personal Metadata. GHGA Central will request that the Data Controller includes the legal basis and the admissibility ground for the transfer in the Data Access Agreement. The GHGA Consortium will also provide generic guidance on suitable legal bases and admissibility grounds, non-specific to any particular transfer request. This generic guidance will be supported by legal analyses provided by the Ethical and Legal Workstream.

The data processing by GHGA Central and the GHGA Data Hubs does not affect authorisation of international transfers of Research Data and Personal Metadata by the relevant Data Controllers.

F. Description of data processing

I. Concepts for access roles and access rights

Access to data will be defined by a person's role, and the type of data they wish to access. In the GHGA Data Infrastructure, three types of personal data will be processed, Research Data which will include genomic information, Personal Metadata which describes the Research Data and is treated with the same controls, and personal administrative data which is generated as users interact with the GHGA Data Infrastructure. It will also be important to consider differences between the roles mediated by the type of data being processed.

1. Research Data

a. Data Controller - The Data Controller is responsible for taking decisions regarding the purposes and essential means of the processing of data which is subject to the GDPR. The Research Data and Personal Metadata submitted to GHGA Central will have a Data Controller as it is considered personal data. A Data Processing Contract will be signed by GHGA Central and the Data Controller to authorise and set out the terms by which GHGA Central may process the Research Data and Personal Metadata. In many cases, the Data Controller will be the legal entity that the Data Submitter is affiliated to, and the Data Submitter will be sharing research data via the GHGA Infrastructure in their function as an affiliate of that Data Controller. A Data Controller will be able to access the Research Data and Personal Metadata that they have archived through GHGA Central upon request.

b. Data Submitter - The Data Submitter is the person or persons who have submitted data to GHGA Central. The legal entity affiliated with the Data Submitter (in its function of submitting the data) will usually also be the Data Controller. In the event that the Data Submitter has deposited data on behalf of a Data Controller, but has no further right to access the data, the Data Submitter will be required to make an application to the Data Controller in order to access the data for research purposes. Thorough documentation of their relationship and rights related to the research data will be documented in the Data Processing Contract.

c. Data Requester - The individual or the individual(s) on behalf of the organisation that requests access to the Research Data and Personal Metadata stored in the GHGA Data Infrastructure for scientific research purposes. They will be required to make an application to a Data Controller in order to be granted access. Once a Data Controller has approved access, the Data Controller will notify GHGA Central. GHGA Central and the GHGA Data Hubs upon its instruction will then make the data available to the approved Data Requesters.

d. Data Stewards - Data Stewards process the data on behalf of GHGA Central under the terms of the Data Processing Contract based on Art. 28 GDPR. They will be responsible for overseeing the submission of Research Data and Personal Metadata and disseminating it to approved researchers upon authorisation by the respective Data Controller. The Research Data and Personal Metadata will be encrypted in transport and at rest. Controls are in place to prevent the unauthorised decryption of the Research Data and personal Metadata

2. Personal Administrative Data

a. *Service Users* - A Service User is any individual that is engaging with the GHGA Data Infrastructure in order to use its service. Service Users will be able to access the data stored about them via the GHGA Data Portal if they have registered. Non-registered users will have to request access to the Administrative Data held about them. Should Service Users wish to exercise their rights as Data Subjects they will contact the Helpdesk who will be able to support their request.

c. *Data Stewards and Project Management workstream* - Data Steward will be responsible for providing a service to Service Users. It will be necessary for them to access personal Administrative Data regarding Service Users to do so. This form of personal Administrative Data will primarily be stored within the GHGA Helpdesk software and may only be accessed through the same. Data Stewards will not be permitted to access the data for any other purpose or through any other method. The Project Management workstream oversees the GHGA Helpdesk, and supports Data Stewards in the triaging of tickets.

Personal Administrative Data belonging to staff working on behalf of GHGA will be accessible to Data Stewards and the Project Management workstream insofar as access to it is required as part of their role.

d. *Staff working on behalf of GHGA* – As a general principle, staff working on behalf of GHGA shall have as little access to personal Administrative Data as is required as part of their role.

II. Creation and collection of data

Research Data and Personal will be collected by researchers and deposited at GHGA Central and its Data Infrastructure. It is expected that for a majority of data submissions, informed consent will be the legal basis that is used (cf. Section D.II). Data Controllers will be required to confirm that they have an appropriate legal basis to share Research Data and Personal Metadata when they deposit it within the GHGA Data Infrastructure. GHGA has developed guidance on suitable consent models but will not be directly involved in the creation or collection of data.

