ICGA DATA POLICY

1. Introduction 2
2. Purpose and Scope 2
3. Important Terms 2
4. Source and Types of Data 4
5. Flow of Data 5
6. Data Processing 7
7. Data Processing Principles 7
8. Triggers for Data Disposal 8
9. Data Disposal Methods 8
10. Roles and Responsibilities of the Compliance Committee 8
11. Data Access Rights 9
12. Managing Open Access to Data 9
13. Data Release and Exemptions 9
14. Data User Agreements 9
15. Withdrawal of Sample Data 10
16. Data Security 10
17. Data Audit 10
18. Data Protection Regulations and Compliances 10
19. Use of Aadhaar Identification 10
20. Data Breach 11
21. Penalties 11

Annexure A to ICGA Data Policy 12

The ICGA Data Policy **"Policy"** has been made effective from 06 June 2022. **The ICGA Foundation reserves the right to periodically review the policy to keep it in sync with the requirements of laws and regulations of the Government of India.**

All stakeholders are expected to read, understand, and adhere to this policy statement as part of their duties and responsibilities at ICGA.

If you have any questions or concerns about this policy statement or any other data protection matters at ICGA, please contact **Dr. Suveera Dhup** (COO, ICGA Foundation) at suveera@icga.co.in.

1. Introduction

The India Cancer Genomics Atlas (ICGA) Foundation, abbreviated as ICGA, is a not-for-profit, public-private-philanthropic partnership with the objective to establish a curated platform and data repository of multi-omics data on Indian cancers. The data, generated through the collaborative efforts of various institutions, will be made available in the public domain for research purposes with applicable safeguards, facilitating the researchers to find cures for cancer. The collection, processing, and storage of biological samples will be carried out in India. Similarly, the data generated from the processing of these biological samples will be processed and stored within India.

2. Purpose and Scope

The core activities conducted by ICGA involve collecting and processing medical data across various geographies with diverse stakeholders. The intention of this policy is to ensure that the stakeholders are aware of the relevant data protection regulations and privacy laws that must be observed and complied with.

3. Important Terms

3.1 **"access"** means retrieval of information/data by a user from a repository for research/patient care/commercial purpose;

3.2 **"anonymization"** in relation to personal data, means the process of encrypting or removing information that can identify an individual from the data, which meets the standards of irreversibility specified by an Authority recognized by the applicable laws in India;

3.3 **"anonymized data"** means data that has undergone the process of anonymization**;**

3.4 **"authority"** means the Data Protection Authority of India or another Authority recognized by the applicable laws in India;

3.5 "**biometric data**" means facial images, fingerprints, iris scans, or any other similar personal data resulting from measurements or technical processing operations carried out on physical, physiological, or behavioral characteristics of a data principal, which allow or confirm the unique identification of that natural person;

3.6 "**biological data**" means information related to living organisms, including their nucleic acids, protein sequence, metabolites, and other molecular and functional characteristics;

3.7 "**consent"** means expressed informed consent granted with full knowledge and understanding of the nature, purpose, and consequences of the collection, use, storage, or disclosure of their data in any written/electronic/video form;

3.8 "**compliance committee**" is a group of persons responsible for the overall data compliance of ICGA. This can be a shared responsibility for an identified group within the ICGA consortium led by a senior member.

3.9 "**data**" includes a representation of information, facts, concepts, opinions, observations, documents, images, charts, tables, figures, or instructions in a manner suitable (digital/analog) for communication, interpretation, or processing by humans or by automated means;

3.10 "**data fiduciary**" means any person, including the State, a company, any juristic entity or any individual who alone or in conjunction with others determines the purpose and means of the processing of personal data;

3.11 "**data principal**" means the natural person to whom the personal data relates;

3.12 "**data processor**" means any person, including the State, a company, any juristic entity or any individual, who processes personal data on behalf of a data fiduciary;

3.13 "**data repository"** means platform for storage of biological data and responsible sharing of data with the user;

