**** **ISO/IEC JTC 1/SC 29/AG 3 N0036**

**ISO/IEC JTC 1/SC 29/AG 3  
MPEG Liaison and Communication   
Convenorship: KATS (Korea, Republic of)**

**Document type:** Output Document

**Title:** **White paper on MPEG-G and its application of regulations and privacy**

**Status:** Approved

**Date of document:** 2021-08-03

**Source:** ISO/IEC JTC 1/SC 29/AG 3

**Expected action:** None

**Action due date:** None

**No. of pages:** 6 (without cover page)

**Email of Convenor:** kyuheonkim@khu.ac.kr

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29ag3>

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/ AG 03 MPEG LIAISON AND COMMUNICATION**

**ISO/IEC JTC1/SC29/AG 3 N0036**

**July 2021, Online**

|  |  |
| --- | --- |
| **Title** | **White paper on MPEG-G and its application of regulations and privacy.** |
| **Source** | **ISO/IEC JTC1/SC29/WG8** |
| **Status** | **Approved** |
| **Serial number** | **20737** |
| **Date** | **2021-08-03** |

# Introduction

The current MPEG-G standard series (ISO/IEC 23092) addresses the representation, compression, and transport of genome sequencing data, metadata and associated annotation data. While the standard has been developed, privacy and security regulations across the world have and are still evolving. Since the MPEG-G standard has the goal to be an important component of infrastructures handling genomics data, it is also expected to be able to support different privacy and security regulations. Existing genomics data representation methods do not support interoperable and standard privacy and security protection of the various data across the genomics workflow. It is one goal of MPEG-G to provide such support.

This white paper consists of a survey of regulations applicable to genomic information in different countries, requirements of security and privacy, and technological tools. It is based on the document “Survey of the regulations of genomic information in different countries, demands of security and privacy and technological tools, for the application guide for MPEG-G - Genomic Information Representation[1] ” which provides additional information and details.

Whole-genome sequencing is fast spreading and is exceeding the pace of Moore’s law. We are approaching the point where whole-genome sequencing will be available to everyone and could be used on a regular basis for the personalized diagnosis and treatment of diseases. Because each genome is specific to each person, it is imperative that this information has to be treated as protected personal medical information. The privacy of the whole-genome sequence and associated information is indeed an especially important issue.

# Evolution of whole-genome sequencing

## Cost reduction in DNA sequencing technique

DNA sequencing technology was first developed in the 1970s. The first fully analyzed human chromosome was published in 1999. The first sequencing of the entire human genome was published by 2001.

In the 20 years after the first whole-genome sequencing, sequencing costs have radically reduced to less than one thousand US dollars today, which is expected to keep falling. The commercialization of whole-genome sequencing is fast developing, and it is expected to be a common and popular process, accessible to everyone in the near future.

## Current and future security function of MPEG-G.

MPEG-G already has built-in functionality to protect the sensitive parts of genomic data. Just as other MPEG standards, MPEG-G is also designed to be flexible and extensible in supporting the usage of current and new standard cryptographic tools. It’s well established, trusted maintenance process ensures that the security functionality of MPEG-G can be updated to meet the changing security requirements of industries and governments, covering use cases in multiple countries in an integrated and interoperable manner.

# Regulations on the privacy of genomic information

Genomic data privacy is an important issue in the digital transformation of recordkeeping including both genome information as well as healthcare records. This white paper summarizes the various regulation for genomic information already established and put in force in various countries.

## Regulation in EU

*GDPR and Genomic Information.*

The GDPR (General Data Protection Regulation) [2] enacted on May 25, 2018, includes items related to the privacy and security of genomic data. Shabani [3] explains the effects of GDPR on genomic data as follows: "any information relating to an identified or identifiable natural person ("data subject")". However, in the catalog of identifiers, the definition provided by the Regulation includes “genetic” (Article 4.1) data, which was not included in the Directive's definition of personal data. Although “genetic” has been generally included as an example of identifier factors, one can consider that this will only apply to identified genetic factors.

*Anonymity and Pseudonymization.*

In the GDPR, anonymity and pseudonymization are given as a condition for exemption to the regulations. Data after anonymization may be outside the scope of the GDPR. In the case of the data after pseudonymization, the data may still be regulated by the GDPR, but some of the regulations are mitigated.

EU GDPR also restricts the use of personal information for profiling. Modern profiling is restricted since it is based on the automatic analysis of massive data, associated with scientific evidence, and in many cases having unexpected effects for the private data subject (person who is the origin of the data).

