**A black and white logo

Description automatically generated ISO/IEC JTC 1/SC 29/WG 2 N0345**

**ISO/IEC JTC 1/SC 29/WG 2  
MPEG Technical requirements   
Convenorship: SFS (Finland)**

**Document type:** Output Document

**Title:** Draft use case and requirements for adaptable private data management of genomic information representation on the large-scale cloud environment.

**Status:** Approved

**Date of document:** 2024-01-26

**Source:** ISO/IEC JTC 1/SC 29/WG 2

**Expected action:** None

**Action due date:** None

**No. of pages:** 13 (with cover page)

**Email of Convenor:** igor.curcio@nokia.com

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg2>

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 2**

**MPEG TECHNICAL REQUIREMENTS**

**ISO/IEC JTC 1/SC 29/WG 2 N00345**

**Online – January 2024**

|  |  |
| --- | --- |
| **Title** | **Draft use case and requirements for adaptable private data management of genomic information representation on the large-scale cloud environment.** |
| **Source** | **WG 2, MPEG Technical Requirements** |
| **Status** | **Approved** |
| **Serial Number** | **23691** |

# Abstract

This document was produced jointly by members of WG08 (MPEG Genomic Coding) and domain experts and is intended for presentation to WG02.

At the 2023/1 meeting, a proposal towards adaptable private data management of genomic information representation on large-scale cloud environment was discussed, and as the next step, it was planned to produce use case and requirement documents.

There is growing interest in the dynamic consent of privacy in biomedical information overall and for genomic information. And given today’s cloud environment, more advanced system design might be possible to directly and promptly reflect such dynamic consent to access the results of analysis. Such mechanism gives greater trust for the data subject (provider of genomic information and health information) to provide their private data.

We have compiled use cases and requirements as a reference to guide future work. This document summarizes the use cases and requirements for such a system to generate broader interest and gather feedback.

# Introduction

The importance of protecting biometric and genomic information privacy is increasingly recognized. However, there is currently no stable and widely agreed-upon policy for managing privacy in multilateral settings, such as international relations[1-9]. To address this issue, WG8 is discussing a proposed approach for a generic, flexible, and portable cloud-based adaptive decentralized protection scheme that can provide adaptive personal data protection in various changing environments [10-14]. This approach aims to provide a comprehensive solution to the complex challenges of protecting sensitive personal information while enabling its flexible use in diverse settings.

To generate broader interest and feedback, this document provides a summary of the use cases and requirements for such a system.

# Use cases

1. **Filtering genomic data based on a dynamic consent for research**   
   This use case refers to conducting research on a cohort of genomic data, which can be shared, exchanged, merged, or filtered based on the dynamic consent of data subjects who provided the genomic samples.
2. **Nationwide electronic health records that include genomic data**This use case involves using nationwide electronic health records that include genomic data to perform big data analysis aimed at improving the health of the general public.
3. **Open-source and/or Interoperable software for genome information repository**   
   This use case refers to an open-source software for genome information repository that allows for the exchange of dynamic consent databases among consortiums and others, and which can be linked with consent databases.
4. **Genomic data base that is designed specifically for pharmaceutical research.**  
   This is a use case of subject database containing genomic data base that is designed specifically for pharmaceutical research.
5. **Multiple countries, seeks to facilitate the exchange of health data including genomic data.**  
    This use case is about treaty organizations that comprise multiple countries and seeks to facilitate the exchange of medical data, including genetic data, among them, while maintaining independent regulations to protect human rights.
6. **Long term protection of sensitive information**  
   This use case relates to the need for appropriate protection of human rights when sensitive information is discovered during the accumulation of genome information for various research purposes or in the public domain.

