

NEWSLETTER

JULY 2021



WELCOME

Resulting from the wave of hiring in spring, we now see an explosion of activity within GHGA. So while the first newsletter introduced you to the GHGA team, the second edition will focus on the different work streams and their progress. You can read an update on the two task area overarching working groups: the architecture and DMP/Legal working groups. In addition, you get to meet the other work forces, ranging from the newly formed outreach working group (combining activities from A1, A2, B1, C1 and C2), the ethico-legal task force (B1), the close collaboration of task area B2 and B3 on metadata to task forces around data stewardship (C1) and workflows (C2).

We also welcome four new faces to the GHGA team! Check out who they are in this newsletter.

In our News section, you can read about the metadata workshop hosted by GHGA, and our virtual booth at the LifeTime 2.0 Conference. We report on a big step for CoGDat and you can see just how sporty the GHGA team is!

Last but not least, the date for the GHGA annual meeting is set. We hope to welcome you in Heidelberg for an - Corona permitting- in person meeting on the 27th and 28th of October! The GHGA team will add one more day, and meet already on the 26th for team day. Please block your calendar on those days.

We hope you enjoy the Newsletter.

The GHGA Board of Directors and the GHGA Coordination Team

CONTENT:

[GHGA Team](#)

[GHGA Working Groups](#)

[News](#)

[Save the Dates](#)

[Jobs in GHGA](#)

GHGA TEAM

The GHGA team keeps on growing. Since May we welcomed four new team member.

Anna is supporting the GHGA outreach efforts, Eric joined the ethics team, and Jordy was hired to support the technical team. We also had Léon - who some may recognise from CoGDat- join GHGA.

The full profiles, where you can also hear from each person themselves can be found on the [GHGA Confluence page](#).

ANNA BENET-PAGES
OUTREACH OFFICER
MÜNCHEN



Main tasks:

Coordination of the Munich hub

Outreach activities, focused on RD and the public

JORDY ORELLANA FIGUERO
SOFTWARE DEVELOPER
TÜBINGEN



Main tasks:

GHGA data management software

Front- and Back-End development

GHGA Data Hub operations

ERIC APONDO
RESEARCH FELLOW
HEIDELBERG



Main tasks:

Patient engagement, focused on cancer community

Concerns regarding patient information, care and management of risk in clinical decision-making

LÉON KUCHENBECKER
SOFTWARE DEVELOPER
TÜBINGEN



Main tasks:

Coordination of technical development in Tübingen

CoGDat coordination

Front- and back-end development

GHGA WORKING GROUPS

Architecture Working Group

By Koray Kirli

The Architecture Working Group welcomed two new members since the last newsletter went out, Jordy and Léon are joining us from Tübingen University. Our team is constantly growing with active job ads for DevOps engineer and software developer positions (ghga.de or see Jobs sections).

The development team has made a lot of progress on many fronts in the past months. We started working on a new roadmap around the designed services and planned technologies. Our first step is to use the software stack with a simplified set of services to test the integration with “the sandbox project”. We completed our first sprint (2 week development effort starting with defining goals and ending with evaluation of progress and a feedback round) for the sandbox project, where we used Jira as our task management solution integrated with code repositories for progress tracking.

The sprint effort also helps establish our development culture. We are working on capturing these guidelines as SOPs, and so far found 4 topics to work on; (1) Digesting milestones to tasks, (2) Ingesting external feedback, (3) Managing development tasks, (4) Going from code to production. We will expand our procedure topics and plan to have a mechanism that would be easy to audit.

The team spent quite some time learning new technologies and also started running education sessions. At the Architecture Working Group meeting on June 29th, we ran a session on deployment solutions and security features presented by Philipp Walz. The content was focused on Kubernetes and related tools, and sparked further discussions for collaborations around these deployment solutions with other projects. Jordy ran another education session on July 13th about Advanced Message Queuing Protocol (AMQP), a technology that is integral for communication within the microservice architecture.

The development team is also exploring the existing research/sensitive data development ecosystem by engaging in discussions with other projects and attending workshops. We held a meeting with lead EGA developers from the EMBL-EBI (UK) and CRG (Spain), learned about the ICGC ARGO stack, and attended GA4GH presentations. We will continue with more discussions with other federated EGA nodes.

