Human Cell Atlas

Ethics and Data Governance Document

No	Version	Date	Description
1	1.0 [external]	Sept. 29, 2022	First version
2	2.0 [external]	October 18, 2023	Amendments to include controlled access

This document was drafted by the Centre of Genomics and Policy (McGill University), and regularly discussed with the Human Cell Atlas Ethics Working Group (EWG) (https://www.humancellatlas.org/learn-more/working-groups/).

TABLE OF CONTENTS

Acronyms

API: Application Programming Interface

DACO: Data Access Compliance Office

DAC: Data Access Committee

DCP: Data Coordination Platform

EWG: Ethics Working Group

GDPR: General Data Protection Regulation

HCA: Human Cell Atlas

OC: Organizing Committee

Definitions

Data Contributor: an individual, institution or research team uploading datasets to the HCA DCP. (In some cases, the individual, institution or research team uploading datasets may not be the entity having collected tissue samples from participants. Collaborative scenarios are envisaged in Section 3.2).

Data User: an individual or organization authorized to access and use HCA Research Data from the DCP.

Data coordination platform (DCP): the open-source platform that hosts HCA Research Data. The DCP is developed and managed by Human Cell Atlas Incorporated (HCA Inc.), in collaboration with scientific and technical experts from numerous institutions. The host institutions that provide technical and scientific support to the HCA DCP include: the Broad Institute (U.S.), the Chan Zuckerberg Initiative (U.S.), the European Molecular Biology Laboratory - European Bioinformatics Institute (an Intergovernmental Organisation), and the University of California, Santa Cruz (U.S.). The DCP cloud server is hosted in the USA, with commercial cloud providers. Data hosted on the DCP may also be mirrored in multiple cloud servers.

Participant: the individual whose tissue samples, data and metadata will be or have been collected, held, used or shared. Participants could include living research participants, deceased individuals (e.g. organ donors, research participants), individuals having provided tissue samples to commercial vendors, or donors of other biological materials (e.g. embryos, fetal tissue).

Metadata: is information accompanying and annotating certain types of Research Data. A current HCA "metadata dictionary" is found on the HCA website[1]. This includes information about

- Biomaterial from which the cells were taken (e.g. organ type, disease state, information about the participant such as ethnicity, sex, and age)
- The Project that the sample came from (e.g. contributor name and lab details, funding sources for the project)
- The Protocols that were applied to generate the sample (e.g. collection method, library construction details)
- The analysis Process applied to the sample (e.g. institution where processing took place and researcher who performed the processing) and
- Information about the Files themselves (e.g. file name or description)

Open access data: means all transcriptomic data and metadata that can be shared under an open-access tier.

[1] Available online at: https://data.humancellatlas.org/metadata

Research Data: data that is derived from samples collected from Participants. This data is contributed to the DCP by Data Contributors. Research data may include, but is not limited to, the following:

- *Single-cell molecular profiles*: this will encompass short-read DNA sequence information, focused particularly on reads that measure the expression ("activity") of all genes in a cell, derived from detecting the RNA molecules within a cell;
- *Count matrices* summarising the above data (e.g., gene-cell matrices, where each row represents an individual gene and each column represents a single cell sampled in the relevant experiment)
- *Metadata* about the tissue of origin, high-level information about the participant (e.g. sex, age, ethnicity and location); for some individuals (e.g., from disease cohorts) non-identifying clinical information may also be provided;
- *Images of tissue sections***,** showing the spatial layouts of cells within the section accompanied with information about genes expressed in each cell.

The types of data which can be included in the DCP may be updated, as HCA evolves.

1. Overview

1.1. Background/rationale

Traditionally, scientists have classified cells by their structure, functions, location, and, more recently, molecular profiles. Surprisingly, however, characterization of cell types and states has remained limited. With the rise of new technologies to profile DNA and proteins in single cells, as well as a combination of DNA, RNA and proteins in the same cell, additional layers of information can now be provided. Furthermore, new spatial analysis techniques allow for high-resolution analysis of large tissues in two (2D) or three (3D) dimensions. In combination with advances in computational algorithms that can help to determine cell types, states, transitionals and locations, these advances allow for increasing scale and resolution to enable understanding of human cells and molecular states within tissues and systems[2].

1.2. Objectives of the HCA

Launched in October 2016, the HCA Consortium is an international, collaborative effort that aims to create comprehensive reference maps of all human cells—the fundamental units of life—as a basis for both understanding human health and diagnosing, monitoring, and treating disease. HCA aims to achieve this by defining all human cells in terms of their distinctive patterns of gene expression, physiological states, developmental trajectories and location. The diverse, international consortium that builds the HCA is open and collaborative, bringing together and aligning experts who have formed networks focused on biological topics.

The HCA initiative will progress in phases to generate reference maps at increasing resolution. Google Maps serves as an analogy: instead of geographical features such as continents, countries, cities, streets and houses, the HCA's maps of the human body will "zoom in" on molecular and organizational features of organs, tissues and cells.

Tissue systems networks will include experts from the various biological systems represented in the atlas (nervous, peripheral nervous, lymphoreticular, immune, urinary, respiratory, female reproductive, male reproductive, hepato-pancreatic-biliary, gastrointestinal, endocrine, skin, musculoskeletal, cardiovascular, breast, organoids). Different development phases will also be represented in the atlas (e.g. developmental cell atlas, pediatric cell atlas). So far, the Chan Zuckerberg Initiative (CZI) has announced support for 17 Pediatric Networks, which will form part of the Human Cell Atlas (HCA). These are research grants enabling funded research consortia to undertake projects to perform the mapping of specified pediatric organs or tissue types. CZI has also announced support for 16 Ancestral Networks, which will form part of the Human Cell Atlas (HCA). These are research grants enabling funded research consortia to develop atlases or mapping exercises directed to ensuring the ancestral diversity of single-cell reference data generated.

[2] Human Cell Atlas Consortium, *White Paper*, October 18 2017, available at: https://www.humancellatlas.org/files/HCA_WhitePaper_18Oct2017.pdf

HCA networks are self-organising, with support from the HCA executive office, and have a degree of autonomy around a number of scientific practices that impact how data across the HCA is generated. For example, individual HCA networks may make decisions about the optimal experimental protocols, normalisation methods and batch correction methods for their biological system of interest.

This reference map aims to be an "open resource", which will dramatically accelerate discoveries by biological researchers, data scientists, and translational scientists and clinicians worldwide. As an open resource, the HCA will ensure that public reference data be made freely and widely available, however in the future access controls and authentication will be required for some subsets of data contributed, in accordance with local ethical and regulatory frameworks.

1.3. Ethical principles

Sharing of samples and data through collaborative research ultimately aims to contribute to the wellbeing of humans and humanity. Given the broad range of tissue sampling sources required to build the atlas, the HCA recognizes the importance of responsible use of human biological materials and in particular the need to develop transparent, respectful and responsible frameworks for tracking provenance, sampling, use, and analysis or secondary use of the data derived from such tissue donations. The principles of dignity, autonomy, privacy, confidentiality, non-discrimination and benefit to society should govern the collection and use of these materials, whether from living or deceased individuals, or from global populations.