3.14 "**data storage"** means retaining biological data in digital form on storage devices;

3.15 "**de-identification"** means the process of removing, obscuring, redacting, or delinking personally identifiable information from an individual's data in a manner that eliminates the risk of unintended disclosure of the identity of the individual;

3.16 "**genetic data**" means personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the behavioral characteristics, physiology, or health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question;

3.17 "**health data**" means the data related to the state of physical or mental health of the data principal and includes records regarding the past, present or future state of the health of such data principal, data collected in the course of registration for, or provision of health services, data associating the data principal to the provision of specific health services;

3.18 "**health facility**" refers to health facilities across the country such as hospitals, clinics, diagnostic centers, health and wellness centers, mobile vans, ambulances, pharmacies, etc.;

3.19 "**Health Information Provider**" or "**HIPs**" means hospitals, diagnostic centers, public health programs, or other such entities, act as information providers (by generating, storing, and distributing health records) in the digital health ecosystem.

3.20 "**metadata"** means the information that describes the data source and the time, place, and conditions under which the data were created. Metadata informs the user of who, when, what, where, why, and how data were generated. Metadata allows the data to be traced to a known origin and known quality.

3.21 "**personal data**" means data about or relating to a natural person who is directly or indirectly identifiable, having regard to any characteristic, trait, attribute, or any other feature of the identity of a such natural person, whether online or offline, or any combination of such features with any other information, and shall include any inference drawn from such data for the purpose of profiling;

3.22 "**personal data breach**" means any unauthorized or accidental disclosure, acquisition, sharing, use, alteration, destruction of, or loss of access to, personal data that compromises the confidentiality, integrity, or availability of personal data to a data principal;

3.23 "**Personal Health Identifier**" or "**PHI**" is the data that could potentially identify a specific data principal and can be used to distinguish such data principals from another. PHIs could also be used for re-identifying previously de-identified data. It could include a data principal's demographic and location information, family and relationship information, and contact details; An anonymous identifier that cannot be used to re-identify previously de-identified data would not be treated as a "**PHI**";

3.24 "**Personal Health Record**" or "**PHR**" is a health record that is initiated and maintained by an individual. An ideal PHR would provide a complete and accurate summary of the health and medical history of an individual by gathering data from many sources and making this accessible online. Generally, such records are maintained in a secure and confidential environment, allowing only people authorized by the ICGA, to access the medical data; under ICGA, the PHR will be converted to data used for analytics and shared openly, in an anonymized manner, which does not include PHI.

3.25 "**processing**" in relation to personal data, means an operation or set of operations performed on personal data and may include operations such as collection, recording, organization, structuring, storage, adaptation, alteration, retrieval, use, alignment, or combination, indexing, disclosure by transmission, dissemination or otherwise making available, restriction, erasure or destruction;

3.26 "**processed data"** means the raw data that has been modified/processed/refined to obtain certain information;

3.27 "**raw data"** means primary data collected from a source and not processed/refined. For example, data coming out from a DNA sequencer will be considered raw data;

3.28 "**security"** refers directly to the protection of data, and specifically to the means used to protect privacy and use of data;

3.29 "**sensitive data**" means data that are sensitive from a personal standpoint (eg., racial or ethical origin, some health or behavior-related data) which may reveal, be related to, or constitute— (i) financial data; (ii) health data; (iii) official identifier; (iv) sex life; (v) sexual orientation; (vi) biometric data; (vii) genetic data; (viii) transgender status; (ix) intersex status; (x) caste or tribe or ethnicity; (xi) religious or political belief or affiliation;

3.30 **"sharing"** means facilitating access and use by users;

3.31 **"users"** refers to individuals or organizations who make use of the data by accessing it from a repository after obtaining necessary permissions, if required.