# Detailed description of use cases

## Use case 1 - Filtering genomic data based on a dynamic consent for research.

A. Use case presentation

|  |  |
| --- | --- |
| A – Use case presentation | |
| A.1 | Please describe the proposed use case illustrating the user experience |
| *This use case refers to conducting research on a cohort of genomic data, which can be shared, exchanged, merged, or filtered based on the dynamic consent of data subjects who provided the genomic samples.*  *This use case proposes the creation of a comprehensive and centralized genomic data repository to facilitate research efforts. The repository would encompass diverse data sources with varying dynamic consents. It is envisioned that the system would enable efficient identification and selection of relevant data for research purposes, while ensuring robust security measures to maintain data integrity and privacy. Such a system would greatly advance research efforts in genomics and drive scientific progress in the field.* |

B. MPEG Technologies

|  |  |
| --- | --- |
| B - MPEG technologies | |
| B.1 | Are there any available MPEG standards/technologies addressing the use case? |
| MPEG-G, MPEG-DASH, MPEG Smart Contract for Media  Note:  MPEG-G and MPEG-DASH are useful for efficient record transmission of genomic data for research. SmartContract helps classify research data. |

C. Additional information

|  |  |
| --- | --- |
| B  C – Additional information | |
| C.1 | Please provide any additional information that you think relevant (e.g. service architecture…) |
| GA4GH and ISO/TC 215 SC1 may be relevant.  fGA4GH is large community of potential users. And ISO TC215 standards is also major standardization community in this domain.  Following diagram shows hypothetical service architecture.     |  |  |  | | --- | --- | --- | | Data Subject provides genomic data and manage their consent | The system, which manage dynamic consent can extend new analyzer by the request of research. Privacy agent will evaluate analyzer and compliant personal information will be applied to the analyzer. New analysis method can be dynamically produce result. And resulting output is also protected by the DRM. | Analyzed data from new analyzer will be provided to the researcher. | |

Related requirements

|  |  |
| --- | --- |
| D – Related requirements | |
| D.1 | Please provide requirements |
| Requirement 1, Requirement 4, Requirement 6, Requirement 7, Requirement 9  Note:  Dynamic consent is required, especially for research data because each research has very specific purpose.  Requirement 1 The standard shall include functionality to assist systems using dynamic consent in demonstrating accountability for the proper implementation of standards-based systems.  .  These are necessary requirements for research data so that they can be used efficiently in various environments.  Requirement 4 The standardization shall define option to be compatible with relevant MPEG standards, such as ISO/IEC 23001-7 Common encryption in ISOBMFF, ISO/IEC 23009-x DASH, ISO/IEC 21000-23 Smart contract for Media, and others.  Requirement 6 The standardization shall ensure the possibility to comply with regulatory and privacy laws by incorporating machine-readable methods of expressing interactions (ISO/IEC 000000 right expression languages).  Requirement 7 The standard shall support different distributed processing environments (e.g., cloud, blockchains, webs, URLs) and be implemented in these environments. |

## Use case 2 - Nationwide electronic health records that include genomic data.

A. Use case presentation

|  |  |
| --- | --- |
| A – Use case presentation | |
| A.1 | Please describe the proposed use case illustrating the user experience |
| *This use case involves using nationwide electronic health records that include genomic data to perform big data analysis aimed at improving the health of the general public.*  *This use case presents a scenario where a national health database is utilized to incorporate genomic data for the purpose of promoting public health. The proposed health database would offer a valuable resource for statistical analysis, which can be utilized to enhance health promotion efforts. Additionally, the incorporation of genomic data would facilitate research on drug efficacy, thereby potentially leading to advancements in medical treatment.*  *Moreover, the implementation of such a health database can help alleviate concerns that users may have about the safety and privacy of their personal health information. By ensuring appropriate measures are taken to secure the data and maintain user privacy, users can be assured that their data is being used for the intended purpose of promoting public health.* |

B. MPEG Technologies

|  |  |
| --- | --- |
| B - MPEG technologies | |
| B.1 | Are there any available MPEG standards/technologies addressing the use case? |
| MPEG-G, MPEG-DASH, MPEG Smart Contract for Media Note:  MPEG-G and MPEG-DASH are useful for efficient storage and transmission of genomic data for national health record. De jure standard is good for this purpose. SmartContract helps classify research data. |

