

Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanagerinnen und -managern

**Genomdaten FAIR und sicher teilen: Das Deutsche
Humangenom-Phänom Archiv (GHGA) als Baustein der
Nationalen Forschungsdateninfrastrukturⁱ**

Jan Eufinger

Jan Korbel

Eva Winkler

Oliver Kohlbacher

Oliver Stegle

2021

Zitiervorschlag

Eufinger, Jan et al. 2021. Genomdaten FAIR und sicher teilen: Das Deutsche Humangenom-Phänom Archiv (GHGA) als Baustein der Nationalen Forschungsdateninfrastruktur. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2021: S. 86-97. DOI: [10.17192.bfdm.2021.2.8349](https://doi.org/10.17192.bfdm.2021.2.8349).

Dieser Beitrag steht unter einer
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

ⁱJan Eufinger (ORCID: [0000-0002-3439-1674](https://orcid.org/0000-0002-3439-1674)), Jan Korbel (ORCID: [0000-0002-2798-3794](https://orcid.org/0000-0002-2798-3794)),
Eva Winkler (ORCID: [0000-0001-7460-0154](https://orcid.org/0000-0001-7460-0154)), Oliver Kohlbacher (ORCID: [0000-0003-1739-4598](https://orcid.org/0000-0003-1739-4598)), Oliver Stegle (ORCID: [0000-0002-8818-7193](https://orcid.org/0000-0002-8818-7193)).

1 Zusammenfassung

Menschliche Genomdaten und andere verwandte sogenannte Omics-Daten, die mithilfe moderner Sequenzierverfahren gewonnen werden, sind integraler Bestandteil der biomedizinischen Forschung. In Zukunft werden diese Daten auch die klinische Versorgung immer stärker prägen. Dabei muss das Bedürfnis, Daten offen und FAIR¹ für die Forschung nutzen zu können immer mit dem Schutz der Privatsphäre der Patient*innen ausbalanciert und gegeneinander abgewogen werden. Zugriff kann dabei nur unter Einhaltung der notwendigen technischen und organisatorischen Schutzmaßnahmen und für legitime Forschungszwecke gewährt werden. Auf europäischer Ebene gibt es für diesen Zweck bereits das Europäische Genom-Phänom-Archiv (EGA). Da die zentrale EGA Infrastruktur die spezifischen nationalen Regelungen zum Datenschutz nur ungenügend abbilden kann, bereitet das EGA aktuell eine Umwandlung in eine föderierte Infrastruktur aus nationalen Knoten („föderiertes EGA“) vor.

Ziel des NFDI-Konsortiums GHGA (**G**erman **H**uman **G**enome-**P**henome **A**rchive) ist der Aufbau eines Genomarchivs als nationaler EGA-Knoten für die sichere Speicherung, den Zugriff und die Analyse menschlicher Omics-Daten (z. B. Genome, Transkriptome) in einem einheitlichen ethisch-rechtlichen Rahmen. GHGA wird dabei auch die Wünsche der Forschungsgemeinde nach effizienten, benutzerfreundlichen Analysen im großen Maßstab und zur Replikation von Ergebnissen auf anderen Kohorten berücksichtigen. GHGA setzt auf existierenden nationalen Omics-Datenlieferanten und deren IT-Infrastrukturen auf, um eine harmonisierte, interoperable Infrastruktur zu schaffen. Ziel ist es, Forschende in Deutschland in die Lage zu versetzen, humane Genomdaten rechtssicher entsprechend der FAIR-Richtlinien austauschen zu können und dabei internationale Standards zum Datenaustausch stärker mitzugestalten. GHGA ist zudem eingebunden in flankierende internationale Forschungsnetzwerke wie etwa die europäische *Beyond One Million Genomes Initiative*.

2 Einleitung

Es besteht kein Zweifel, dass die technologische Revolution bei der Entwicklung von DNA-Sequenziertechnologien, die etwa seit der Jahrtausendwende eingesetzt hat, zu einer fundamentalen Veränderung der biomedizinischen Forschung geführt hat. Die Möglichkeit, das gesamte Genom eines Menschen innerhalb weniger Stunden oder Tage zu immer geringer werdenden Kosten zu entschlüsseln, hat eine Vielzahl von neuen Erkenntnissen über die genetische Grundlage von Gesundheit und die Ursache von Erkrankungen erbracht. Die auch als Genomik bezeichnete Genomforschung, die

¹FAIR: Findable, Accessible, Interoperable, Reusable, siehe Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, u. a. *The FAIR Guiding Principles for Scientific Data Management and Stewardship*. Scientific Data 3, Nr. 1 (15. März 2016): 160018. <https://doi.org/10.1038/sdata.2016.18>.

auf der Sequenzierung der Erbinformation im Zellkern basiert, wird dabei häufig kombiniert mit der Erstellung umfassender Profile von anderen Parametern der Zellen. Diese werden oft unter der Bezeichnung „Omics-Technologien“ zusammengefasst². Dazu gehören z. B. die möglichst vollständige Erfassung aller Gen-Transkripte einer Zelle zu einem bestimmten Zeitpunkt (Transkriptom) oder die Erfassung von epigenetischen Änderungen von Zellen (Epigenom). Neben diesen auf DNA-Sequenzieretechnologien basierenden Methoden wird auch die parallele Messung anderer Biomoleküle in Zellen den Omics-Technologien zugerechnet, etwa die Erfassung von Proteinen oder Lipiden von Zellen eines bestimmten Typs oder Zustandes. Diese Ansätze werden analog als Proteomik bzw. Lipidomik bezeichnet.