4. Source and Types of Data

Defining all biological data types generated by biotechnological methods is a challenging task due to their vast and continuously evolving nature, particularly with the emergence of high-throughput technologies. Nevertheless, there are certain broad categories that can be used to classify such data, although these should not be considered exhaustive. ICGA is provided with various types of biological and/or molecular data that are generated using biotechnological methods. The data is classified into broad classes which provides an overview of the types of data.

4.1 **DNA Sequence Data** – Such data can be at the level of a whole genome, or single genes. Such data can be a single sequence or multiple fragmented sequences from a genomic region with a high depth of coverage (such as those generated by a massively parallel DNA sequencer).

4.2 **RNA Sequence Transcriptomic Data** – The nature of the data is similar to those generated by a massively-parallel DNA sequencer since usually cDNA synthesis is performed before sequencing.

4.3 **Genotype Data**- Modern methods use high-density microarrays to genotype individuals at a large number of loci spread across the entire genome. Genotyping by sequencing (GBS) is being increasingly used for genome-wide association studies. However, for various specific purposes, small-scale genotyping using PCR-RFLP and other similar technologies continues to be used.

4.4 **Epigenomic Data**- These data are also primarily generated using a BeadChip (that is similar to a DNA microarray). However, epigenomic data may also be generated using sequencing methods after bisulfite conversion.

4.5 **Protein Structure Data**- Atomic coordinates and other information that describes a protein and other important biological macromolecules comprise such data. These data provide 3D shapes of proteins, nucleic acids, and complex assemblies that help understand various aspects of protein synthesis under different conditions.

4.6 **Mass Spectrometry Data**- Mass spectrometry (MS) is a key analytical technology in current proteomics and mass spectrometers are widely used to generate data that allow protein identification, annotation of secondary modifications, and determination of the absolute or relative abundance of individual proteins.

4.7 Imaging Data- Images of individual cells, organs, or body parts, for example, chest X-rays or images of human eyes or mouth cavity.

4.8 Metabolome Data- Metabolomics is increasingly used in conjunction with microbiome data to better understand host-microbiome interaction. Small molecule metabolite patterns are generated using either LC-MS, GC MS, or CE MS.

4.9 Microbiome Data - These data are also nucleic acid sequence data and currently are of three major subtypes (a) Amplicon sequencing data from which specific groups of microorganisms present in any sample (e.g., human stool, soil, sediment, etc.) can be identified, or (b) Shotgun metagenomic sequence data that allows comprehensive assessment of all microbial organisms present in a sample and (c) genome sequences of individual isolates In addition, there is also data in the form of individual gene sequences used for Multi Locus Sequence Analysis and Multi Locus Sequence Typing, or for taxonomic purpose like 16S rRNA, gyrase and many other genes.

4.10 Flow Cytometry Data – Flow cytometry is a technique used to detect and measure physical and chemical characteristics of a population of cells or particles. Flow cytometry data pertain to counts and multi-parameter profiles of different cells in a heterogeneous fluid mixture.

4.11 Clinical Data – Related to the patient demography, their medical and family history, co-morbidity profile, radiology images, clinicopathology, surgery, chemo-radiation, clinical outcomes, and oncological follow-up, as captured by the onco-clinician interfacing with the patient.