Personal Administrative Data will be collected through a number of routes. Service Users may choose to register with GHGA for the purposes of accessing support, and and/or submitting or requesting Research Data and Personal Metadata. Service Users may also contact the GHGA Helpdesk directly, by email or phone, and it may be necessary for the Data Steward who is processing the request to transmit the information into the Helpdesk software. This processing will rely on legitimate interest as a legal basis (cf. Section D.IV).

Personal Administrative Data relating to Data Submitters and Data Requesters will also be processed by the GHGA Operations Consortium. Personal Administrative Data regarding Data Submitters will be collected via the Data Processing Contract agreed between a Data Submitter and GHGA Central. This data is required to ensure that there is a clear designation of who may act on behalf of the Data Controller for the duration of the Data Processing

Contract. This data will be collected using 'processing is necessary for the performance of a contract' as a legal basis, Art. 6 (1) (b) GDPR.

Additionally, Personal Administrative Data relating to Data Requesters will be collected via the Data Access Agreement agreed between the Data Requester and the Data Controller. This data is required by the GHGA Operations Consortium in order to make Research Data available to approved researchers. It is important that the GHGA Operations Consortium is able to verify the identity of the approved researchers before sharing the personal Research Data. GHGA will supply an Information Sheet to Data Controllers for them to include in their Data Access Agreements describing the processing of the Data Requester's data. This data will be collected using 'processing is necessary for the performance of a contract' as a legal basis, Art. 6 (1) (b) GDPR.

Personal Administrative Data relating to staff working on behalf of GHGA will be collected as part of their onboarding process to the Project.

III. Storage and deletion

Research Data and Personal will be stored by GHGA Central at one of a number of GHGA Data Hubs. All GHGA Data Hubs will be governed by the GHGA Data Hub Cooperation Agreement and Bilateral Contracts which will describe the data protection standards required by GHGA Central based on the Data Processing Contract. GHGA Data Hubs will store the Research Data and Personal Metadata on the de.NBI cloud or similar secure and specialist infrastructure. The data will be stored for as long as GHGA is able to support the service, or until such time that the Data Controller wishes to terminate their agreement with GHGA, or whilst the Data Controller is able to meet their obligations.

Data Controllers may request the destruction of the Research Data and Personal Metadata. If they wish for this data to be removed from the GHGA Data infrastructure the Data Destruction SOP will be followed by a Data Steward. This will result in the secure and irretrievable destruction of the Research Data and Personal Metadata and the issuing of a Data Destruction Certificate to confirm that the data is no longer held within the GHGA Data Infrastructure.

Personal Administrative Data will be stored according to the terms of the GHGA Data Deletion Concept. The designated retention period varies depending on the nature of the personal Administrative Data and its processing purpose.

G. Data subjects' rights

I. Related to Research Data and Personal Metadata

Sufficient guarantees are given to the Data Subjects related to their rights and freedoms. The Data Controllers carry out pseudonymisation through established processes prior to the submission of Research Data and Personal Metadata into the GHGA Data Infrastructure. The Data Controllers are responsible for the technical and organisational measures to secure

these pseudonyms. The present data protection concept only applies from the encrypted transmission of data to GHGA Central as a processor. From this point on, the IT security measures described in Section H.III apply.

Against the background of the relationship between the actors involved (cf. section C), and the envisaged contracts (which will determine the obligations under data protection law), as well as because of the pseudonymised forwarding of data by the Data Controllers to the GHGA Central, the Data Controllers remain contact points for Data Subjects to implement their data subject rights according to Art. 15 et seq. GDPR. GHGA Central supports the Data Controllers in answering data subjects' requests to exercise their rights. Any requests from Data Subjects submitted to the GHGA Central will be forwarded to the respective Data Controller. If necessary, the Data Controllers provide the required information and data or initiate the required implementation.

GHGA Central supports transparency for data subjects by providing information about research projects and the scientific research and data processing carried out via the GHGA Data Infrastructure on its website.

Exact details of the distribution of the responsibilities for the implementation of data subject rights and the notification and information obligations between the Data Controllers, the data processor and the data sub-processors, including auditing, as well as the fulfilment of data protection obligations will be defined in the contracts according to Art. 28 (3) GDPR.

II. Related to personal Administrative Data

All data subject rights related to personal Administrative Data will be implemented and managed by the GHGA Data Hubs and GHGA Central as joint controllers for this data. A Data Subjects' Request Procedure has been developed and will be enacted in the event that a Data Subject seeks to exercise their rights. The Joint Controller Agreement detailing their relationship internally as joint controllers, based on Art. 26 GDPR, has defined the distribution of responsibilities for data subjects' rights.

H. Annexes: Data protection documentation

Administrative Data

- Record of Processing Activity

Non-personal Metadata

- Privacy Impact Assessment