Mit den immer breiter und kostengünstiger verfügbaren Technologien zur Genomsequenzierung³ haben sich diese zu einer Standardmethode der biomedizinischen Forschung entwickelt. In zahlreichen Labors werden, aktuell noch meist zu Forschungszwecken, in Zukunft aber immer mehr auch für diagnostische Ansätze, humane Omics-Daten erhoben. Sie stammen meist von Patient*innen, aber immer häufiger auch von gesunden Proband*innen im Rahmen von Kohortenstudien. Die Forschungsfragen, die mit Omics-Technologien untersucht werden, sind dabei nahezu unbegrenzt. Beispiele sind die Auffindung von angeborenen genetischen Varianten, die zur Manifestation einer seltenen Erkrankung führen, die Analyse von genetischen Veränderungen in Tumorzellen, die für die Entstehung oder den besonders bösartigen Verlauf von Krebserkrankungen ursächlich sind, bis hin zur Auffindung von Korrelationen zwischen genetischen oder epigenetischen Merkmalen mit der Entstehung von neurodegenerativen oder Herz-Kreislauf-Erkrankungen. Gemein ist diesen Ansätzen, dass die produzierten Daten in der Regel nicht nur für die initialen Forschungsfragen wertvoll sind, sondern durch Kombination mit anderen Datentypen oder durch Validierung in gleichartigen Datentypen aus anderen Datenbeständen erst ihr volles Potential für die Forschung entfalten. Ein breites Teilen dieser biomedizinischen Forschungsdaten ist daher unbedingt geboten um den bestmöglichen Nutzen für die Forschung und die spätere klinische Anwendung zu erreichen.

2.1 Besondere Herausforderungen an den Datenschutz und die Aufklärung

Demgegenüber steht jedoch die besonders schützenswerte Natur von humanen Genomdaten: diese fallen unter die Kategorie der personenbezogenen Daten, da sie

²Nationale Akademie der Wissenschaften Leopoldina. *Zukunftsreport Wissenschaft. Lebenswissenschaften im Umbruch – Herausforderungen der Omics-Technologien für Deutschlands Infrastrukturen in Forschung und Lehre*. (Deutsche Akademie der Naturforscher Leopoldina e.V., 2014). <https://www.leopoldina.org/publikationen/detailansicht/publication/lebenswissenschaften-im-umbruch-2014/>, (zuletzt abgerufen: 30.04.2021).

³The Cost of Sequencing a Human Genome. Genome.gov. <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>, (zuletzt abgerufen: 07. 01. 2021).

sensible Informationen über die betroffenen Personen beinhalten können. Im Gegensatz zu anderen Forschungsdaten aus der Biomedizin ist die Möglichkeit der Re-Identifizierbarkeit nicht durch übliche Verfahren wie etwa eine Pseudonymisierung oder eine Einschränkung der in Suchanfragen möglichen Abfragen zu verhindern⁴. Das Risiko der Re-Identifizierbarkeit gilt dabei nicht nur für Genomdaten, sondern auch für viele andere Omics-Datentypen. Zudem sind genetische Informationen nicht nur für die Person relevant, von der sie stammen, sondern können in bestimmten Konstellationen auch über blutsverwandte Angehörige aussagekräftig sein. Mit der Eigenschaft als personenbezogenen Daten sind für den Austausch von humanen Omics-Daten daher zwingend die Vorgaben der europäischen Datenschutz-Grundverordnung (DSGVO) beziehungsweise ihrer nationalen Umsetzungen zu erfüllen. Ein Zugriff auf, beziehungsweise der Austausch von, humanen Omics-Daten kann daher nur über vertraglich geregelte kontrollierte Datenzugänge, sogenannten „controlled access“⁵, ermöglicht werden. So wird sichergestellt, dass alle durch die DSGVO vorgegebenen Auflagen erfüllt werden. Diese Vorgabe bedeutet für die genomische Gesundheitsforschung eine große Herausforderung, da zum einen die DSGVO innerhalb der EU-Mitgliedsstaaten unterschiedlich interpretiert und umgesetzt wird, zum anderen ein Austausch mit nicht EU-Staaten daran geknüpft ist, dass mindestens die Ansprüche der DSGVO im Empfängerland erfüllt werden.⁶