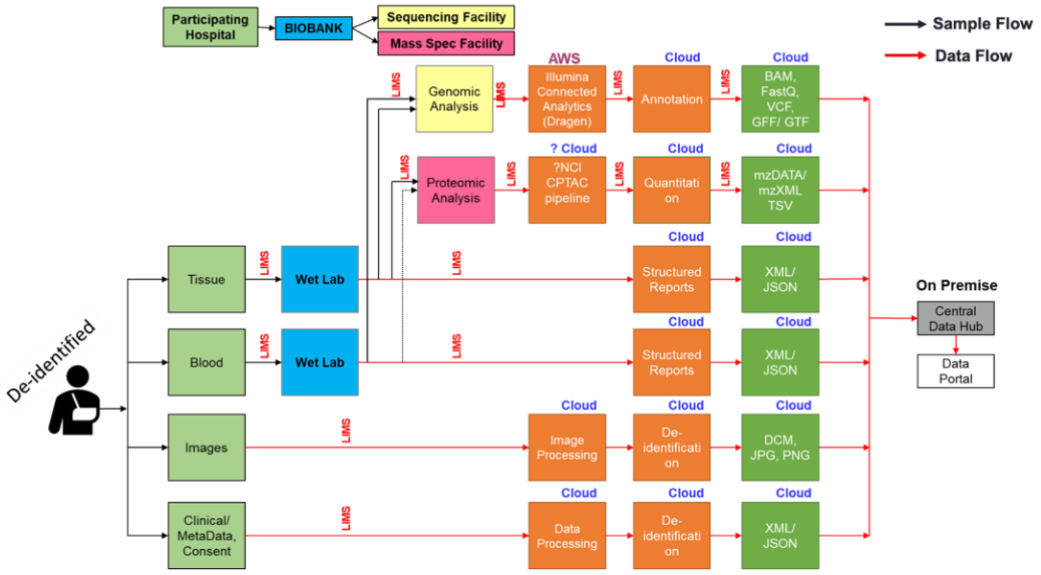
4.12 Any other data that may be relevant to the overall stated plan/aims of ICGA.

5. Flow of Data

ICGA is provided with de-identified clinical metadata and biospecimen samples from various authorized stakeholders across India. Such data is further processed and will be finally released in the public domain with applicable safeguards and made available to the research community. The flow of data mentioned below provides an overview of the nature of data, the processing activities, and the entities/organization involved:


Anonymized Clinical Data and / or Biospecimen Samples
The clinical metadata and biospecimen samples including but not limited to clinical metadata, blood samples, fresh/fresh-frozen tissue (tumor and healthy), Formalin fixed paraffin embedded (FFPE), and preliminary pathology review are ethically collected from cancer hospitals/clinics in India and shipped in a secure manner for pathology review,QC, biobanking and molecular studies. Study identification codes will be generated at local sites for de-identification of patients.
Biobank
The samples received are revalidated by onco-pathologist. Activites related derivatisation of DNA, RNA, Protein analysis, biospecimen QC analysis, biobanking of images, tissues, DNA, RNA are performed and derivatives/analytes sent to multi-omics sequencing labs.
Multi-Omics Labs
On receiving the QC passed molecular analytes, processes related to NGS and generation of raw big data, storage and archival commences. Such data is sent for data management and analytics to the IT partners.
Data Management and Analytics
The IT partner(s) using a variety of tools will conduct RAW to Processed data conversion for use in Analytics by users of the ICGA Portal. The partner(s) will handle the data storage, processing, retrieval, access control and publicly visible web portal, as well as internal applications for clinical data collection, sample/ specimen/ information tracking, and general IT such as emails.
Open Source Access to Global Research Community
The processed data is released in the public domain to share with cancer consortia with applicable safeguards.

Figure 1-Draft detailed data flow in various process types



6. Data Processing

There are Four key stages involved here:

6.1 Clinical metadata Collection & Storage: Clinical Data will be collected by the participating organizations and de-identified data will be stored preferably via an ICGA-provided interface. This interface will allow input, storage, and retrieval of data from a centralized ICGA setup hosted in India. Patient history and case details can be retrieved by authorized clinicians. Further, for existing patients, it is envisaged that patient history will be provided to the clinician in an easy-to-use manner.

6.2 Sample Tracking: Applies to Physical Samples and Digital Data being exchanged among members of the ICGA.

6.3 Data Processing (Anonymization): Bioinformatics and related data processing pipelines, clinical data conversion to anonymized metadata.

6.4 ICGA Portal: Access Requests, Access Governance, Monitoring, and Reporting on the data access taken by members and registered users of the ICGA.