C. Additional information

|  |  |
| --- | --- |
| B  C – Additional information | |
| C.1 | Please provide any additional information that you think relevant (e.g. service architecture…) |
| Given that the national electronic health record system is integral to the health administration of each country, it is imperative to develop a standard that accounts for the significant variability in this system across different countries. In order to safeguard personal information related to genomic data, it is necessary to incorporate additional processing methods into diverse health information databases. Such methods would serve to protect sensitive information and ensure its appropriate usage.  Following diagram shows hypothetical service architecture.     |  |  |  | | --- | --- | --- | | Data Subject provides genomic data and manage their consent | The system is integrated with national health record but manage protection of genomic data from inappropriate use. Use of genomic data will be managed by privacy agent and may be controlled by dynamic consent of data subjects. | Government may use subset of health record for the analysis combining health data with genomic data. | |

D. Related requirements

|  |  |
| --- | --- |
| D – Related requirements | |
| D.1 | Please provide requirements |
| Requirement 1, Requirement 2, Requirement 3, Requirement 4, Requirement 5, Requirement 6, Requirement 7, Requirement 8, Requirement 9  Note:  All requirements are relevant in this use case because it is large database operated by public organization so all requirements are necessary. |

## Use case 3 - Open-source software for genome information repository.

A. Use case presentation

|  |  |
| --- | --- |
| A – Use case presentation | |
| A.1 | Please describe the proposed use case illustrating the user experience |
| *This use case refers to an open-source software for a genome information repository that allows for the exchange of dynamic consent databases among consortiums and others, and which can be linked with consent databases. This particular use case involves the creation of a tool, based on established standards, for a genomic data database that operates on an open-source platform. The development of such open-source tools significantly enhances the accessibility of genomic information, enabling diverse users to leverage and tailor it to their specific needs.* |

B. MPEG Technologies

|  |  |
| --- | --- |
| B - MPEG technologies | |
| B.1 | Are there any available MPEG standards/technologies addressing the use case? |
| MPEG-G, MPEG-DASH, MPEG Smart Contract for Media |

C. Additional information

|  |  |
| --- | --- |
| B  C – Additional information | |
| C.1 | Please provide any additional information that you think relevant (e.g. service architecture…) |
| As a reference software, this open-source tool offers a valuable resource for the standard under consideration. By presenting the core functionalities of the standard in compact and easily comprehensible implementation modules, the expectation is that the scope of the standard's applications will be significantly broadened.  Following diagram shows hypothetical service architecture.     |  |  |  | | --- | --- | --- | | Data Subject provides genomic data to various services. | The genomic database can be used any service and integrated into their system. | The service can also use analysis data but analysis is managed to adopt dynamic consent. | |

D. Related requirements

|  |  |
| --- | --- |
| D – Related requirements | |
| D.1 | Please provide requirements |
| Requirement 1, Requirement 2, Requirement 3, Requirement 4  Note:  The open source software is intended to provide adaptive privacy management therefore Requirement 1 is meant to necessary.  Requirement 1 The standard shall include functionality to assist systems using dynamic consent in demonstrating accountability for the proper implementation of standards-based systems.  There are all kinds of possible usage environments, and many other requirements may not be necessary, but it is surmised that Requirement 2,3,4 would be very useful in any case.  Requirement 2 The standard shall provide the capability to safeguard the analytical outcomes of genomic information, using dynamic consent, in the same manner as the genomic information.  Requirement 3 The standardization shall support the technologies that aid in dynamic consent and ensure the privacy of genomic data and metadata, satisfying the compatibility to the ISO/IEC 23092-x standard series (MPEG-G).  Requirement 4 The standardization shall define option to be compatible with relevant MPEG standards, such as ISO/IEC 23001-7 Common encryption in ISOBMFF, ISO/IEC 23009-x DASH, ISO/IEC 21000-23 Smart contract for Media, and others. |