Eine zweite Herausforderung ist die adäquate Aufklärung derer, von denen die Daten stammen. Der herkömmliche ethische Standard für die informierte Zustimmung war bisher eine spezifische Information über die Ziele, Methoden, Vorteile und Risiken einer Studie. Das Sammeln von Patient*innendaten für zukünftige Forschungsprojekte mit noch nicht festgelegten Zielen macht daher neue Wege der Aufklärung, Information und Kontrolle notwendig. Die World Medical Association hat auch aus diesem Grund in ihrer Überarbeitung der Deklaration von Helsinki im Jahr 2013 die Bestimmungen zur Forschung mit personenbeziehbaren Daten geändert. Die Erklärung erfordert weiterhin eine Einverständniserklärung zur Sammlung, Lagerung und Wiederverwendung. Jedoch ist es für eine Nachnutzung der Daten in künftigen Studien nun nicht mehr zwingend notwendig, diese in der Einverständniserklärung konkret zu benennen.⁷ Ebenso erkannte 2016 der Rat für Internationale Organisationen der medizinischen Wissenschaft (CIOMS) gemeinsam mit der WHO eine breite Zustimmung als

⁴Siehe z. B. Thenen, Nora von, Erman Ayday, und A Ercument Cicek. *Re-Identification of Individuals in Genomic Data-Sharing Beacons via Allele Inference*. *Bioinformatics* 35, Nr. 3 (February 1, 2019): 365–71. <https://doi.org/10.1093/bioinformatics/bty643>.

⁵Shabani, Mahsa, Stephanie O. M. Dyke, Yann Joly, und Pascal Borry. *Controlled Access under Review: Improving the Governance of Genomic Data Access*. *PLOS Biology* 13, Nr. 12 (31. Dezember 2015): e1002339. <https://doi.org/10.1371/journal.pbio.1002339>.

⁶Molnár-Gábor, Fruzsina, and Jan O Korbel. *Genomic Data Sharing in Europe Is Stumbling—Could a Code of Conduct Prevent Its Fall?* *EMBO Molecular Medicine* 12, Nr. 3 (March 6, 2020). <https://doi.org/10.15252/emmm.201911421>.

⁷World Medical Association. *World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects*. *JAMA* 310, Nr. 20 (27. November 2013): 2191–94, <https://doi.org/10.1001/jama.2013.281053>.

angemessen für solche Studien an⁸. Eine breite Zustimmung ist zu einer umsetzbaren Lösung geworden, um die Studienteilnehmenden generell über den Umfang der Forschungsprojekte zu informieren, die von ihren Daten gespeist werden, sowie über die Forschungspartner und Länder, die künftig am Datenaustausch beteiligt sind. Eine breite Zustimmung verlagert jedoch die Verantwortung für die Kontrolle und Überwachung der Gültigkeit dieser Zustimmung auf die Forschungseinrichtung, Forschungsethikkommissionen, Datenzugriffskomitees und die Forscher*innen selbst. Sie fungieren als Treuhänder*innen für die Interessen der Patienten*innen. Die Erfüllung dieser Treuhandpflicht kann auf lokaler oder nationaler Ebene funktionieren, wird jedoch für den globalen Datenaustausch ein ehrgeiziges Unterfangen. Daher ist die Vernetzung von GHGA mit den europäischen Initiativen und dem gerade entstehenden Regelwerk des europäischen Datenraums auch mit Blick auf die Governance von GHGA notwendig.

2.2 GHGA als Bestandteil eines föderierten europäischen Archivs für Genomdaten

Insbesondere die national unterschiedlichen Auslegungen der DSGVO innerhalb der EU stellt das existierende europäische Humangenom-Archiv (EGA)⁹, welches gemeinsam vom European Bioinformatics Institute (EBI) in Hinxton (England) und dem Centre for Genomic Regulation (CRG) in Barcelona (Spanien) betrieben wird, vor große Herausforderungen. Um diese zu bewältigen, wurde kürzlich beschlossen, EGA in ein neues Modell aus föderierten nationalen Knoten zu überführen.^{10,11} Im Rahmen dieser föderierten Struktur sollen nationale EGA-Knoten entstehen, die Daten archivieren und in kontrollierter Weise zur Verfügung stellen. Dies ermöglicht den Umgang mit Daten, die etwa aufgrund von Vorgaben der Einverständniserklärung nur lokal archiviert werden dürfen. Die nationalen Knoten tauschen mit dem zentralen EGA lediglich Metadaten über die Studien und Datensätze aus, die keine Rückschlüsse auf die betroffenen Individuen zulassen. So wird eine Auffindbarkeit der Daten über alle Knoten hinweg gewährleistet. Nach der Identifikation von relevanten Studien können Forschende Anfragen zur Nutzung der aufgefundenen Datensätze stellen, die vom zentralen EGA an die datenhaltenden nationalen Knoten weitergegeben werden (s. [Abbildung 1](#)). Die Entscheidung über die Gewährung des Zugriffs liegt bei den jeweils mit den Datensätzen

⁸Council for International Organizations of Medical Sciences (CIOMS). *International Ethical Guidelines for Health-Related Research Involving Humans*. 4th edition, 2016. <https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf>, (zuletzt abgerufen: 30.04.2021).