6.5 Any other processing/analysis that may be relevant.

7. Data Processing Principles

The stakeholders of ICGA which deal with or process personal data including health data, genetic data, and sensitive data shall abide by the following data processing principles:

7.1 Fairness and Transparency - All stakeholders, including employees, who have access to personal data need to process such data lawfully, fairly, and in a transparent manner in relation to the data principal. In all cases, the processing must be for the purposes for which the data is collected. In case any employee/ stakeholder has any queries in relation to this Policy, they may approach the Compliance Officer who may approach any member of the DPCT in case the Compliance Officer is unable to answer the queries.

7.2 Instruction and Internal Policies – The stakeholders, including employees, shall act in accordance with the instructions given by the Compliance Officer in relation to the collection, use, retention, transfer, disclosure, and destruction of any personal data given by the Compliance Officer and the Policy.

7.3 Data Retention - The employees/ stakeholders shall not retain any personal data for any longer than is necessary in light of the purpose(s) for which that data is collected, held, and processed. Different types

of personal data will necessarily be retained for different periods (and its retention periodically be reviewed at minimum 12-month intervals), as set out in the data retention policy. Compliance Officer shall be in charge of maintaining the data retention schedules as per the data retention policy.

7.4 Data Erasure - The personal data needs to be erased upon expiry of the data retention period, request received by the data principal, or the personal data is no longer necessary in relation to the purpose for which they were collected. Upon receiving any request from the data principal for erasing personal data. However, this does not apply to data that has egressed the ICGA IT environment in an authorized/ approved manner.

7.5 Controls on Data Sharing- The employees/ stakeholders shall not transfer/ share any personal data internally with other employees, members of any other department, third parties, and outsiders unless instructed by their Compliance Officer, who in turn, must have attained the necessary approvals.

7.6 General Data Protection Measures- The employees/ stakeholders shall adopt the following minimum data protection measures as given below to ensure the security of personal data:

7.6.1 The employee/ stakeholder shall prevent unauthorized persons from gaining access to data processing systems in which personal data are processed.

7.6.2 The employee/ stakeholder shall prevent persons entitled to use a data processing system from accessing personal data beyond their needs and authorizations.

7.6.3 The employee/ stakeholder shall ensure that the personal data is protected against undesired destruction or loss.

7.6.4 The employee/ stakeholder shall ensure adequate security of personal data and take appropriate measures against unlawful processing.

7.6.5 The employee/ stakeholder shall inform his/ her supervisor/ officer in charge and the Compliance Officer immediately after being aware of any violation of this Policy.

8. Triggers for Data Disposal

The trigger events for the destruction/ deletion of personal data are as follows–

8.1 If the data principal has made a direct request for deletion of personal data.

8.2 If the purpose for which the data was collected has been completed.

8.3 When the data fiduciary has requested deletion of the personal data.

The Compliance Committee shall be in charge of determining whether a trigger event has occurred or not. The above-mentioned trigger events pursuant to the deletion of personal data shall not be applicable to anonymized data.

9. Data Disposal Methods

The employees/stakeholders shall upon the happening of any of the trigger events specified in Section 10 ("**Triggers for Data Disposal**") delete, destroy, or otherwise dispose of as follows under the supervision of the Compliance Officer:

9.1 personal data stored electronically (including any and all accessible backups thereof) shall be deleted securely;

9.2 personal data, if stored in hardcopy form, shall be shredded.

10. Roles and Responsibilities of the Compliance Committee

10.1 The Compliance Committee shall be in charge of identifying the personal data being handled by employees/stakeholders and preparing the requisite Data retention schedule. A format is provided in the data retention policy.

10.2 The Compliance Committee shall also determine the manner in which the personal data is maintained, accessed, and shared.

10.3 The employees/ stakeholders shall be responsible for informing the Compliance Committee as and when they believe a Data Disposal Trigger event has been achieved. The Compliance Officer shall assess the situation and determine whether such trigger event has actually been achieved.