Data Management Plan/ Legal Working Group

by Simon Parker

In recent months, the DMP and Legal Working Group has been focussed on producing the agreements and contracts that GHGA will need to process research and user data. These documents will set out the responsibilities for those institutions that are processing data, and how those institutions will be governed within the broader GHGA consortium. We have presented the draft versions of agreements to the data hubs on July 8th. Our intention is to now engage with external legal professionals to develop our drafts into formal legal documents.

We have also been working with the Data Protection and Legal teams at DKFZ to produce the notifications of our proposed data processing and to determine our data protection requirements. This has been a very involved process with many aspects to work on. You may see our progress so far on the workstream’s [Confluence page](#).

Ethico-legal Framework and Patient Participation

By Eric Apondo, Andreas Bruns, Lisa Kaldowski

Within Task area B1 Eva Winkler, Eric Apondo and Andreas Bruns attend to ethical consideration, currently laying the groundwork for information material, interviews and discussions that will be conducted with a diverse set of German patients (cancer patients and patients with genetic and rare disorders), patient representatives, self-help group members and relatives to find out what their concerns and expectations are for GHGA. We hope to use these findings to involve these groups in governance within GHGA and develop a framework which measures our success. At the same time, we have reached out to a contact within Cancer Research UK, which has had experience with patient and public involvement strategies and are eager to learn from them! Feel free to point us in the direction of such groups or individuals...wherever they are!

B1.M2 is currently working towards such a balance between the need for broad consent and the need for appropriately informed consent. We aim to develop patient information and consent documents that could be used to obtain prospectively robust consent. Here, our ethics team is reliant on input from GHGA's technical workstreams on matters such as data flow and data governance. For this purpose, we have set up a Task Force (with an initial meeting planned for mid-July) whose main task is to determine GHGA's needs regarding these documents. Moreover, we have started a close collaboration with B1.M1 as well as GHGA's outreach workstreams in order to coordinate common interests and needs regarding patient engagement.

Fruzsina Molnár-Gábor, Lisa Kaldowski and Ameli Schenk are working on the legal questions surrounding GHGA, currently focusing on Legacy Consent. Data previously collected for scientific research and clinical studies can present a valuable source for scientific research conducted with the help of the GHGA infrastructure. However, the question remains whether such data can be processed lawfully. The legal team has been working on a tool to assess consent forms based on which data has been collected before the current EU data protection law has come into force (legacy consent). According to the GDPR, where processing is based on consent pursuant to its predecessor, the Data Protection Directive, it is not necessary for the data subject to give their consent again if the manner in which the consent has been given is in line with the conditions of the GDPR.

The legal team has checked consent forms developed under the Data Protection Directive and have identified key issues that might lead to a limitation of further data processing within the realm of GHGA based on these consents. Both issues around the adherence to the GDPR standard and the comparability of the scope of data processing covered by the legacy consent with that of the envisaged further processing have been taken into account. To guide the assessment of legacy consent as a legal basis for processing data in GHGA, the legal team has developed a decision tree which will now be presented to GHGA in various forums.

B2 and B3 collaboration on Metadata Implementation

by Anandhi Iyappan, Florian Kraus, and Deepak Unni

Task area B2 is overseen by Sven Nahnsen together with Martin Lablans, and aims at establishing best practice guidelines for the FAIRness of access-controlled data, as well as enabling the interoperability of GHGA and clinical data sources by developing concepts for harmonized, privacy-preserving record linkage. Standardization and harmonization at GHGA is driven by task area B3, and is overseen by Peer Bork and Joachim Schulze. The aim is to define, harmonize and standardize primary data and metadata within GHGA, as well as establishing best practice guidelines and monitoring requirements and standards for data submission.

In a collaborative effort by task area B2, B3 and the GHGA Core Team, we are working on the implementation of the metadata model for GHGA. In a first step, the ["Metadata Implementation Taskforce"](#) evaluates existing data portals and models in order to guide the creation and implementation of our metadata model. A preliminary version of the GHGA metadata model can be found in the respective [GitHub repository](#). The model is built using LinkML, a framework that provides a developer-friendly way to federate across different levels of semantic expressivity. This taskforce meets every Wednesday from 3 PM to 4 PM and is open to all GHGA Team members who want to contribute.