## Regulations in Japan

*Anonymity and Pseudonymization.*

In Japan, the Personal Information Protection Law was established in 2003. Following enactment of this law, each local government has formulated a Privacy protection "ordinance", and the regulations in Japan are different depending on the organization and place. There are about 2,000 variations, each slightly different, of the Personal Information Protection Ordinances. Meanwhile, in 2019, the EU and Japan each issued an adequacy decision on their respective regulations on this topic. This means the status of Japan at large is accepted to be compliant with the EU regulations and vice versa. Therefore, it is thought that private and the public organization can flexibly respond to these various privacy regulations according to the rules in each region and the business content in Japan and the EU, respectively.

*Adequacy Decision Between EU and Japan.*

Between Japan and the EU, Japan has been certified to have appropriate data protection levels to be able to transfer personal data, which may include the exchange of genetic information. In this way, The GDPR is also a possible common standard to regulate the privacy of genetic information as described above. Itakura, Terada, and others reported on the issues and prospects of the GDPR accreditation in Japan. [4] [5] [6] .

*Secondary use of EHR.*

In Japan, the secondary use of health information is planned [7] The anonymity process is assumed in such a system.

The law for protection of personal information is revised again in 2018.

Genomic information is interpreted as an individual’s identifier in Japan as described in a cabinet order in 2003[8] . Even if a part of the genomic information is not sufficient to identify individuals, such information cannot be delivered as pseudonymized information according to the current law. Some experts think it should be reconsidered because part of DNA is not possible to identify an individual[9] .

## Regulations in the USA

National Human Genome Research Institute under NIH has detailed information explaining privacy in genomics. The website is available in the following URL. [10] [11] .

Privacy in Genomics. <https://www.genome.gov/about-genomics/policy-issues/Privacy>

On this site, privacy in research is first described as follows: “When conducting genomics research, two essential values of science research need to be balanced - the need to share data broadly to maximize its utility for ongoing scientific exploration, and the need to protect research participants' privacy.” This aligns with broad consensus on privacy and scientific merit.

In the current research protocol, DNA samples are exchanged in such a way that the individual cannot be identified by name. However, it has been shown that there is a possibility that the individual may still be identified from the gene according to the report of Gemrek [12] . In this way, privacy is not protected. sufficiently. Therefore, a balance of the benefit of research and privacy is necessary.

*Common Rule.*

Federal Regulations Governing Human Subjects Research, aka “Common Rule”, published 1991, establishes the standard of ethics for government-funded research in the United States [11] and it is also applied to research on genetic information. Federally funded research projects are required to obtain informed consent from each participant prior to their participation.

*Genetic non-discrimination law (GINA) and Other Law.*

GINA is a United States law, enacted in 2008, to protect genetic privacy. The Health Insurance Portability and Accountability (HIPAA) regulates genetic information as health information, which is individually identifiable, as it may not be used or disclosed for underwriting purposes. Certificates of Confidentiality are issued by NIH and can be used to safeguard the research participants. Newborn Screening Saves Lives Act Reauthorization was signed in 2014 and requires parental consent for the use of newborn blood spot samples in research.

# Technological aspects in Privacy Protection

## Secondary use

The term secondary use is often used in the regulations which refers use of the privacy information once captured for the primary usage (e.g., health record) and then used as database for the statistical analysis. Accumulation of genomic information and use of statistical analysis is recognized as secondary use and is not allowed without complying with related regulations.

## Anonymization and pseudonymization

Both anonymization and pseudonymization are referred to as the process to remove information to be used to identify individuals from the target data. Those two processes provide different level of anonymity. They are distinguished in the regulations and technologies.

## Various encryption technique

There are a variety of encryption technique available and can be used to secure the privacy information. In addition to standard encryption algorithms e.g., public key or shared key encryption, there are emerging technologies e.g., Homomorphic encryption, Partial encryption, , Boolean tensor representation as well as ongoing research in encryption techniques that can be a candidate solution for specific requirements in Privacy Protection of genomic information. Refer to [1] for further details.

# MPEG-G as infrastructure to manage privacy

One of the essential benefits of MPEG-G, having a standard for genomic information representation, is exactly the ability to balance the privacy of the genetic data, the personal information protection legislation, and the public interest by the medical use of the gene information. This has not been possible using the existing genomics workflow tools which each store genomics data in their own format.

The management of the gene information as currently implemented uses the approach that the generator of the gene sample protects the information because the gene sample might be identified personally. Because there is no thought currently given to downstream secondary and tertiary use of the data, this approach is not optimal. The result is that it is very difficult to safely share and analyze a large amount of gene data. MPEG-G, which makes use of a single data representation method, provides a single data file with support for existing privacy regulations and supporting secondary and tertiary use.