⁹Ilkka Lappalainen et al. *The European Genome-Phenome Archive of Human Data Consented for Biomedical Research*. *Nature Genetics* 47, Nr. 7 (Juli 2015): 692–95, <https://doi.org/10.1038/ng.3312>.

¹⁰*Federated EGA - EGA European Genome-Phenome Archive*. <https://ega-archive.org/federated>, (zuletzt abgerufen: 18.01.2021).

¹¹Saunders, Gary et al. *Leveraging European Infrastructures to Access 1 Million Human Genomes by 2022*. *Nature Reviews. Genetics* 20, Nr. 11 (November 2019): 693–701, <https://doi.org/10.1038/s41576-019-0156-9>.

verbundenen Datenzugriffskomitees (*Data Access Committees, DACs*). Die DACs sind in der Regel an den Institutionen angesiedelt, die die Datenverantwortung tragen. Das DAC entscheidet unter Berücksichtigung der zur Datenerhebung erteilten Einverständniserklärung, ob die Daten an den Anfragenden weitergegeben werden können und ist für die notwendigen vertraglichen Regelungen des Datenzugriffs und -austauschs zuständig.

Mit GHGA beteiligt sich Deutschland als eines der ersten EU-Länder an diesem Modell und entwickelt gemeinsam mit den Vertreter*innen des zentralen EGA und anderen nationalen Knoten den Rahmen für die Inbetriebnahme des föderierten EGA.

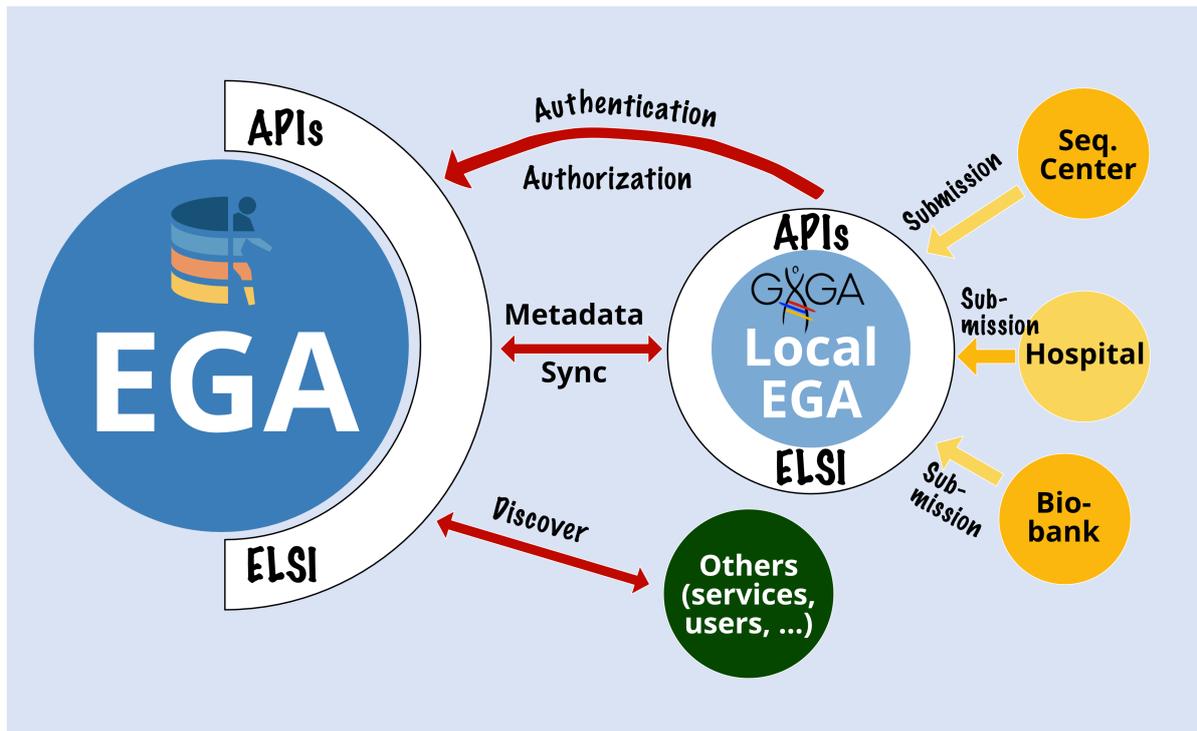


Abbildung 1: **Modell zum Datenaustausch im föderierten European Genome-Phenome Archive (EGA)**. Daten, zum Beispiel von Sequenzierzentren, Krankenhäusern oder Biobanken werden in lokalen Knoten, unter Einhaltung der nationalen Datenschutzbestimmungen, archiviert (rechte Seite). Mit der zentralen EGA-Instanz (im Bild links) werden über Programmierschnittstellen (API) lediglich Metadaten ausgetauscht, die eine Auffindbarkeit der Daten ermöglichen. Darüber werden auch die ethischen und legalen Aspekte des Datenaustauschs (ELSI) abgeklärt (Quelle: European Genome-Phenome Archive (EGA))