As an international genomic data sharing consortium, the HCA abides by the Global Alliance for Genomics and Health *Framework for Responsible Sharing of Health Related Data*[3], and in particular implementing the following foundational principles:

- Respect Individuals, Families and Communities
- Advance Research and Scientific Knowledge
- Promote Health, Wellbeing and the Fair Distribution of Benefits
- Foster Trust, Integrity and Reciprocity

These principles are implemented by the HCA through the development of tools and resources which foster[4]:

- Transparency;
- Accountability;
- Data quality and security;

[3] Global Alliance For Genomics and Health, *Framework for Responsible Sharing of Genomic and Health Related Data*, December 9, 2014, available online at: https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/framework-for-responsible-sharing-of-genomic-and-health-related-data/

[4] Ibid.

- Privacy, data protection and confidentiality;
- Risk-benefit analysis;
- Sustainability;
- Education and training;
- Accessibility and dissemination.

1.4. Fostering global, equitable participation in the HCA

As a global initiative, the HCA recognizes the importance of involving representatives from all parts of the world and welcomes participation from as wide a range of communities as possible.

In particular, the HCA strives to adopt and implement the following overarching principles (ref: H3Africa, MalariaGen, 1000 Genomes project, HapMap):

- Reciprocity;
- Community consultation and engagement;
- Appropriate informed consent;
- Benefit sharing;
- Accountability.

In order to encourage participation from all communities and implement appropriate ethical and equity principles, this Ethics and Data Governance Document strives to adopt approaches that are interoperable and relevant across different groups and communities of participants and researchers. As this is a living document, the HCA welcomes participation, engagement and input to further develop these ethics governance principles, in view of ensuring this document remains relevant and interoperable across a wide range of research participants and tissue donors. In particular, the HCA's Equity Working Group is implementing an "equity in action" approach to involving the scientific community around the world. The Ethics Working group aims to work with the Equity Working group to adapt and adjust the HCA ethics toolkit, as needed, to reflect the growing diversity of the HCA donor and research communities.

In a related effort, members of the Human Cell Atlas Equity Working Group (EqWG) and of the Human Cell Atlas Ethics Working Group (EWG)[5] have drafted a public-facing position statement entitled "The Commitment of the Human Cell Atlas to Humanity"[6]. This statement and a companion publication[7] have been made available in open access. These documents describe the role of the HCA in helping to understand the biological and cellular variation arising across healthy human populations, and in providing local communities with resources to facilitate their participation in scientific research.

[5] See working groups composition online at: https://www.humancellatlas.org/learn-more/working-groups/

[6] Available online at: https://www.humancellatlas.org/wp-content/uploads/2021/10/25-OCT-2021-The-Commitment-of-the-Human-Cell-Atlas-to-Humanity-1.pdf

[7] Majumder, P., Mhlanga, M., Shalek, A., Guigó, R., Knoppers, B. M., & Wold, B. (2022). How to ensure the Human Cell Atlas benefits humanity. *Nature*, *605*(7908), 30–30. https://doi.org/10.1038/d41586-022-01186-0

1.5. About this Ethics and Data Governance Document

This Ethics and Data Governance Document provides a basis for framing ethico-legal issues within the organizational context of the HCA. It is cognizant of the consortium's international composition, and of potential divergences in cultures, languages, practices as well as diverse legal, societal, and ethical requirements. Therefore, it should be read in conjunction with other policies of the HCA, as well as national legislation or policies, which take precedence over this document.

More specifically, this document aims to:

- Provide guidance on ethico-legal issues for researchers taking part in the HCA, whether collecting tissue samples and/or submitting data as a Data Contributor, managing the DCP or accessing/downloading DCP data as a Data User;
- Provide guidance on the underlying ethical considerations governing the HCA to research ethics committees, local institutions or other regulatory bodies responsible for authorising deposition of research data in the HCA DCP;
- Clarify the technical and organizational measures in place to demonstrate transparency and accountability of the HCA towards participants, the scientific community and other stakeholders.

Other policies may be developed by HCA Inc., the HCA Organizing Committee, and other working groups: this document may evolve in time to align with other such policies.

Finally, this document is intended to provide guidance to different stakeholders, and does not aim to set any legal requirements on Data Contributors and/or Data Users. Data Contributors and/or Data Users are ultimately responsible to ensure that they have obtained the necessary approvals and are in compliance with local laws, regulations or policies, prior to contributing or using data from the HCA DCP.

2. HCA governance and organizational structure

2.1. Governance overview

The HCA is structured through a non-profit legal entity, **HCA Incorporated (HCA Inc.)**, established in the United States. It has applied for government recognition as a 501(c)(3) nonprofit entity in the United States. HCA Inc. ensures the regulatory compliance of the Human Cell Atlas, including its data protection compliance mandate. HCA Inc. also coordinates the administrative and communications activities of the larger HCA project, and is responsible for the administration of its finances. The HCA Inc. Board of Directors administers the activities of HCA Inc., and is composed of 7 to 15 members. HCA Inc. will, in the future, be supplemented by a Netherlands-incorporated legal entity in the European Union.

The scientific objectives of the HCA are steered and governed by an **Organizing Committee (OC)**. The responsibilities of the OC include convening the community through regular meetings, workshops and jamborees; coordinating and authoring key documents; defining scientific direction and purpose; providing ethical guidance; defining and upholding processes including quality-control standards and analytic standards; coordinating the HCA work products; and polling the HCA community at regular intervals for input on issues, including the performance of the OC.

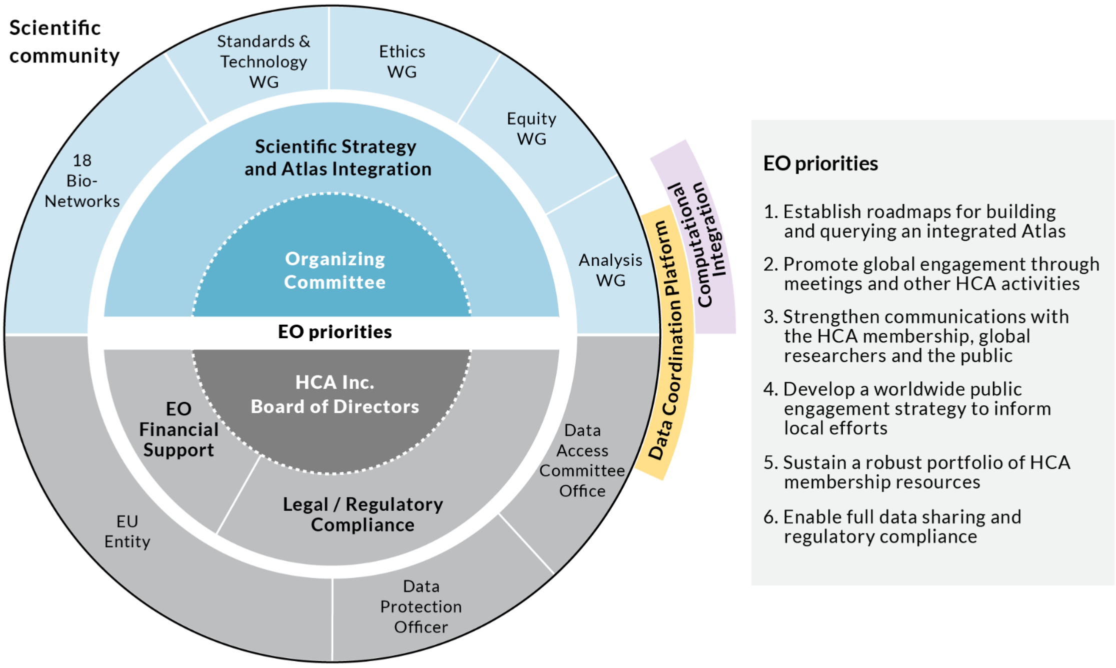
The OC has an **Executive Committee** responsible for performing routine tasks between OC meetings; preparing agendas, and providing guidance to the executive offices.