In terms of the privacy law system, it is thought that it is necessary to treat gene information as personal information because the DNA sample contains enough information to identify the individual. However, if the exchange of DNA samples is regulated as if they are regular personal data the cost will be too large compared to the benefit.

It has also to be remarked that, it seems that personal information obtained from DNA is a piece of virtually semipublic information at its source, even if analyzed results are protected. Whether it is compliant with the law or not, it is virtually impossible to restrict someone from getting cells from others to analyze their genes. There is no means to confirm if the DNA sample was acquired properly and if the analysis is requested for an appropriate sample. Once genomics data is encapsulated within MPEG-G what can be assured is that privacy and security is maintained in a consistent manner throughout the genomics workflow, from pipeline processing through secondary and tertiary analysis.

Although the privacy of DNA should be properly protected, we think that it is necessary to optimize the implementation of the restrictions in consideration of the usefulness of DNA characteristics and the information obtained by DNA analysis.

# Conclusion

Since the first human genome analysis was achieved in 2003, the speed of gene decoding continued to progress phenomenally, and we have now entered an age when gene decoding is trusted by individuals. The cost of gene decoding is expected to drop rapidly in the future. While there are privacy concerns, if the results of genetic analysis and other statistical information can be effectively combined appropriately, the potential effect is immense. At the same time, negative social issues such as the invasion of privacy and genetic discrimination need to be prevented as much as possible. MPEG-G, through the incorporation of support for privacy and security, can help to protect the privacy of genomic data while also allowing for the application of genomic data for advanced diagnosis and treatment. We hope this White Paper will encourage the reading of MPEG WG8 output document “A survey on the regulations and privacy issues in the application of MPEG-G - Genomic Information Representation (Draft 1)[1] ”.

# References

1. ISO/IEC JTC1/SC29/WG8, A survey on the regulations and privacy issues in the application of MPEG-G - Genomic Information Representation (Draft 1), ISO/IEC JTC1/SC29/WG 8 N00054. (2021)
2. Wikipedia. 2019. General Data Protection Regulation. https://en.wikipedia.org/wiki/General\_Data\_Protection\_Regulation, 2019-03-18 20:35 (UTC).
3. Shabani, M. Borry, P. 2018. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. European Journal of Human Genetics, Vol. 26, Pages 149-156. https://www.nature.com/articles/s41431-017-0045-7.
4. Itakura, Y. Fujimura, A. Kameishi, K. Magata, F. 2019. Legal position of SECURE MULTI-PARTY COMPUTATION in foreign countries. Information Processing Society of Japan, 2019-EIP-83(1),1-4 (2019-02-08), 2188-8647. Permanent link: Permalink: http://id.nii.ac.jp/1001/00194182/
5. Itakura, Y. Terada, M. 2019. Present Situation and Prospect of the EU's "Adequacy Decision" -A Study Based on the Opinions of the European Data Protection Board (EDPB). Information Processing Society of Japan, 2019-EIP-83(2),1-4 (2019-02-08), 2188-8647. Permalink: Permalink: http://id.nii.ac.jp/1001/00194183/
6. Terada, M. Itakura, Y. 2019. Current Situation and Issues over the Draft e-Privacy Regulation in the EU―Based on its relation to the GDPR (General Data Protection Regulation) ―. Information Processing Society of Japan, 2019-EIP-83(3),1-6 (2019-02-08), 2188-8647. Permalink: http://id.nii.ac.jp/1001/00194184/
7. Yoshihara, H. 2016. EHR Project: Nation-wide EHR Implementation in JAPAN. Kyoto Smart City Expo. https://expo.smartcity.kyoto/2016/doc/ksce2016\_doc\_yoshihara.pdf.
8. Cabinet order: Individual identifier (in 2017), https://www.ppc.go.jp/files/pdf/290530\_personal\_cabinetorder.pdf
9. Yuki Kuroda, Possibility of Medical research using pseudonymized medical data, Julist, winter 2018, p. 120.
10. NIH. 2019. Privacy in Genomics. https://www.genome.gov/27561246/privacy-in-genomics/. acquisition on March 19, 2019.
11. NIH. 2019. Privacy in Genomics. https://www.genome.gov/about-genomics/policy-issues/Privacy. acquisition on April 23, 2019.
12. Gymrek, M. et al. 2013. Identifying Personal Genomes by Surname Inference. Science 18 Jan 2013: Vol. 339, Issue 6117, pp. 321-324, DOI: 10.1126/science.1229566. https://science.sciencemag.org/content/339/6117/321.