3 Struktur und Arbeitsweise von GHGA

Im Rahmen des seit Oktober 2020 von der DFG geförderten NFDI-Konsortiums GHGA haben sich 38 Antragsteller*innen aus 16 Institutionen zusammengetan. Ergänzt werden sie durch weitere Teilnehmer*innen aus fünf nicht direkt geförderten Institutionen. GHGA repräsentiert damit die große Breite der medizinischen Genomforschung in Deutschland. Unter anderem beteiligt an GHGA sind die vier DFG Kompetenzzentren für Hochdurchsatzsequenzierung¹² in Dresden, Kiel, Köln und Tübingen sowie die Sequenzierzentren in Heidelberg (DKFZ) und München (HGMU/TUM). Über die Einbindung von Zentren für Hochleistungsrechnen in Köln, Tübingen, Dresden und München sowie dem kürzlich gegründeten CISPA Helmholtz-Zentrum für Informationssicherheit in Saarbrücken wird essentielle IT-Expertise in das Konsortium eingebracht.

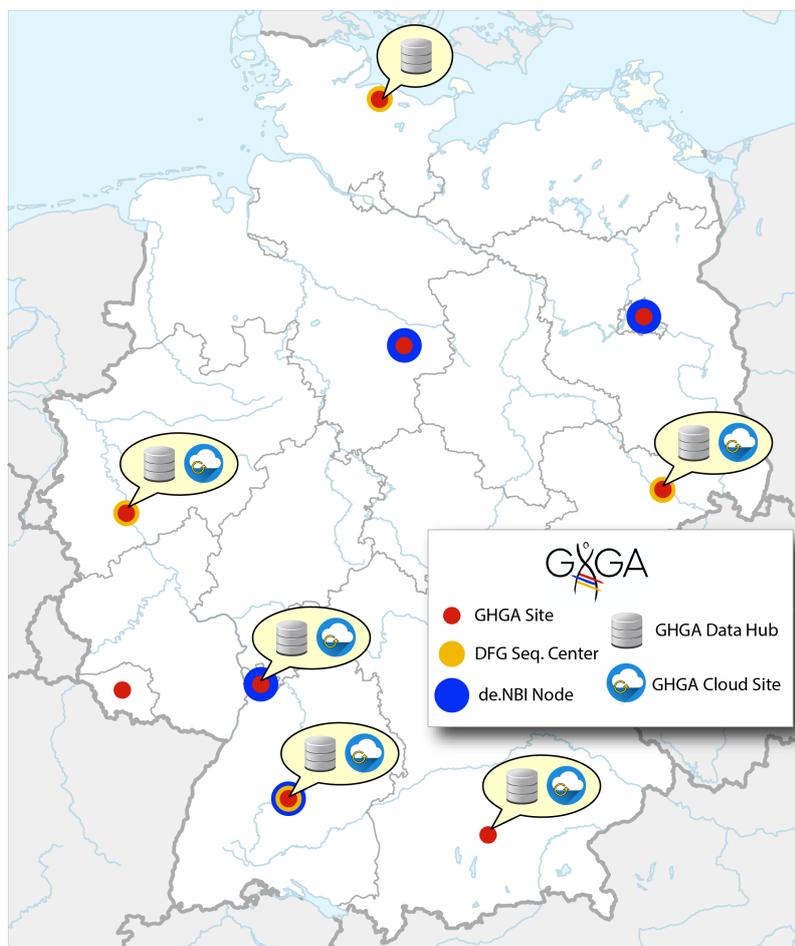


Abbildung 2: Übersicht der an GHGA beteiligten Standorten und Funktionalitäten der Standorte

¹²Deutsche Forschungsgemeinschaft. *Hochdurchsatzsequenzierung: DFG richtet vier Kompetenzzentren ein*. 16. März 2018. https://www.dfg.de/service/presse/pressemitteilungen/2018/pressemitteilung_nr_06, (zuletzt abgerufen: 30.04.2021).

Das GHGA Netzwerk ist dabei mit zahlreichen anderen Initiativen verknüpft, die im Rahmen der medizinischen Genomforschung tätig sind. So sind unter anderem vier Standorte auch Knoten im Rahmen des Deutschen Netzwerk für Bioinformatik-Infrastruktur – de.NBI¹³, welches wiederum Deutschland im Rahmen des europäischen Bioinformatiknetzwerks ELIXIR¹⁴ vertritt.