The OC establishes Working Groups and mandates them to take on specific key areas. At the moment, these include the following:

- Analysis Working Group (AWG)
- Standards and Technology Working Group (STWG)
- Ethics Working Group (EWG)
- Equity Working Group (EqWG)

The HCA **Executive Office (EO)** is responsible for providing administrative and logistics support to the HCA. Its principal responsibilities include developing partnerships with external organisations, performing communications and outreach activities, and convening HCA conferences and meetings.

HCA INCORPORATED: CENTRALIZED AND SUSTAINABLE SUPPORT FOR THE GLOBAL CONSORTIUM


Scientific community
Standards & Technology WG
Ethics WG
Equity WG
Analysis WG
18 Bio-Networks
Scientific Strategy and Atlas Integration
Organizing Committee
EO priorities
HCA Inc. Board of Directors
EO Financial Support
Legal / Regulatory Compliance
EU Entity
Data Protection Officer
Data Access Committee Office
Data Coordination Platform
Computational Integration

EO priorities

1. Establish roadmaps for building and querying an integrated Atlas
2. Promote global engagement through meetings and other HCA activities
3. Strengthen communications with the HCA membership, global researchers and the public
4. Develop a worldwide public engagement strategy to inform local efforts
5. Sustain a robust portfolio of HCA membership resources
6. Enable full data sharing and regulatory compliance

2.2. HCA Biological Networks

HCA Biological Networks are communities of collaborators responsible for the study of a particular tissue group, organ, or theme. Each Biological Network is responsible for efforts related to the imaging, spatial mapping, and single-cell analysis for its concerned area of research. These efforts are performed through a combination of volunteer efforts and the provision of targeted research funding to stimulate productivity and expedite Atlas development.

The HCA OC establishes the HCA Biological Networks, and the EO supports their activities. Eighteen HCA Biological Networks have been created thus far. A team of Coordinators is responsible for convening and for structuring the activities of each Biological Network. Coordinators are volunteer members with topic-relevant expertise.

The biological networks address the following research topics: Adipose, Breast, Development, Eye, Genetic Diversity, Gut, Heart, Immune, Kidney, Liver, Lung, Musculoskeletal, Nervous System, Oral and Craniofacial, Organoid, Pancrease, Reproduction, Skin.

2.3. Data Coordination Platform

HCA Inc. governs the **Data Coordination Platform (DCP)**, which includes making all policy decisions concerning the DCP, approving the overall plan for the DCP, and ensuring the plan's successful execution by the major developers of the DCP. HCA Inc., in collaboration with relevant scientific, technical, and legal experts, oversees the implementation of these policies, by providing guidance and making decisions concerning certain key topics, that include definition of the data manifest; official analysis pipelines; required metadata to reflect data collection standards; common coordinate framework; and any formal 'release portal'.

2.4. HCA Members and Projects

Any individual may become a **HCA Member** by registering at the HCA Member Registry and agreeing to abide by the principles of HCA. In addition, a HCA member who is a collaborator on an HCA project, or a member of an OC-designated HCA group is designated as an HCA Collaborating Member.

Any scientific project related to systematic biological characterization at single-cell resolution may become an **HCA Project** by registering in the HCA Project Registry.

2.5. Ethical Oversight

2.5.1. HCA Ethics Working Group

The **Ethics Working Group (EWG)** has an advisory role with respect to the ethical governance issues related to the HCA. It does not provide legal opinions, nor is it responsible for decisions made with respect to the structure or governance of the HCA. The EWG reports to the OC. It shall regularly report any issue it believes should be brought to the attention of, or which requires decision-making by the OC.

The HCA community can submit ad hoc ethics policy questions to the EWG via the ethics helpdesk (ethics-help@humancellatlas.org).

2.5.2. Local research ethics committees, institutions or other regulatory approval bodies

Oversight of the **collection of tissue samples, contribution and upload of research data by Data Contributors** should be ensured by the local institutional research ethics committee (where applicable), or the legal department of the Data Contributor's institution or organization. Indeed, local regulations or policies may require that ethics or legal approval be obtained prior to transferring datasets to external entities, or to other jurisdictions. Furthermore, additional regulatory approvals may be required for certain tissue sampling contexts (e.g. organ donation/anatomical gift, research with embryos/fetal tissue, etc.). Data Contributors should consult their local research ethics committee, data protection officers and/or legal departments to ascertain requirements, prior to sharing data with the HCA DCP.

With respect to the **download and use of research data by Data Users**, whether local institutional research ethics committee approval or other regulatory approval is required or not may depend on local regulatory or institutional requirements. For instance, this might vary based on the type and nature of the data accessed (e.g. aggregate data, raw sequencing data, etc.). Some jurisdictions consider use of certain types of coded data as constituting human research, while others do not (meaning that research is exempted from requiring ethics approval). Where in doubt, HCA Data Users should consult their local research ethics committee or other institutional regulatory approval body, prior to downloading HCA data.

2.6. Data access and release oversight

The HCA encourages contributors to submit datasets that have appropriate permissions for open access data sharing. However, it is understood that not all datasets can be released in open access and therefore, the HCA has also implemented controlled access to a subset of datasets. For datasets held in controlled access, applicants are required to submit an application to access data. A Data Access Compliance Office (DACO) administers the data access request process. A Data Access Committee (DAC) performs the review of these applications prior to granting access to HCA controlled access data. The DAC is responsible for determining that the applicants are affiliated to an accredited research institution, and that the proposed research project is consistent with HCA data access policies. Further information on requirements to apply for access to controlled access and the DACO review process are available on the HCA website.

2.7. Sustainability

Worldwide, the HCA involves collaboration with governmental agencies, partnerships with philanthropic foundations, organizations, and individuals, and collaborations with technology and pharmaceutical companies that can contribute equipment, expertise, or financial resources in support of HCA data collection, analysis, storage and distribution.

3. Tissue sampling and data collection by Data Contributors

3.1. Tissue type prioritization and data suitability

Where applicable, eligibility and enrolment criteria related to the types of tissues collected by a Data Contributor should be specified in that Data Contributor's scientific protocol or equivalent document.

However, as a whole, the HCA welcomes data meeting data suitability requirements[8], as updated from time to time.

3.2. Collaboration scenarios: sample collection and data generation

In this document, the term Data Contributor refers to the researcher or research group generally undertaking all phases from tissue sample collection, to data generation, to upload of research datasets to the HCA DCP.

In practice, however, this process may not always be undertaken by a single researcher or research group, but rather different collaboration scenarios may arise. For instance, a research group collecting tissue samples may collaborate with one or more research groups to analyse the sample and generate data and subsequently upload this data to the HCA DCP.

In cases where these research collaborations involve the exchange of tissue samples between institutions, the HCA ethics toolkit[9] contains a *template material/data transfer agreement* to assist collaborators in preparing agreements that enable streamlined contribution to the HCA.

3.3. Recruitment of participants

Participants are recruited by Data Contributors from multiple institutions around the world. Data Contributors are ultimately responsible for oversight of participant recruitment and tissue procurement procedures, and in particular their compliance with local and international ethical standards[10]. Different HCA tissue atlases may set priorities in terms of sampling and recruitment.