## Use case 4 - Genomic data base that is designed specifically for pharmaceutical research.

A. Use case presentation

|  |  |
| --- | --- |
| A – Use case presentation | |
| A.1 | Please describe the proposed use case illustrating the user experience |
| *This is a use case of a subject database containing genomic data base that is designed specifically for pharmaceutical research. The aim of this use case is to propose a genomic database dedicated to pharmaceutical research. In this context, the genomic data of individuals can be leveraged for the development of new pharmaceuticals, and individual consent is obtained specifically for research purposes. Systematic processing of this data is deemed to be highly beneficial.* |

MPEG Technologies

|  |  |
| --- | --- |
| B - MPEG technologies | |
| B.1 | Are there any available MPEG standards/technologies addressing the use case? |
| MPEG-G, MPEG-DASH, MPEG Smart Contract for Media |

C. Additional information

|  |  |
| --- | --- |
| B  C – Additional information | |
| C.1 | Please provide any additional information that you think relevant (e.g. service architecture…) |
| PharmaLedger and other established information exchange formats are increasingly gaining traction in the pharmaceutical industry. To expand the reach of this standard, it is crucial to improve its interoperability with these widely-used formats. This would enhance the accessibility and integration of the standard within existing industry practices.  Ref - https://pharmaledger.eu/ |

D. Related requirements

|  |  |
| --- | --- |
| D – Related requirements | |
| D.1 | Please provide requirements |
| Requirement 1, Requirement 2, Requirement 3 |

## Use case 5 - Multiple countries, seeks to facilitate the exchange of health data including genomic data.

A. Use case presentation

|  |  |
| --- | --- |
| A – Use case presentation | |
| A.1 | Please describe the proposed use case illustrating the user experience |
| *This use case is about treaty organizations that comprise multiple countries and seeks to facilitate the exchange of medical data, including genetic data, among them, while maintaining independent regulations to protect human rights. As regulations pertaining to genomic data vary across different countries, the transfer of such data across borders is subject to different sets of rules and regulations. Given that genomic data constitutes personal data, the transfer of this information is subject to various restrictions. Therefore, the development of systems that facilitate the cross-border transfer of genomic data while ensuring compliance with relevant regulations is an essential use case for this standard.* |

B. MPEG Technologies

|  |  |
| --- | --- |
| B - MPEG technologies | |
| B.1 | Are there any available MPEG standards/technologies addressing the use case? |
| MPEG-G, MPEG-DASH, MPEG Smart Contract for Media |

C. Additional information

|  |  |
| --- | --- |
| B  C – Additional information | |
| C.1 | Please provide any additional information that you think relevant (e.g. service architecture…) |
| Given the sensitivity of medical and health data, it is important that transfers of such data across borders occur in a secure and appropriate manner. On the one hand, the ability to transfer such information across countries is necessary to protect the lives and health of individuals. To this end, each country establishes regulations that stipulate the acceptable forms of data transfer. In this context, it is imperative that standards define a set of interoperability requirements that genomic databases shall meet while adhering to the various rules and regulations governing cross-border transfers of personal information.  Following diagram shows hypothetical service architecture.     |  |  |  | | --- | --- | --- | | Clinical trial applicant is data subject in this case. They voluntarily provides genomic data for the analysis but may manage consent. | The genomic database can be used for the analysis of the clinical trial. Thus, clinical trial can be analyzed against genetical attribute. But the genetical attribute can be sensitive attribute in meantime. So adaptive privacy management is beneficial to execute clinical trial in more detailed but ethically appropriate. | Clinical analysis result can be managed appropriately. | |

D. Related requirements

|  |  |
| --- | --- |
| D – Related requirements | |
| D.1 | Please provide requirements |
| In this case, all requirements are relevant because such application need high level of compatibility.  Requirement 1, Requirement 2, Requirement 3, Requirement 4, Requirement 5, Requirement 6, Requirement 7, Requirement 8, Requirement 9 |