Thematisch sind die Arbeiten in GHGA in fünf Task Areas aufgegliedert (s. auch [Abbildung 3](#)):

- Über die *Task Area A - Communities* wird sichergestellt, dass GHGA entlang des Bedarfs der unterschiedlichen Forschungsgemeinden entwickelt wird. Dazu engagieren sich in GHGA zahlreiche Forschende aus der Krebsforschung und der Erforschung von seltenen Erkrankungen, den beiden initialen Forschungsgemeinden in GHGA. In beiden Forschungsgebieten sind Omics-Ansätze weit verbreitet. Durch gezielte Interaktion mit diesen und weiteren Forschungsgemeinden, etwa durch spezifische Umfragen, Interaktion mit entsprechenden Fachgemeinschaften und Teilnahme auf Fachkongressen soll sichergestellt werden, dass die zu entwickelnden Dienste dem Bedarf entsprechen. Eine Aufnahme von weiteren Forschungsthemen ist vorgesehen. GHGA wird außerdem Trainingsmaßnahmen für die Nutzung der entwickelten Dienste anbieten.
- In *Task Area B - Concepts* werden für den Datenaustausch essentielle Konzepte entwickelt. Diese umfassen zum einen die Entwicklung eines ethisch-legalen Rahmens unter besonderer Berücksichtigung von Patient*inneninteressen. Hierzu ist der enge Austausch mit anderen, thematisch benachbarten NFDI-Konsortien wie etwa NFDI4Health¹⁵ vorgesehen. Insbesondere im Bereich von ethischen und rechtlichen Fragestellungen beim Forschungsdatenmanagement plant das Konsortium, im Rahmen der NFDI gemeinsam Positionen und Strategien zu entwickeln. Zum anderen sollen Konzepte für die Anwendung der FAIR-Prinzipien und für die Harmonisierung und Standardisierung entwickelt werden. Hierbei wird GHGA eng mit existierenden Initiativen in diesem Bereich zusammenarbeiten. Auf nationaler Ebene ist dies zum Beispiel das de.NBI-Netzwerk und die Medizininformatik-Initiative¹⁶, im internationalen Bereich ist hier neben den europäischen Initiativen wie ELIXIR und EGA insbesondere die Global Alliance for Genomics and Health (GA4GH) zu nennen.
- *Task Area C - Data* befasst sich mit der Entwicklung der eigentlichen GHGA-Plattform, dem Aufbau einer Struktur für den Support von Datenanbietern über Data Stewards und der Auswahl von kuriierten Datensätzen und Workflows für den Einsatz in den Forschungs-Communities.
- In der vierten *Task Area D - Operations* werden schließlich die notwendigen Strukturen für einen Betrieb sowohl der zentralen wie auch der dezentralen Dienste an den einzelnen GHGA Data Hubs geschaffen.

¹³<https://www.denbi.de>.

¹⁴<https://elixir-europe.org/>.

¹⁵<https://www.nfdi4health.de/>.

¹⁶<https://www.medizininformatik-initiative.de>.

- Koordiniert wird das Projekt dabei über *Task Area E - Management*, in der neben dem zentralen Projektmanagement auch ein Modul für die Vergaben von flexiblen Mittel für die Ermöglichung von sogenannten „*Innovation and Implementation Projects*“ angesiedelt ist.