[Data Stewardship & First-Level User Support](#)

By Susanne Motameny

Task Area C1 is responsible for managing the data submission and data access processes at GHGA. For this purpose, one data steward will be appointed at each data hub. The six data stewards will form the GHGA helpdesk and support users with data submission and data access requests - they will be the face and voice of GHGA towards its users.

PIs from all GHGA data hubs are members of the TA C1: Ivo Buchhalter (Heidelberg) and Susanne Motameny (Cologne), who lead the task area, Andreas Dahl (Dresden), Julien Gagneur (Munich), Sven Nahnsen (Tübingen), and Philip Rosenstiel (Kiel). With Simon Parker (Heidelberg) and Christian Mertes (Munich), two data stewards are already recruited, the others will follow as the project progresses.

Since the start of the project, TA C1 has been active in many areas. As the collection of metadata is key to the GHGA submission process, we compiled a first version of a GHGA core metadata set together with TA B3. We also started a series of meetings with the NGS-CN sequencing centers that are co-located with the data hubs to include our major data submitters early on and collect their opinion about how data submission into GHGA can be supported by them. A subgroup of TA C1 meets regularly to define [SOPs](#) for the various processes that the data stewards will handle. Currently we focus on the details of the helpdesk structure and mode of operation. Our goal is to have a working helpdesk in time for the MVP being launched (expected in summer 2022 according to the overall GHGA roadmap). To make this happen, we depend on the work contributed by many other task areas and are open to further cooperation in this respect. For more details, check out the TA C1 page on [Confluence](#).

[Workflows workstream](#)

By Christian Mertes and Nick Smith

The workflow workstream, formerly known as task area C2, has been tasked with the implementation and benchmarking of the bioinformatic pipelines for GHGA to work alongside curated datasets compiled with the community's research interests in mind. This team is led by Julien Gagneur and Stephan Ossowski, and a growing team of capable and driven bioinformaticians: Christian Mertes, Christopher Hakkaart, Christian Schudoma, and Nicholas Smith.

In recent months the workflow workstream team has been present in many planning meetings and working to set a road map in order to achieve our goals. We have been holding regular meetings and coordinating internally on the best ways to achieve our goals. Recently we have been focusing on learning more about nextflow workflows, GA4GH APIs, and the FAIR principles extending into the context of workflows. In addition, we are reaching out to experts, especially to the nf-core community, to fill the gap in our expertise.

The detailed requirements of the workflows, e.g. input, output, and QC metrics, will be worked on over the coming months with prototypes coming later this year, and a usable

product ready on the first release of the GHGA Archive. During the prototyping phase, any suggestions from the community are welcome. We plan on abiding by the best practices and providing useful community features such as population-based variant frequency tables, count matrices, gVCF files for joint variant calling, standardized variant annotations, and thorough quality control metrics to ensure that the data is reliable from start to finish.

Outreach Strategy

by Ulrike Träger

Interaction with parties outside of GHGA are part of different task areas. Led by the Community Engagement and Community Development task areas (A1/A2), GHGAs outreach activities are being refocused and defined. An Outreach team, involving GHGA team members from A1/A2 but also from task areas Patient Participation (B1) and Community Data Curation and Processing Workflows (C2) has been formed. The team is meeting biweekly on Monday mornings. If you like to discuss an outreach idea, feel free to contact [Ulrike Träger](#). We happily take your input, ideas and initiative on board.

Currently, the team is working on an outreach strategy defining the different target groups GHGA want to reach. After establishing key messages we want to convey to each target group, we then decide on activities to be prioritised. Ongoing projects include the relaunch of the GHGA website including an external newsletter and blog, the planning of the GHGA Annual Meeting, a monthly lecture series starting in fall and workshops with different research communities to establish their needs and use cases. In our aim to expand GHGA outreach efforts to the general public and to lobby for genome research in Germany in general, we are also conceptualising a podcast and video series. Working with a PR firm, we hope to soon establish a style guide and information material for and on GHGA. These materials (flyers, slides, etc.) can be used by all GHGA members when talking about GHGA at a function or conference.