[8] Available online at: https://data.humancellatlas.org/contribute/contributing-data-suitability#data-suitability

[9] Available online at: https://www.humancellatlas.org/ethics/

[10] For example: World Medical Association (WMA), Declaration of Helsinki. At: http://www.wma.net/en/30publications/10policies/b3/index.html; World Medical Association (WMA), Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks (2016), available online at: https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/

From an ethical standpoint, it is understood that research data will have different provenances and routes of entry into the HCA DCP. Furthermore, different legal requirements, ethical frameworks and standards may apply depending on the tissue sampling and provenance contexts. An HCA ethics toolkit[11] is made available to assist Data Contributors in explaining the HCA to their local research ethics committee or other regulatory body, and appropriately recruiting and sampling tissues from participants. Nonetheless, Data Contributors are responsible to ensure that their recruitment procedures comply with such requirements, and make any necessary adjustments to template recruitment and consent materials developed by the HCA.

It is expected that sampling of tissues will involve the following **contexts** as part of different phases of HCA implementation:

- autopsy/organ donation;
- research project;
- leftover clinical tissue (e.g. biopsy, resection);
- fetal/embryonic tissue;
- tissue from commercial vendors;
- tissue from other repositories (e.g. existing biobanks).

Individual participants from which such samples will be sought may also include:

- Deceased participants;
- Healthy participants;
- Diseased participants;
- Pediatric participants;
- Participants from communities where there are particular considerations to take into account for biomedical research (e.g. indigenous communities, communities in resource-poor settings, etc.).

Data Contributors are responsible for the implementation of appropriate recruitment procedures, tissue sampling and consent practices, based on the specificity of the context in which tissues are sampled, local regulatory/legal requirements and local institutional or community best practices. Details on specific recruitment procedures and sampling can, for example, be described by the Data Contributors as part of their scientific protocol or similar document, that is specific to their local project.

3.4. Consent to research & health databases

International ethics guidelines applicable to health databases recommend that participants be adequately informed of the following elements[12]:

[11] Available online at: https://www.humancellatlas.org/ethics/

[12] See for example: World Medical Association (WMA), Declaration of Helsinki. At: http://www.wma.net/en/30publications/10policies/b3/index.html; World Medical Association (WMA),

- Purpose of the database;
- Risks associated with the collection, storage, use and sharing (e.g. open, controlled access) of data and materials;
- Nature of the data or materials to be collected;
- Procedures for the return of research results (where applicable);
- Rules of access to the database and governance arrangements;
- Privacy protections;
- That the participant may not be able to know what is done with their material or data;
- That if the sample or data is de-identified, the participant may not have the option to withdraw consent;
- Information on commercial use and benefit-share, intellectual property issues;
- Information on transfer and sharing of data and materials to other institutions or third countries.

As with other types of health research, institutional and jurisdictional requirements may vary and therefore Data Contributors should comply with local requirements.

In addition to these general elements, the HCA has developed core consent elements (see section below), that are reflective of data sharing and research undertaken through the HCA DCP.

3.5. Core elements specific to the HCA

Consent to take part in research involving the collection and analysis of tissue samples and sharing of research data with the HCA, should be obtained by the Data Contributors from the participant (first person consent) or, where applicable, by obtaining consent from their legal representative (substituted consent). Some types of tissues (e.g. embryos/fetal tissues), could require consent from each of the gamete providers, depending on the country's laws and regulations. Data Contributors should comply with all applicable local laws, guidelines and institutional policies. In particular, attention should be paid to requirements or limitations related to sampling of tissues and obtaining of data in certain special or vulnerable populations (e.g. research with human developmental tissues, research with paediatric participants, etc.).

The HCA ethics toolkit includes guidance on consent forms and consent clauses and provides examples of language formulating the core consent requirements. However, Data Contributors are free to use their own consent language, if it includes elements similar to the Core Consent Elements listed below. The elements listed below are not to be used as actual consent language/text, but rather, are high-level themes to include.

Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks (2016), available online at: https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/

Irrespective of the legal framework under which consent/substituted consent to tissue sampling is sought, it is suggested that, in order to foster interoperable, ethical, broad, and, transparent public (open) data sharing in the HCA, participant information and consent material minimally include the following **CORE CONSENT ELEMENTS for <u>public (open) data sharing of "data":</u>**

	Public (open) data sharing **If a Data Contributor wishes to <u>deposit data in the HCA public (open) access tier</u>, consent should be obtained for:**
Research data	Genetic analysis of data from the tissue sample and the collection of metadata related to the sample.
International sharing	International sharing of data
Future use	Any future, unspecified use of data
Commercial use	Use of data for commercial purposes
Public (open) access	No access controls or tracking over who is able to access data or for what use
Storage on cloud servers	Storage of data outside of the country where it was collected [i.e. consent language does not restrict storage of data on cloud servers, including private/commercial cloud service providers]
Duration of storage	Indefinite data storage
Data withdrawal	Not possible to withdraw data that has already been distributed and used
Re-identification	Risk that the participant could be re-identified in the future, including through linkage with external databases

If any of the items listed above are not included in the consent, datasets should not be deposited as-is in the open access tier of the HCA DCP without obtaining appropriate approvals (for example, research ethics committees, institutional representative, etc.).

Finally, Data Contributors should note that the requirements for informed consent to donate tissue samples for research may differ from other types of consent requirements (e.g. consent to the processing of personal data under data protection laws). Therefore, for the sake of transparency towards participants, *informed consent for the use of tissues in the context of research* as an international bioethical norm (and enacted in several national laws on research with human subjects), is treated here as distinct and separate from other consent requirements that may be applicable. Please consult with your appropriate institutional representatives on matters related to compliance with data protection regulations, particularly regarding sharing of data in a public database (see also the Global Alliance for Genomics and Health, GDPR Forum[13] for further guidance).

3.5.1. Controlled access (in development)

While the HCA DCP encourages contributors to submit datasets that have the appropriate permissions to be released in open access, some projects collecting data to build the next phases of the HCA may not be able to deposit all datasets in public (open) databases and thus need to obtain consent for controlled access (for example, due to ethical or legal constraints). The HCA has developed a controlled access portal, which will enable the HCA to host more sensitive

[13] Available online at: https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/gdpr-forum/

data. We suggest that Data Contributors use the consent elements below, for datasets requiring controlled access:

	Controlled data sharing If the Data Contributor wishes to deposit datasets in **a controlled-access tier** research consent should be obtained for:
Research data *[same as for public sharing]*	Genetic analysis of data from the tissue sample and the collection of metadata related to the sample
International sharing *[same as for public sharing]*	International sharing of data
Future use	Future, unspecified research use of data
Commercial use *[same as for public sharing]*	Use of data for commercial purposes
Controlled access tier	Sharing of sensitive data** through a controlled-access mechanism, meaning that data users are required to register, create a researcher account and agree to access conditions before accessing such data* **If the Data Contributor is depositing some data in the controlled-access tier and other subsets of data in the open access tier (e.g. gene count data), please also add consent element for the open access tier, and specify the general types of data deposited in each tier.* *For example: Sharing of a subset of data through a public database with an open-access mechanism, meaning that it is totally transparent to the donor that there are no controls or tracking over who is able to access data or for what use* *** The definition of what constitutes sensitive data may vary from one jurisdiction to another. Data Contributors should refer to their ethics committee of data protection officer to determine whether a particular dataset is sensitive and requires controlled access.*
Storage on cloud servers *[same as for public sharing]*	Storage of data outside of the country where it was collected [i.e. consent language does not restrict storage of data on cloud servers, including private/commercial cloud service providers]
Duration of storage *[same as for public sharing]*	Indefinite data storage
Data withdrawal *[same as for public sharing]*	Not possible to withdraw data that has already been distributed and used
Re-identification *[same as for public sharing]*	Risk that the participant could be re-identified in the future, including through linkage with external databases

3.6. Additional consent considerations

The collection of tissue samples and metadata from participants is under the responsibility of the Data Contributor. The HCA does not have any direct interaction with participants.