## Use case 6 - Long term protection of sensitive information.

A. Use case presentation

|  |  |
| --- | --- |
| A – Use case presentation | |
| A.1 | Please describe the proposed use case illustrating the user experience |
| *This use case relates to the need for appropriate protection of human rights when sensitive information is discovered during the accumulation of genome information for various research purposes or in the public domain. Given the rapid pace of progress in genomic research, it is difficult to predict the nature of the sensitive information that may be derived from genomic data in the future. Therefore, in order to safeguard human rights and protect against any potential negative consequences arising from the long-term sharing of genomic data, it is crucial to establish mechanisms for controlling the terms of consent on an ongoing basis, through the use of dynamic consent. In this context, the long-term protection of sensitive genomic information emerges as an important use case.* |

B. MPEG Technologies

|  |  |
| --- | --- |
| B - MPEG technologies | |
| B.1 | Are there any available MPEG standards/technologies addressing the use case? |
| MPEG-G, MPEG-DASH, MPEG Smart Contract for Media |

C. Additional information

|  |  |
| --- | --- |
| B  C – Additional information | |
| C.1 | Please provide any additional information that you think relevant (e.g. service architecture…) |
| The protection of personal information, particularly in cases where the value of the data changes over time, such as with genomic information, is still an evolving field. Both legal and technical developments in this area are expected to continue. In order to explore this use case, it is pertinent to consider and integrate various privacy protection technologies and the latest research pertaining to the privacy of genomic information.  Following diagram shows hypothetical service architecture.     |  |  |  | | --- | --- | --- | | In this scenario, anyone who wants to ensure long-term protection should use this system. | Therefore, all services, organizations, and laboratories that wish to work with those people's genomic data should also use systems of dynamic privacy. | Genomic data and all analytical results will be protected so that even if new sensitive attributes are identified, data subjects may withdraw their consent to the use of such attributes. | |

D. Related requirements

|  |  |
| --- | --- |
| D – Related requirements | |
| D.1 | Please provide requirements |
| Requirement 1, Requirement 2, Requirement 9  Note:  For this functionality, Requirement 1 and Requirement 2 are minimum necessary functionality.  Requirement 1 The standard shall include functionality to assist systems using dynamic consent in demonstrating accountability for the proper implementation of standards-based systems.  Requirement 2 The standard shall provide the capability to safeguard the analytical outcomes of genomic information, using dynamic consent, in the same manner as the genomic information. |

# Requirements

1. The standard shall include functionality to assist systems using dynamic consent in demonstrating accountability for the proper implementation of standards-based systems.
2. The standard shall provide the capability to safeguard the analytical outcomes of genomic information, using dynamic consent, in the same manner as the genomic information.
3. The standardization shall support the technologies that aid in dynamic consent and ensure the privacy of genomic data and metadata, satisfying the compatibility to the ISO/IEC 23092-x standard series (MPEG-G).
4. The standardization shall define option to be compatible with relevant MPEG standards, such as ISO/IEC 23001-7 Common encryption in ISOBMFF, ISO/IEC 23009-x DASH, ISO/IEC 21000-23 Smart contract for Media, and others.
5. The standardization shall incorporate anonymization/pseudonymization methods specified by various regulatory authorities and provide the flexibility to add new specifications as they become available.
6. The standardization shall ensure the possibility to comply with regulatory and privacy laws by incorporating machine-readable methods of expressing interactions (ISO/IEC 000000 right expression languages).
7. The standard shall support different distributed processing environments (e.g., cloud, blockchains, webs, URLs) and be implemented in these environments.
8. The standard shall support different distributed processing environments (e.g., cloud, blockchains, webs, URLs) and be implemented in these environments.
9. The standard shall make available factors such as risk sharing, defense in depth, and security updates that are sufficient to protect security.