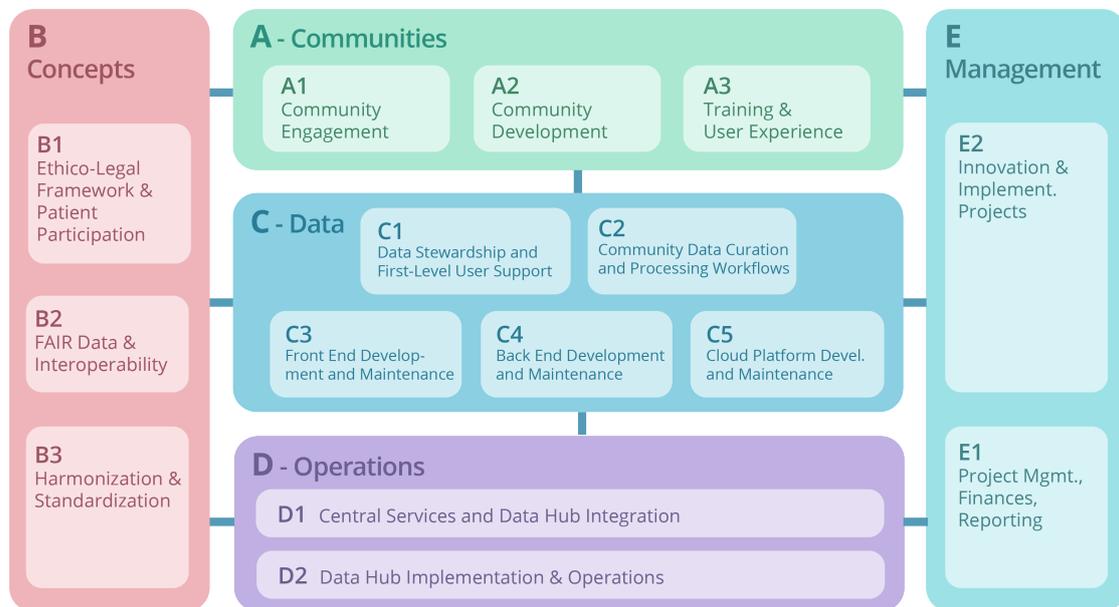


Abbildung 3: **Struktur der Task Areas des GHGA Konsortiums**

4 Kernziele von GHGA als Bestandteil der NFDI

4.1 Humane Omics-Daten FAIR teilen

Wie oben erwähnt, stellen die Anforderungen des Datenschutzes hohe Hürden für den Austausch von humanen Omics-Daten dar. Ein wesentliches Element von GHGA wird daher sein, legale, organisatorische und technische Konzepte zu entwickeln, wie Genomdaten möglichst gewinnbringend entlang der FAIR-Prinzipien für die Forschung eingesetzt werden können, ohne dabei das Schutzbedürfnis der sensiblen Daten zu kompromittieren. Hierzu beteiligen sich an GHGA ein Gruppe von Expert*innen, die sich seit langem mit solchen Fragen beschäftigen und bereits zahlreiche Konzepte, etwa für die Formulierung von Patienten-Einwilligungserklärungen für die Genomfor-

schung¹⁷, ausgearbeitet haben. GHGA interagiert dabei eng mit anderen Konsortien in diesem Bereich wie etwa der Medizininformatik-Initiative¹⁸ oder dem Netzwerk des Nationalen Centrum für Tumorerkrankungen (NCT)¹⁹ sowie anderen NFDI-Konsortien. Durch eine enge Interaktion zwischen den Expert*innen für ethische und legale Fragen und den technischen Gruppen in GHGA wird sichergestellt, dass die Konzepte, etwa zur technischen Unterstützung von Data Access Committees, im Rahmen der GHGA-Plattform standortübergreifend technisch unterstützt werden.

4.2 Aufbau einer föderierten Dateninfrastruktur für Genomdaten

Kern der GHGA-Infrastruktur wird der Aufbau einer Föderation zwischen den sechs an GHGA beteiligten Data Hubs (s. [Abbildung 2](#)) sein. Diese werden ihre bestehenden und künftigen Datensätze für eine Aufnahme in GHGA bereitstellen. Durch die Einhaltung von gemeinsamen, z. B. in der Global Alliance for Genomics and Health (GA4GH)²⁰ entwickelten Standards für Forschungs- und Metadaten und den Einsatz von interoperablen Schnittstellen sollen Datensätze über Standorte hinweg auffindbar und später auch analysierbar werden. Wie oben erläutert, bildet diese nationale Föderation dann die Grundlage für die Schaffung der deutschen Beteiligung an dem geplanten föderierten europäischen Genomarchiv EGA (s. [Abbildung 1](#)).

GHGA wird dabei so konzipiert, dass ein Austausch von Daten auf Basis des deutschen Datenschutzes ermöglicht wird. Anwendende werden kontinuierlich unterstützt, um die notwendigen Schritte für den Datenupload, die Suche nach Datensätzen, die Beantragung des Zugriffs und den Zugriff auf Datensätze so einfach wie möglich zu gestalten. Die Funktionalität in der ersten Ausbaustufe von GHGA wird sich dabei auf die DSGVO-konforme Archivierung und Zurverfügungstellung von Daten fokussieren. Die Implementierung von weiteren Funktionalitäten ist fester Bestandteil der Roadmap von GHGA: so sollen themenspezifische Datenportale bereitgestellt und in die GHGA-Plattform integriert werden. Diese sollen unter anderem kuratierte Referenzdatensätze und Analysetools zur Verfügung stellen. Hierbei wird GHGA eng mit den erwähnten Forschungsgemeinden zusammenarbeiten. Damit soll ein breiterer Zugang zur Analyse und Weiterverarbeitung von großen Omics-Daten ermöglicht werden. Schließlich soll der Wert von Forschungsdaten durch die Integration mehrerer Omics-Modalitäten und der Verknüpfung von Omics-Daten mit Phänotyp-Daten erhöht werden.

¹⁷Projektgruppe EURAT. *Eckpunkte für eine Heidelberger Praxis der Ganzgenomsequenzierung*. 2. Auflage. Universität Heidelberg, November 2015. https://www.uni-heidelberg.de/md/totalsequenzierung/mk_eurat_journal_d_2016_web.pdf, (zuletzt abgerufen: 30.04.2021).

¹⁸<https://www.medizininformatik-initiative.de/>.

¹⁹<https://www.bmbf.de/de/nationales-centrum-fuer-tumorerkrankungen-nct-12580.html>.

²⁰<https://www.ga4gh.org/>.