In particular, as part of the project's consent materials the Data Contributor should consider providing participants with additional information that may be specific to the project or tissue sampling context, for example:

- Type of tissue collected;
- Type of metadata collected (for example, indicate whether access to the medical record is required);
- Procedures and risks involved in the collection of the tissue samples;
- Where tissue analysis will be undertaken (for example, in cases where the tissue will be sent to other sites for sequencing);
- Fate of the tissue sample, once it is analyzed (e.g. destruction, banking locally, banking at another site, further sharing of tissue samples, etc.).

The HCA itself will not receive or bank tissue samples (although Data Contributors or their collaborators may do so). Data Contributors are ultimately responsible for the management and oversight of the tissue sample itself.

3.7. Secondary research use of retrospective (legacy) samples and datasets for HCA

Where available, the HCA welcomes the contribution of retrospective (legacy) datasets, as well as datasets obtained through the use of legacy samples (e.g. from leftover clinical samples, pre-existing repositories, etc.). This includes, for instance:

- samples collected/datasets generated prior to the creation of the HCA;
- samples collected/datasets generated after the creation of the HCA, but designed for projects or uses other than the HCA, including clinical purposes (also called "secondary use").

Prior to the contribution of legacy samples/data, the Data Contributor should consult with their local research ethics committee or, where applicable, other regulatory body to determine whether legacy samples/datasets are suitable for contribution to the HCA DCP. The HCA Ethics Toolkit[14] contains a *retrospective assessment tool*, as guidance to help Data Contributors undertake a preliminary assessment as to whether additional steps may be required (e.g. re-consent of participants, obtaining an ethics waiver of consent, etc.) prior to uploading these datasets to the HCA DCP.

3.8. Risks and benefits to participants

Participants will not benefit personally from sharing data with the HCA because research usually takes a long time to produce medically useful results. However, their contribution may help others in the future and open data sharing has the potential to reduce barriers in data sharing, foster international collaboration, and speed scientific discoveries.

Sharing data through an **open access** means that the HCA does not control who is accessing the data or for what purpose(s). Data in the open access tier is made publicly available to anyone, through the HCA DCP. As opposed to a controlled or registered access system, which involves some level of oversight over the individuals accessing the data or over the nature of the research project proposing to use the data. Open access presents some additional risks, due to the unknown future use of the data.

Much like fingerprints, it is possible to identify someone if certain data are put together from different sources. While appropriate data security measures are in place to protect the privacy of participants and their biological relatives (see Section 5), there is always a risk that their data may lead to participants or their families being re-identified. As technology advances, there

[14] Available online at: https://www.humancellatlas.org/ethics/

may be new ways of linking data back to participants that cannot be foreseen today. Although the HCA does not hold directly identifying information about participants, if participants or their biological relatives have provided genomic data to another database or to social media, it may be possible to re-identify participants by comparing and linking data from different sources and putting data together.

In addition, while the transcriptomic data stored in the HCA DCP may not currently reveal any sensitive data about individuals, as technology evolves, it is possible that this type of information may reveal information that was not foreseen. Individual-level metadata collected about participants may also reveal certain characteristics about the individual, some of which may be sensitive in nature, such as age, geographical location, limited clinical information (disease, medication, treatments, etc.), cause of death (where applicable), ethnicity, etc. In the HCA DCP, the exact nature and richness of metadata contributed may vary form one Data Contributor to another.

To evaluate the risks related to the project, the HCA has undertaken a Data Protection Impact Assessment, which will be periodically revised, in order to continue assessing and monitoring risks relating to this new technology.

3.9. Withdrawal of participant data from the HCA

Withdrawal of participant data is managed entirely by the Data Contributors. Since the HCA does not collect any identifying data about participants, it is unable to receive requests from participants to remove individual-level datasets.

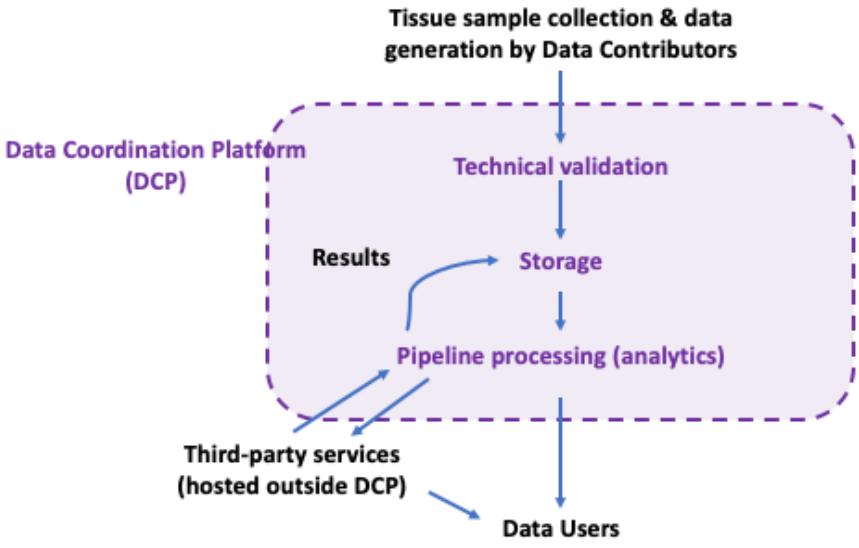
Data Contributors are responsible for receiving withdrawal requests from participants, and should ensure that their coding process used to submit data to the HCA enables individual-level dataset removal. In the event of participant withdrawal, the Data Contributor is responsible for contacting the HCA DCP to have relevant datasets removed.

Some Data Contributors may also choose to irreversibly remove codes in order to further de-identify data (this is related to concepts such as “anonymization”, “pseudonymization” and “irreversible de-identification”) (see Section 5 for further explanations). This process should be undertaken in consultation with the Data Contributor’s local research ethics committee or other regulatory approvals, as the process of de-identification may be subject to different regulatory interpretations and implementations, depending on the country and institution. However, it should be noted that this process could make it difficult or impossible for a participant to request that their data be removed from the HCA, for example if the Data Contributor has uploaded datasets from several participants. Consent forms should therefore be worded appropriately.

After the removal of individual-level data, aggregate and summary data are recalculated without the data from the removed individual or individuals. However, data in the DCP that has been accessed by or sent to other researchers around the world cannot be withdrawn if it has already been used or published.