10.4 Once a Data Disposal Trigger has been achieved the Compliance Committee shall ensure that the specific personal data is to be disposed of as per this Policy.

11. Data Access Rights

Detailed information about the data access rights, privileges, and restrictions applicable to various stakeholders pertaining to access of data within the ICGA network, and outside the network including data with third parties will be described once the roles and responsibilities are adequately defined. To enable data sharing with suitable measures and security, the ICGA will frame a detailed Data User Agreement as appropriate for the type and category of data.

12. Managing Open Access to Data

ICGA is committed to facilitating open access to data sets while ensuring maximum security and privacy measures. The open-access frameworks will be designed basis the principles of transparency, accountability, and responsibility.

ICGA will manage and design open access frameworks with the utmost care, implementing robust security protocols to safeguard against unauthorized access and data breaches. Data sets will be de-identified or anonymized to protect patient privacy, and access will be granted only to authorized stakeholders through authentication and access controls. Ongoing monitoring and auditing will ensure that data sets continue to meet the latest security standards and best practices.

13. Data Release and Exemptions

The release of data and access to information related to research and records to the members of ICGA shall be in accordance with the membership rules, Memorandum of Association (MOA), and Articles of Association (AOA) of ICGA.

14. Data User Agreements

On request, ICGA will provide access to raw or processed de-identified data including but not limited to clinical metadata, and multi-omics data to its members or any third parties after approval by the Compliance Committee with appropriate justification. This process will be performed by executing a data user agreement ("**DUA**"). Processing of such data by ICGA members and third parties shall be in accordance with the terms of the DUA.

The following terms shall be included in the Data User Agreement (**DUA**):

14.1 Acknowledgment: Users must agree to acknowledge the ICGA in all oral and written presentations, disclosures, and publications resulting from any analyses of the data.

14.2 Prohibition on re-identification of individuals using shared human data: In accordance with the Government of India guidelines and ICGA policies, re-identification must not be attempted, and appropriate legal provisions will be imposed via the DUA.

14.3 Purpose of access: The user requesting access to data must clearly state the purpose of access while applying.

14.4 Confidentiality and security of shared data: The application for accessing data must clearly describe the plan to uphold the confidentiality of the data and the security of the data to prevent access by unauthorized users.

15. Withdrawal of Sample Data

Details will be provided about the procedure to be followed in the event the donor request for the return/deletion of its data and the processes involved in complying with the request received from the donor.

16. Data Security

ICGA is committed to implementing rigorous security measures and taking full responsibility for the protection of any data we process. ICGA undertakes the following activities to ensure data security:

16.1 Implement robust security protocols to protect against unauthorized access, data breaches, and cyber-attacks.
16.2 Regularly assess and update our security measures to ensure they meet the latest standards and best practices.
16.3 De-identify or anonymize data sets to protect patient privacy.
16.4 Limit access to data to authorized individuals through authentication and access controls.
16.5 Monitor and audit data access and usage to identify and address any potential vulnerabilities or breaches.
16.6 Take steps to ensure stakeholders understand and comply with our data security policies and procedures.
16.7 Provide training and resources to our stakeholders to ensure they understand their role in data security.

ICGA takes extensive measures to protect the privacy and security of data. It will take prompt action to address any issues that may arise and continue to improve its security measures to minimize the risk of data breaches.

17. Data Audit

ICGA and its partners are responsible for the collection and processing of data. The data will reside solely in ICGA's enabled facility for sharing and providing access with applicable safeguards. A proper audit trail of record creation, access, updating, and deletion of clinical and other forms of important data attributes relevant to ICGA would be maintained. Similarly, all configurations and configuration changes should be maintained in an accessible and wherever possible a replicated form for audit, traceability, and compliance purposes. The Information Technology internal assets of ICGA would be subject to information annual security audits. Further, the data security posture at ICGA will be designed with industry standards such as ISO 27001 or newer standards from time to time as the guiding principles as would be guided by the Board.