4.3 Stärkung der deutschen Genomforschung durch Bündelung der verfügbaren Ressourcen

GHGA strebt an, durch die Bündelung von Omics-Datensätzen in einer zentralen Plattform die Bedeutung der deutschen Omics-Forschung in der internationalen Forschungslandschaft zu erhöhen. Forschende aus Deutschland sind seit Langem an internationalen Aktivitäten in führenden Rollen beteiligt, als Beispiele können z. B. das International Cancer Genome Consortium (ICGC)²¹ oder die PCAWG-Initiative²² gelten. Allerdings ist diesen Forschungsaktivitäten bisher keine nationale Infrastruktur beigegeben, die die in Deutschland in großer Zahl erzeugten Genomdaten zentral sammelt und - in datenschutzkonformer Form - zur Verfügung stellt. Demgegenüber stehen erfolgreiche Ansätze anderer Länder²³ wie zum Beispiel das britische 100,000 Genomes Project²⁴ oder die finnische nationale Genomstrategie Finngen²⁵, die in nationalen Initiativen große genomische Datensätze erzeugen und zentral integrieren. Durch die Einrichtung von GHGA soll es Forschenden in Deutschland leichter möglich sein, vorhandene und zukünftige Datensätze effizienter in internationale Kollaborationen einzubringen und so besser zur internationalen Genomforschung beizutragen. Hierbei wird auch die kürzlich von Bundesgesundheitsminister Jens Spahn angekündigte deutsche Genominitiative genomDE²⁶ eine entscheidende Rolle spielen, die ein wesentlicher Teil des deutschen Beitrags zur europäischen 1+Million Genomes Initiative (1+MG) ist²⁷. GHGA-Forschende wurden in die Entwicklung des genomDE-Konzeptes eingebunden, und möchten durch eine Integration der Daten aus Deutschland in das EGA eine Nachnutzbarkeit der Genomdaten für Forschungszwecke sicherstellen. Neben diesen europäischen Initiativen wird sich GHGA auch im Rahmen der Global Alliance for Genomics and Health (GA4GH)²⁸ bei der Etablierung von globalen Standards für die Genomforschung einbringen.

²¹Hudson (Chairperson), Thomas J. et al. *International Network of Cancer Genome Projects*. Nature 464, Nr. 7291 (April 2010): 993–98. <https://doi.org/10.1038/nature08987>.

²²Campbell, Peter J. et al. *Pan-Cancer Analysis of Whole Genomes*. Nature 578, Nr. 7793 (Februar 2020): 82–93. <https://doi.org/10.1038/s41586-020-1969-6>.

²³Siehe: Stark, Zornitza et al. *Integrating Genomics into Healthcare: A Global Responsibility*. The American Journal of Human Genetics 104, Nr. 1 (3. Januar 2019): 13–20. <https://doi.org/10.1016/j.ajhg.2018.11.014>.

²⁴Genomics England. *The National Genomics Research and Healthcare Knowledgebase*. figshare, 21. Dezember 2017. <https://doi.org/10.6084/m9.figshare.4530893.v5>.

²⁵<https://www.finngen.fi/en/about>.

²⁶Bundesgesundheitsministerium. *Die deutsche Genom-Initiative - genomDE*. <https://www.bundesgesundheitsministerium.de/themen/gesundheitswesen/personalisierte-medizin/genomde-de.html>, (zuletzt abgerufen: 18. 01. 2021).

²⁷Bundesministerium für Bildung und Forschung - BMBF. *Deutschland tritt Genomprojekt der EU bei - BMBF*, 16. Januar 2020. <https://www.bmbf.de/de/deutschland-tritt-genomprojekt-der-eu-bei-10676.html>, (zuletzt abgerufen: 30.04.2021).

²⁸<https://www.ga4gh.org/>.

5 Auf dem Weg: Aufbau einer neuen nationalen Infrastruktur im europäischen Kontext

Mit dem Beginn der Förderung im Oktober 2020 wurden die Arbeiten zur Entwicklung von GHGA aufgenommen. Ähnlich wie in anderen IT-lastigen Forschungsprojekten steht zunächst der Aufbau eines standortübergreifenden Entwicklungsteams im Vordergrund. Aufgrund der aktuell schwierigen Rekrutierungslage für entsprechend qualifizierte IT-Expert*innen stellt dies eine erhebliche Herausforderung dar.

Gleichzeitig mit der Etablierung einer deutschen Föderation der sechs aktuell geplanten GHGA-Data Hubs muss die Einbindung in die europäischen Aktivitäten rund um das föderierte EGA und die 1+MG-Initiative entwickelt werden. Daher stehen hier zunächst weitreichende Abstimmungen etwa über gemeinsame Metadaten-Standards und Schnittstellen an. Diese Einbindung wird es ermöglichen, in GHGA entwickelte Konzepte zum Datenschutz und zum Austausch von Daten auf europäischer Ebene einzubringen und zu diskutieren. Hierdurch wird es möglich sein, aktuelle Hindernisse in der medizinischen Genomforschung in Zukunft immer besser aus dem Weg zu räumen.

Förderhinweis:

Das Forschungsprojekt GHGA – Deutsches Humangenom-Phänomarchiv (NFDI 1/1) wird gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Nationalen Forschungsdateninfrastruktur NFDI.