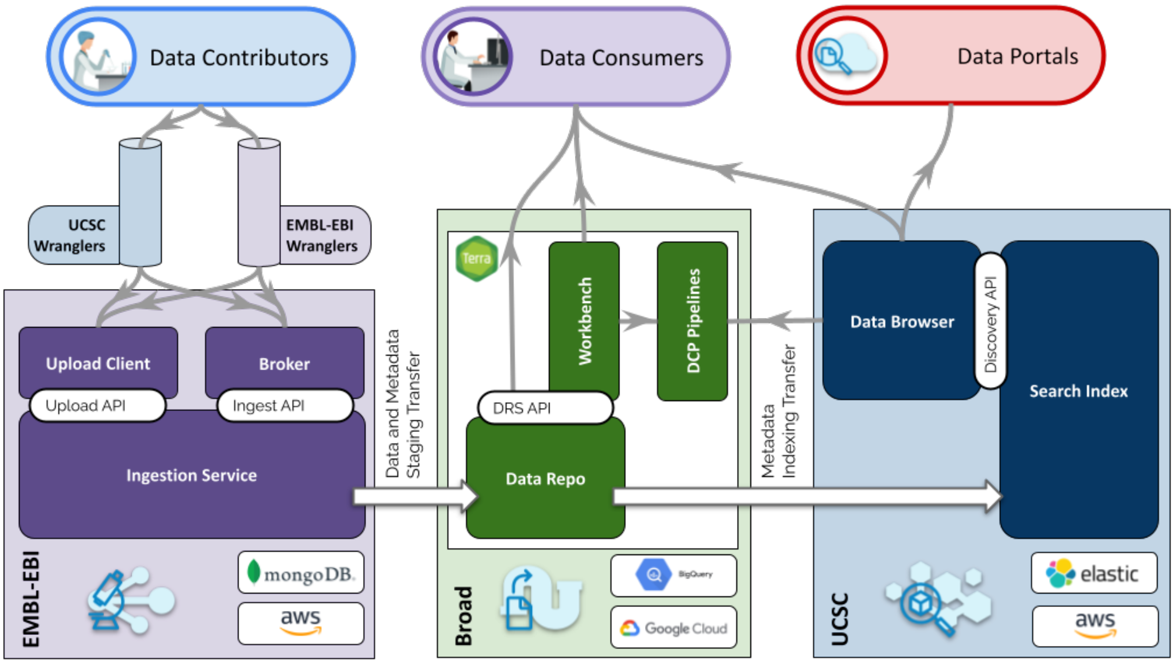
4. Data contribution, storage and use in the HCA DCP

The HCA DCP is an open source, cloud-based platform developed to organize, standardize, and make accessible HCA research data.


Tissue sample collection & data generation by Data Contributors
Data Coordination Platform (DCP)
Technical validation
Results
Storage
Pipeline processing (analytics)
Third-party services (hosted outside DCP)
Data Users

The DCP data flow architecture is organized around four key steps [1] a data ingestion service, [2] a synchronized data repository with multiple cloud replicas (the Terra Data Repository) [3] a collection of secondary analysis pipelines for basic data processing and [4] a collection of tertiary portals for analyses, visualizations, and rich forms of data access[15].

[15] Full information is available at: https://www.humancellatlas.org/data-coordination/



The following points summarize key steps[16]:

1. Data ingestion service
   - *Data generation and collection*: Data Contributors are responsible for the collection of tissue samples at their local institutions, as well as the generation of sequencing data (either at their local institution or through collaboration with other groups) and sometimes derived data (e.g. expression matrices, see "Quantifications" section 1.1.1 below), as well as the collection of metadata for projects, experiments and samples. Accepted data types can then be submitted via the user interface or by contacting data wranglers.
   - *Ingestion*: Data flow begins with the ingestion of raw experimental data files and associated metadata. The ingestion process is supported by HCA data wranglers, as well as an interactive user interface, and a Ingestion Service API. The Ingestion Service processes, validates and assembles biomolecular data and metadata before pushing them to the Terra Data Repository, where all HCA data are stored.
2. The Terra Data Repository
   - *Depositing data into the Terra Data Repository:* The Terra Data Repository is a platform that hosts HCA data and HCA data analysis software that is contributed to the DCP. Prospective Data Users create Terra workspaces that enable them to import data and/or data analysis tools, which reside on the Terra data repository and the Terra methods repository, respectively. Controlled access data can be imported to a Terra workspace once the DAC has authorised

[16] These steps are fully detailed online at: https://www.humancellatlas.org/data-coordination/

the Data User to access the data. All data remains on the Terra Data Repository, which is hosted on the Google Cloud Platform (GCP), and is made available to users through the GA4GH Data Repository Service (DRS) API.

3. Secondary analysis services
   - *Data processing:* Processing of the datasets is then done via the secondary analysis pipeline and derived results from the secondary analysis pipelines are also deposited into the Terra Data Repository. Secondary analysis pipeline processes raw data using community-vetted algorithms to generate intermediate derived results that will be deposited back into the Terra Data Repository.
4. Tertiary portal
   - *Data access:* Data Users can access both raw and processed data from a wide range of downstream user-facing portals, including analyses, visualizations and other forms of access. New portals can eventually be developed by anyone in the scientific or computational community, for a wide diversity of use cases.
   - To encourage a fully open ecosystem, there will be no requirements or governance around portal development, and instead the HCA encourages the community to work together to develop best practices and share resources.

4.1. Data upload and processing

4.1.1. Types of data accepted

The HCA will eventually accept all types of single-cell data, but it currently accepts and stores the following types of data, from Data Contributors:

- **Transcriptome data identifies and quantifies** RNA molecules in different kinds of cells. Currently in the HCA, transcriptomic data-types include:
  - Single cell sequence data (scRNA seq)
    - These are sequencing data in raw form that results from experimental workflows, such as 10x Genomics; Fluidigm, etc. These are strings of raw sequence (i.e. reads) stored in a FASTQ or FASTA file format.
  - Sequence aligned scRNAseq data
    - These are scRNAseq FASTA/FASTQ read data that have been mapped to human reference genome and/or transcript model set. These are text files stored in Sequence Alignment Maps (SAM) format or Binary Alignment Maps (BAM) containing the same information stored in compressed binary form.
  - Quantifications
    - Quantification values are the number of aligned reads assigned to a given gene (i.e. gene quantifications) or transcript (i.e. transcript quantifications) in raw (e.g. counts, effective counts) or normalized (e.g. TPM, FPKM) form. Quantification values do not contain base-by-base sequence information. They are stored and exchanged as Expression

        Matrices that enable a data user to compare quantifications across different cells and samples.
        - Spatial transcriptomic data, derived by probe hybridization, are captured as primary image data and are processed into quantification matrices with associated positional coordinate metadata. Hybridization data do not contain base-by-base gene sequence information.
- Biological material metadata is information accompanying and annotating the above files. This metadata makes the sequence files more useful for other researchers because it explains, for example, how the data was generated or the organ from which it comes from. Biological material metadata contains metadata about samples (including tissue information as well as details about the cell suspension that is the material being sequences) and the donors that provided these samples. Donor information may include details about patients that is sensitive information, whereas sample information does not.
    - A current "metadata dictionary" is found online[17]. This includes information about:
        - Biomaterial from which the cells were taken (e.g. organ type, disease state, information about the participant such as ethnicity, sex, and age)
        - The Project that the sample came from (e.g. contributor name and lab details, funding sources for the project)
        - The Protocols that were applied to generate the sample (e.g. collection method, library construction details)
        - The analysis Process applied to the sample (e.g. institution where processing took place and researcher who performed the processing) and
        - Information about the Files themselves (e.g. file name or description)

The extent and nature of metadata attached to a data file can, on their own, present risk to the participant of unwanted re-identification.

Further information about the process for data contribution can be found on the HCA website[18].

4.1.2. Public database (open access)

Data Contributors are required to determine whether their datasets can be released in a public (open access) database or whether datasets require controlled access.