Project Management

By Jan Eufinger and Nicole Schatlowski

The GHGA Project Management team is coordinating all activities within and around GHGA. Besides many administrative tasks around budget and recruiting, this includes the engagement with the many other workstreams mentioned above, especially the legal working group. Together with the outreach team we are coordinating all GHGA meetings, especially the upcoming GHGA All-Hands Meeting on July 28 and the [GHGA Annual Meeting](#). Regular consultation with the GHGA Board of Directors is carried out via a weekly Jour fixe with the four directors.

A major activity was the establishment of the [GHGA reporting scheme](#) and the first collection of reports from all task areas. Thanks to all for the valuable input which we are currently reviewing. Based on this we will be working on a strategy for a revised GHGA work program for the next few years. In addition, together with many other GHGA members we are also actively engaging with other NFDI consortia and the newly founded NFDI e.V..

GHGA at the LifeTime 2.0 Conference

GHGA celebrated a first in June. The outreach team manned GHGA's first - virtual - booth at a conference. Following a promotional shout-out during Oliver Stegle's talk, the breakout room saw quite a few visitors on the second day. Researchers were keen to know more about what GHGA can offer and when data can be submitted. These interactions will become even more essential once GHGA is operational.



GHGA held a meta data workshop



Hosted by Susanne Motameny and Koray Kirli, invited external guest speakers and specialists from within GHGA presented different aspects around the usage of metadata and ontologies. Pier Luigi Buttigieg, Melanie Forche and Beatrix Tettmann (all from HMC) were joined by Mallory Freeberg from EGA to share their experience, use cases of ontologies and metadata models they work with. GHGA team members Koray Kirli and Deepak Unni presented work from their previous employment on the 4DN data portal and the monarch initiative, respectively.

The input and discussions around the relevance of these models for the design of GHGA's infrastructure will inform the strategy for the implementation of metadata models within GHGA. Notes from the meeting can be found on [Confluence](#).

GHGA - We Run The Genomes

Things got athletic in July. Ten members of the GHGA team took part in the 10th NCT Run - a charitable run against cancer! The virtual event took place over the span of three days from 02-04 of July. Collecting as many sponsored steps as possible, the team averaged 29,658 steps each! That's a whopping 23,8 kilometers each. Under the motto "Alone. Together.", most kilometers were run, walked or hiked alone - but the Heidelberg office had an official kick off run on Friday. Along the Neckar, the squad did 6,5 km together in the best weather.



Defining a Best Practice for Data Governance for Omics Research

In a recent publication GHGA members Fruzsina Molnar-Gabor, Lisa Kaldowski and Jan Korbel propose a code of conduct for omics research. The goal of GHGA in the NFDI is the creation of a technically secure and legally compatible national infrastructure for processing omics data. In so doing, data protection law challenges, e.g. the legal obligations of the data controller and processor have to be approached with application-oriented solutions. The solutions found provide potential for generalization and, therefore, could be implemented in a code of conduct pursuant to Art. 40 GDPR. Accordingly, the code could function as a Best Practice for data governance for omics research by adjusting, generalizing and

formalizing rules, in order to subsequently communicate them in a manner raising public awareness. The development of such best practice can take due respect of sector-specific data processing needs and aims, while coordinating these with data protection principles and the abstract-general requirements of applicable law. Furthermore, a code can serve as a basis not just to incorporate the established omics infrastructure into the national research data infrastructure, but to open it up on a European and international level and, thus, to promote cross-border omics research.

The paper will be available online (open access) from October onwards. The text has been circulated to registered users of the GHGA-all-hands list on the 21st of June 2021 and is available upon request from the corresponding author (Molnar-Gabor).

Full Paper Reference: Fruzsina Molnar-Gabor, Lisa Kaldowski and Jan Korbel. Code of Conduct for Omics Research: An instrument based on Art. 40 GDPR as an element of the data governance in the NFDI-GHGA consortium. In: Zeitschrift für Datenschutz, 6/2021, 313-319 (in German).