18. Data Protection Regulations and Compliances

The data processing activities of ICGA will be governed to the extent the mandatory and applicable by-rules and regulations under the Indian legal framework including but not limited to ICMR guidelines, The Information Technology Act 2000 ("IT Act"), Information Technology (Reasonable security practices and procedures and sensitive personal data or information), Rules 2011, Data Protection of Personally Identifiable Information, and new policies updates, circulars, rules and regulations issued by the Authorities from time to time that can apply to healthcare specific data attributes.

19. Use of Aadhaar Identification

If Aadhaar identification information is used in any manner, then ICGA will follow the compliances set out under the Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits and Services) Act, 2016 and

Unique Identification Authority of India (UIDAI) Information Security Policy; along with other applicable Aadhaar related rules, regulations, circulars and notifications.

20. Data Breach

In any event of identifiable or non-anonymized personal data/ PHI/ PHR breach that has occurred due to theft of personal data or any other reason, the employee/stakeholders shall immediately inform the Compliance Committee and provide them with all the required information about the data breach incident. The Compliance Committee shall inform the key stakeholders to undertake necessary mitigation and reporting activities.

20.1 Contents of data breach notification-

20.1.1 Nature of personal data breach;
20.1.2 Categories and approximate number of data principals concerned;
20.1.3 Categories and approximate number of personal data records concerned;
20.1.4 Description of the likely consequences of the data breach;
20.1.5 Measures taken or proposed to be taken by data processor/data fiduciary to address the personal data breach and measures to mitigate its possible adverse effects.

21. Penalties

ICGA Foundation is committed to complying with all applicable data protection laws and regulations of the Government of India. This policy applies to all stakeholders, including employees, contractors, consultants, partners, and affiliates of ICGA who handle personal data in any form or manner.

Any stakeholder who fails to comply with this policy or any other data protection policies issued by ICGA will be subject to penalties under the applicable data protection laws and disciplinary actions by ICGA. The course of action will be guided by the Compliance Committee of ICGA and will follow the Government of India's data protection laws and regulations.

Annexure A to ICGA Data Policy

Data Request Information

I. Controlled Data:

1. The Controlled data includes information that is unique to an individual (Please note, the individual is always de-identified in data sets). This includes most raw data files and some processed data such as:

    - Raw clinical metadata and digital images
        - Primary sequencing data (BAM and FA1STQ files) from DNA, RNA, miRNA, or bisulfite sequencing studies
    - Raw and processed SNP6 array data
    - Raw and processed Exon array data
    - Somatic and germline mutation calls for an individual (VCF and MAF files)

2. Guidelines for Member Organizations to access Controlled Data

a) Member organizations while accessing the data will ensure that only authorized individuals will have access to the data.
b) Member organizations need to designate a contact person with the responsibility to oversee compliance and authorization of individuals within the organization.
c) Member organizations must ensure that adequate security measures are implemented with respect to the data accessed, there is no attempt to identify or contact individual participants from whom these data were collected without appropriate approvals, the data must not be shared without due authorization, and maintain compliance with guidelines in the ICGA Policy.

II. Open Data:

1. Open Data includes information that is not unique to an individual. This includes information such as:

    - De-identified clinical and demographic data
    - Imaging archive
    - Gene expression data
    - Copy number alterations in regions of the genome
    - Epigenetic data
    - Summaries of data across individuals

2. Guidelines to access Open Data
   Organizations and individuals must ensure that adequate security measures are implemented with respect to the data accessed, there is no attempt to identify or contact individual participants from whom these data were collected without appropriate approvals, the data must not be shared without due authorization and to maintain compliance with guidelines in the ICGA Policy.

III. Acknowledgment to ICGA

Users must agree to acknowledge the ICGA in all oral and written presentations, disclosures, and publications resulting from any analyses of the data. The following format may be used for this purpose:

The results <published or shown> here are in whole or part based upon data generated by the ICGA Network: https://www.icga.in.