In particular, Data Contributors and their institutions should be aware of the following regarding the open access tier:

- All datasets submitted and stored in the open access tier of the DCP (raw sequencing data, gene expression profiles, metadata) are public;
- There are no access controls (e.g. no data access committee) to access datasets in the open access tier of the HCA DCP;

[17] Available at: https://www.humancellatlas.org/data-coordination/

[18] Available at: https://data.humancellatlas.org/contribute

- Users are not required to register to access open access datasets, they are not required to provide credentials and there is no data access agreement requirement prior to accessing and using datasets in the HCA DCP;
- Data in the open access tier can be used for any type of research (the HCA DCP does not review the type of research undertaken by the Data Users);
- Datasets in the open access tier can be freely downloaded.

4.1.3. Coding of datasets by Data Contributors

Prior to uploading research data to the HCA DCP, Data Contributors should ensure that all research data is appropriately coded (whether controlled access data or open access data). Coding, de-identification, pseudonymization or other data protection requirements may vary between jurisdictions (e.g. GDPR, HIPAA), and institutions - therefore, where in doubt, Data Contributors should consult their local research ethics committee, data protection office, or legal department to determine the appropriate standard.

It is suggested that coding done by Data Contributors minimally meet the following elements (in addition to any other local regulatory requirement):

- Removal of all directly identifying data (e.g. name of participants, contact information, official identification numbers (e.g. healthcare number), etc.) or other types of personal data (e.g. date of birth, area code, etc.) from the research data and the metadata contributed to the HCA's DCP.
- Data Contributors should consider single- or double-coding the data – replacing all identifiers with an alphanumeric code – prior to submission to the HCA DCP. Data Contributors can, for example, consult the European Medicines Agency Good Clinical Practices, ICH E15[19] and ICH E18[20]) for guidance on single or double coding:
    - *Single coding*: Single coded data and samples are usually labelled with a single specific code and do not carry any personal identifiers. It is possible to trace the data or samples back to a given individual with the use of a single coding key. The Data Contributor is responsible for keeping the coding key.
    - *Double coding:* Double coded data and samples are initially labelled by the Data Contributor with a single specific code and do not carry any personal identifiers. The data are then relabelled with a second code before being submitted to the HCA. This second code is linked to the first code via a second coding key. It is possible to trace the data or samples back to the individual by the use of both coding keys, which are held by the Data Contributor. The use of the second code

[19] European Medicines Agency, ICH E15, Definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories, Step 4 (November 2007), available at: https://www.ema.europa.eu/en/ich-e15-definitions-genomic-biomarkers-pharmacogenomics-pharmacogenetics-genomic-data-sample-coding

[20] European Medicines Agency, ICH E18, Guideline on genomic sampling and management of genomic data, Step 5 (October 2017), available at: https://www.ema.europa.eu/en/ich-e18-guideline-genomic-sampling-management-genomic-data#current-effective-version-section

provides additional confidentiality and privacy protection for participants over the use of a single code, as access to both coding keys is needed to link any data or samples back to a participant identifier.

- The linking-log (or logs) to link the code to the participant's identity should only be kept by the Data Contributor.

Data Contributors should note that *full anonymization*, as defined in ICH E15, in which coding does not allow for participants to be re-identified as the coding keys have been deleted, has limitations in the context of genomic research. Indeed, with the increasing availability of genomic information and analysis methods, it is not always possible to prevent re-identification of each participant by deleting the link between the participant's identifiers and the unique code(s). Anonymization often needs to be determined on a case-by-case basis, taking into account, for instance, all the means of identification, direct or indirect, reasonably likely to be used by any person, and objective factors, including the costs of and the amount of time required for identification, the available technology at the time of the processing, and technological developments.

4.1.4. Metadata contribution considerations

Metadata refers to secondary data elements that describe the primary data contributed and are useful for better understanding the content of the data and the context of its collection. Metadata elements can, for example, denote date and location of sample collection, unique participant study code, age, or gender, phenotypic data, among other data features.

Metadata is a useful and necessary element to allow data users to perform useful research with the data. It is especially relevant for drawing comparisons across datasets and data points, and drawing inferences from the data that can be generalized across the population. Rich metadata, meaning metadata that contains a lot of information about the concerned individual, does present some privacy considerations.

Data Contributors should note that some metadata fields create a direct risk of individual identification. For instance, if metadata includes the concerned participant's name or home address, this could lead to a breach of privacy. Other metadata fields are individually low risk but can present privacy issues if a certain metadata fields are combined together. For instance, a participant's gender, age, health condition, and city of residence are individually non-identifying pieces of information. However, if a dataset's metadata contains all of this information linked together, there is a heightened chance that the participant could be identified by the combination and linkage of all this metadata.

To reduce the risks of metadata leading to privacy harms or re-identification, the following general approaches are recommended (in case of doubt, Data Contributors should consult their local research ethics committee, data protection officer, or equivalent):

- Free text fields should be avoided where possible as a Data Contributor could inadvertently contribute directly identifying metadata such as a person's name.
- Do not include metadata elements that could directly identify the underlying participant. These include obvious identifiers such as name, but in some circumstances can also include less obvious identifiers such as disease name or location of data collection. For instance, denoting particularly rare diseases or collection sites in sparsely inhabited locations could risk directly identifying the underlying participant.
- The total number of metadata fields should be limited to necessary data. If a Data Contributor has provided enough metadata elements to risk individual re-identification, some of them should be eliminated.
- Certain data modification techniques can be used to reduce identifiability. *Generalization*, for example, means taking variables that are highly specific (i.e. an age) and representing them in a more general manner (i.e. a range of ages such as 5-10, 10-15, 15-20, etc.). Another method, *suppression*, means eliminating highly unique or identifying variables from a record to protect against identification.
- Where possible, individual-level data should be de-identified or made anonymous using methodologies that preserve data utility but restrict individual identifiability and mitigate other privacy risks. Suppressing records or variables, generalizing variables, and relocating potentially identifying information across records are among possible de-identification methodologies. Clinical trial agencies in Europe and Canada have recommended a residual re-identification risk of 5% as the threshold for data to be considered anonymized when performing fully open public release. Certain statistical methodologies have been proposed for determining if de-identified data remains identifiable.
- Statistical methods include "k-anonymization," "l-diversity," and "t-closeness" which measure if certain members in a dataset have a unique combination of data elements that could cause them to be identified, and also if certain subgroups (e.g. rare disease patients, specific communities) within the dataset have similar sensitive data attributes, that create sensitive inferences about those subgroups.

4.1.5. Data Contributor Agreement

When uploading data to the HCA DCP Data Contributors will be required to enter into a Data Contributor Agreement. The Data Contributor Agreement is under development. This agreement may, amongst other things, require that the following ethico-legal conditions be met:

- Ensure that datasets have the **appropriate consent, exemptions, ethics waiver or regulatory approval** to allow them to be deposited in the appropriate tiers (either open access or controlled acccess tier);
- Ensure that data coding/de-identification standards and **removal of participant identifiers** (direct and indirect) **meet local data protection requirements**;
- Retain all **documentation required to demonstrate compliance** with the agreement, and with all relevant laws, contracts and professional obligations

- Provide the HCA with contact information (of principal investigator or other similar actor and of affiliated Institution), ensure the contact information is kept up to date, and act as contact point for requests by participants and other concerned data subjects;
- Warrant that data meets the HCA's policies, including the Data Release Policy;
- Notify the HCA of any breach of its terms on the part of the Data Contributor.