# UML description of the system

Figure 1 depicts the Use Case diagram. Registrants of DNA sequences provide consent to multiple consent items. The system's basic structure can be described using the Unified Modeling Language (UML) Figure 2. There exists an n:m relationship between consent items and registrants. Moreover, there is an n:m relationship between the specific permission item and the consent item, which is managed by the Privacy Agent.

The Permission object is responsible for specifying the following processing actions: query, anonymization, analyzer, and DRM. To use genomic sequence data, all specified processing actions shall be applied, and appropriate anonymization and DRM measures shall be applied. The execution of queries and analyzers is only permitted upon necessary condition for using genomic sequence data.

# 

Figure 1 Initial UML use case diagram



Figure 2 Initial UML class diagram

# References

1. PharmaLedger, website, https://pharmaledger.org/
2. Makoto Suematsu, Dynamic consent, Deliverable in The 14th Genomic Medicine Realization Promotion Council hosted by PM office, <https://www.kantei.go.jp/jp/singi/kenkouiryou/genome/dai14/siryou4.pdf>
3. GA4GH, Framework for Responsible Sharing of Genomic and Health-Related Data, <https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/>
4. Itaru Kaneko, Emi Yuda, On the Privacy of Genomic Big Data and HER Standardization and Regulation, ICBSP '19: Proceedings of the 2019 4th International Conference on Biomedical Imaging, Signal Processing October 2019
5. Cristian Pattaro et al, The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results, J Transl Med. 2015; 13: 348.
6. Deborah Mascalzoni, Roberto Melotti, Cristian Pattaro, Peter Paul Pramstaller, Martin Gögele, Alessandro De Grandi & Roberta Biasiotto, Ten years of dynamic consent in the CHRIS study: informed consent as a dynamic process, European Journal of Human Genetics volume 30, pages1391–1397 (2022)
7. IEEE, Ethically Aligned Design version 2, https://standards.ieee.org/news/ead\_v2/
8. GA4GH, Consent Policy, https://www.ga4gh.org/wp-content/uploads/Consent-Clauses-for-Genomic-Research.pdf
9. A R Manu; Jitendra Kumar Patel; Shakil Akhtar; V K Agrawal; K N Bala Subramanya Murthy, Docker container security via heuristics-based multilateral security-conceptual and pragmatic study, 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)
10. NIST, IR 8432 Cybersecurity of Genomic Data, https://csrc.nist.gov/pubs/ir/8432/ipd (2023)
11. NIST, IR 8467 (Initial Public Draft) Cybersecurity Framework Profile for Genomic Data, https://csrc.nist.gov/pubs/ir/8467/ (2023)
12. Marco Mattaveli (EPFL), Raymond Krasinski (Philips), Paolo Ribeca (Bioss), Itaru Kaneko (Hitotsubashi Univ, Tohoku Univ.), Emi Yuda (Tohoku Univ.), Mirko Zichichi(IOTA Foundation) Mayu Terada(Hitotsubashi University), m61786, Proposal for the new work item towards adaptable private data management of genomic information representation on the large-scale cloud environment. (note: this is internal document and only available for WG8 members)
13. Itaru Kaneko, Emi Yuda, Hitoshi Okada, Docker Vectorization, a Cloud-Native Privacy Agent—The Analysis of Demand and Feasibility for Era of Developing Complexity of Privacy Management, Journals Appl Science, Vol. 13, Issue 15
14. Mirko Zichichi, Stefano Ferretti, Gabriele D'Angelo, Victor Rodríguez-Doncel; Data governance through a multi-DLT architecture in view of the GDPR, Cluster Computing, Springer. July 2022.
15. Mirko Zichichi; Stefano Ferretti; Gabriele D’Angelo, On the Efficiency of Decentralized File Storage for Personal Information Management Systems, 2020 IEEE Symposium on Computers and Communications (ISCC)
16. Itaru Kaneko, Emi Yuda, Docker vectorization of privacy agents as a defense against the Sherlock problem of AI machine learning in the age of mass wearable devices and large-scale bio information, Information Processing Society of Japan, DPS-192,ISSN 2188-8647 (2022)