New Precision Oncology Approach Improves Therapeutic Interventions for Rare Cancer Patients

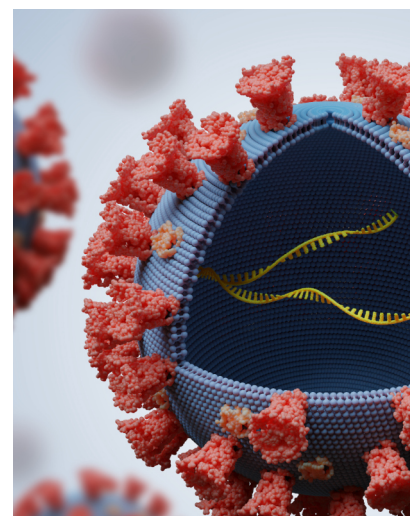


Comprehensive genomic and transcriptional analysis can enable personalised medicine and improve patient care. The ongoing NCT/DKFZ/DKTK [MASTER](#) study established a standardised precision oncology workflow for patients suffering from rare cancer types. The molecular and clinical results for the first 1310 patients have now been [published](#). Assessing 472 clinically relevant genes and 6 composite biomarkers, a molecular tumour board gave evidence-based recommendations for clinical care. For 86.9% of patients a new therapy could be suggested. When administered, the recommended therapies were effective, resulting in significantly improved overall response and disease control rates compared to previous therapies. For one third of patients, the progression-free survival rate increased.

This study was headed by GHGA members Stefan Fröhling and Hanno Glimm. In the future, GHGA will support studies such as the MASTER program by enabling access to reference datasets and genome analysis pipelines.

CoGDat-Development reaches milestone

The SARS-CoV-2 Genomics Data Platform (CoGDat) is an initiative launched by researchers across Germany. Central goals of CoGDat are the collection, analysis and publication of the sequencing raw data that are being generated in the context of the SARS-CoV-2 genomic surveillance efforts initiated by the German government in January 2021. GHGA contributes to the CoGDat effort both on a conceptual level in the data privacy domain as well as through the development of the data portal software DataMeta, which enables the submission, tracking and processing of the aforementioned data by contributing sequencing laboratories. Marking a big milestone, the first stable release of DataMeta was published and the CoGDat data protection plan received organization level approval.



SAVE THE DATE

GHGA ALL HANDS MEETING

28.07.2021
3 PM

[AGENDA](#)

GHGA ANNUAL MEETING

26-28. 10.2021

[DETAILS](#)

GA4GH 9TH PLENARY MEETING

28 - 29.09. 2020

[DETAILS](#)

JOBS IN GHGA

The **GHGA Office in Heidelberg** is looking to recruit four people to join the technical team:

- ▷ An experienced [DevOps Engineer/Cloud Specialist](#) to develop, deploy and maintain the cloud platform of the GHGA infrastructure. The person will also implement system security measures and build a hybrid cloud solution to span multiple cloud providers.
- ▷ A [Software Developer](#) to contribute to front and back end development of the GHGA software and data portal. The person will design and specify features and align them with international standards.
- ▷ A senior [Full Stack Developer](#) to coordinate and contribute to the design, implementation and operation of the GHGA core software architecture. The person will further participate in developing the federated EGA network.
- ▷ A [bioinformatician or computational biologist](#) to implement and maintain data processing and analysis workflows as part of GHGA. The person will develop and improve workflows and subsequently test/benchmark them with cloud deployment.

The **University of Cologne (RRZK)** is looking to hire two people:

- ▷ A [Technical Project Coordinator](#) to coordinate the installation and operation of data storage infrastructure at the GHGA data hubs as well as the harmonization of this infrastructure throughout the project.
 - ▷ An [Infrastructure DevOps developer](#) for the development, installation and operation of the storage infrastructure at the GHGA Cologne hub. The person will also implement the interfaces to the West German Genome Center (WGGC) sequencing platform in Cologne to streamlined data submission are also part of the tasks.
-

The **German National Cohort (NAKO)** is looking for a [Project Manager and Data Engineer](#) (focusing on Omics). Managing NAKOs tasks within the context of GHGA, the person will develop a concept for the storage and use of NAKO-OMICs data in the context of national and international consortia.

In addition, Jan Korbel's group at **EMBL** is soon advertising for a Data Manager & Steward / Cloud Data Manager.

If you are interested in any of these opportunities, want to know more or know someone interested, please contact the GHGA Office.

DID YOU LIKE THE GHGA NEWSLETTER?

Please send feedback or suggestions, and any events, jobs or news items for the next edition to [Ulrike Träger](#) or the [GHGA Office](#).