4.2. Data storage in the DCP

Once uploaded to the DCP, research data will be stored indefinitely by the HCA, or, until it is no longer useful for research purposes or, the Data Contributor requests that it be withdrawn.

The data contributed to the HCA is stored in the Terra Data Repository, which is a cloud-native space for storage that leverages multiple cloud platforms. Currently, all data in the Terra Data Repository is stored in the Google Cloud Platform (GCP), ensuring that the data is accessible in each environment.

4.3. Data Access by Data Users

4.3.1. Terms of use and agreement

4.3.1.1. Open-access terms of use

When accessing open access data from the HCA DCP, Data Users may be required to abide by the Open Access Terms of Use. The terms of use are under development. Amongst other matters, these terms of use will require that the Data User agrees to the following ethico-legal requirements:

- **Compliance with HCA policies**, as may be adopted from time-to-time;
- If applicable, hold and maintain any required **approvals from regulatory authorities** (e.g. ethics, other…);
- Ensure that data is used in compliance with local data protection requirements;
- Potential to include the following obligations (subject to discussion): Not to sell data; to notify HCA of user breach of terms; not to harm / discriminate; not to make IP claims on primary data; stop using / destroy data at HCA request; impose same conditions on data shared with third parties.

4.3.1.2. Controlled access

Researchers requesting access to controlled access datasets will be required to follow the Data Access Compliance Office (DACO) Procedures, as posted on the XXXX website to apply for access. Following approval by the DAC, will be required to enter into a Data Access Agreement with HCA Inc.

5. Data Protection and privacy

5.1. Compliance with regional data protection regulations

5.1.1. Research ethics and data protection

Given that the HCA is building a global data intensive resource with the help of thousands of participants, Data Contributors and Data Users should be aware that different frameworks may apply to the regulation of human research and/or the donation of tissue samples on the one hand, and regulations pertaining to privacy/data protection, on the other. They should also be aware that certain regulatory frameworks may have effects and implications outside of the country in which they are enacted (for example, the European *General Data Protection Regulation*).

Indeed, standards may vary across different types of regulatory sectors. As an example, informed consent to participate in a research project is a generally accepted ethico-legal requirement in most biomedical research involving human participants, both in national laws and international guidelines. However, consent to participate in research is a distinct notion from consent under certain data protection laws, such as the European *General Data Protection Regulation* (GDPR). While consent is a fundamental concept in the field of human biomedical research, data protection law provides different mechanisms for the processing of personal data (including personal data generated in the course of research projects)[21].

For this reason, HCA Data Contributors and Data Users are encouraged to familiarize themselves with their local regulatory requirements regarding both research ethics and data protection/privacy, in order to understand how their activities in contributing or using HCA data may be regulated in their country. In case of doubt, it is suggested that Data Contributors and Data Users consult with their research ethics committee, their institutional data protection officers (or equivalent), or with legal counsel.

5.2. Protection of participant data

To protect participants' information, Data Contributors should code research data prior to uploading to the DCP. If the Data Contributor's research ethics committee requires data to be double-coded prior to depositing data the HCA DCP, double coding should be applied.

The key to link the participant's HCA research data (e.g. linking log) is only to be held locally by the Data Contributor, and subject to local practices. The HCA will not collect information that directly identifies participants (e.g. names, contact information).

[21] For further reading, see *Global Alliance for Genomics and Health*, "What is the difference between research ethics consent and data protection consent" : https://www.ga4gh.org/news/what-is-the-difference-between-research-ethics-consent-and-data-protection-consent/

Participants will not be identified as part of scientific conferences or appear in scientific publications.

5.3. Data Protection Impact Assessment (DPIA)

The HCA is conducting a Data Protection Impact Assessment (DPIA), which examines the data handling, analysis and storage practices of the HCA in order to identify and minimise the data protection risks related to this project. This DPIA aims to comply with the requirements established in the European *General Data Protection Regulation*. Its contents will include data flow mapping, risk assessment and mitigation, internal and external stakeholder consultation, assessment of technical measures (including the DCP construction, and data intake), and a determination of how to best preserve data subject rights. Given that the HCA is piloting innovative technology, this DPIA will be revisited over the course of the HCA's implementation and lifespan, so as to reassess risks.

A DPIA is a process designed to describe the processing, assess its necessity and proportionality and help manage the risks to the rights and freedoms of individuals resulting from the processing of personal data by assessing them and determining the measures to address them. The DPIA helps HCA not only to comply with requirements of applicable privacy and data protection laws, but also to demonstrate that appropriate measures have been taken to ensure compliance.

6. Database closure/decommissioning

Data housed in the HCA DCP is under the custodianship of HCA Inc.

If the HCA database were to cease activities at any time in the future, a decision made with regard to the transfer or closure of the database will involve the Board of HCA Inc. Efforts will be made to transfer the coded data to a third party that agrees to comply with the HCA policies and the terms of participants' consent. As a note, transfer of server storage entities does not constitute a transfer of data custodianship). Prior to any transfer of data custodianship and responsibilities with respect to the maintenance of HCA, Data Contributors will be notified of such decision to transfer data and provided with the opportunity to accept the terms of transfer, or to withdraw data from cohorts under its purview.

Appendix A: Supporting resources

Human Cell Atlas Consortium, *White Paper*, October 18 2017, available at: https://www.humancellatlas.org/files/HCA_WhitePaper_18Oct2017.pdf

Human Cell Atlas, *Ethics Toolkit*, available at: https://www.humancellatlas.org/ethics/

Global Alliance for Genomics and Health, *Privacy and Security Policy*, Version 2, August 2019, available at: https://www.ga4gh.org/wp-content/uploads/GA4GH-Data-Privacy-and-Security-Policy_FINAL-August-2019_wPolicyVersions.pdf

Global Alliance for Genomics and Health, *Security Technology Infrastructure*, Version 3, August 15, 2019, available at: https://drive.google.com/file/d/1kfy96aKbamXsgxvri1CnNgjdFNHc1ufa/view

Global Alliance for Genomics and Health, *Consent Policy*, Version 2, September 2019, available at: https://www.ga4gh.org/wp-content/uploads/GA4GH-Final-Revised-Consent-Policy_16Sept2019.pdf

Global Alliance for Genomics and Health, *Consent Clauses for Genomic Research*, Version 1.0, July 2020, available at: https://www.ga4gh.org/wp-content/uploads/Consent-Clauses-for-Genomic-Research.pdf

Global Alliance for Genomics and Health, *Data Use Ontology,* available at: https://github.com/EBISPOT/DUO

World Medical Association (WMA), *Declaration of Helsinki*, available online at: http://www.wma.net/en/30publications/10policies/b3/index.html

International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH), *Good Clinical Practice (GCP) Guidelines*, available online at: http://www.ich.org/products/guidelines.html

Council for International Organizations of Medical Sciences (CIOMS) in collaboration with the World Health Organization (WHO), *International Ethical Guidelines for Health-related Research Involving Human* (2016), available online at: http://cioms.ch/ethical-guidelines-2016/WEB-CIOMS-EthicalGuidelines.pdf

International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH), *Guidelines on Genomic Sampling and Management of Genomic Data (E18) (Step 4)*, (August 2017), available online at: http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/genomic-sampling-and-management-of-genomic-data.html

United Nations Educational, Scientific and Cultural Organization (UNESCO), *UNESCO Recommendation on Open Science*, November 2021, available online at: https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en

World Medical Association (WMA), *Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks* (2016), available online at: https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/