INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2016/m37802
February 2016, San Diego, United States

Source DMAG-UPC (Distributed Multimedia Applications Group – Universitat Politècnica de Catalunya)
Status Proposal
Title Follow up on application scenarios for privacy and security requirements on genome usage, compression, transmission and storage
Authors Jaime Delgado, Silvia Llorente

1 Introduction

In m35838 [1], we already contributed on an initial survey of possible privacy and security requirements on genome compression, transmission and storage. However, this document is more a follow up of m36405 [2], a second contribution more on the security and privacy issues for the handling of genomic information in different application scenarios.

As indicated in those contributions, genomic information can be used for many different purposes: Legal and forensics, find out the susceptibility to develop some diseases, relatives search, paternity tests, rare disease research, etc. These different scenarios could have different implications in the access to and distribution of genomic information.

Individuals should be "owners" of their genomic data and should be able to define its usage and access rules, even its removal after usage. For instance, people can donate their DNA information for a specific scientific study but they might not want that this same DNA information be used afterwards to find out the suspects from a murder thanks to an ancestor finder service.

The continuous research in genomics leads to everyday new findings inside currently stored information that may affect the persons whose genomic information is used, especially when used for different purposes from the ones initially expected. Some of these effects may imply for example the refusal from a health insurance company, not being accepted due to some tendency to suffer from a disease, etc. On the other hand, access to as many as possible DNA profiles is desirable in order to discover common features for different kinds of diseases, which will lead to better and more specific treatments based on the research done over that genomic information.

To this end, what we would like to highlight again in this document is the need to include usage and access rules inside genomic information files. That is not necessarily to deny access but to control information usage and allow some authorized uses based on individuals' consents. The final objective is to reach a compromise between providing privacy and security to individuals whilst permitting advances in research through controlled sharing of genomic information. This document is a follow up of the work presented in [1] and [2].

2 More possible security requirements

To complement the current version of requirements for Genome compression and storage in MPEG [3], we identified in [1] some additional security requirements for genomic information coming from different sources. In [2], we continued with this task, identifying some more security and privacy issues coming mainly from [4] and other relevant papers. These issues include:

- The need to guarantee genomic information integrity, which in fact has been already discussed in MPEG.
- The need to protect phenotype inference (including predisposition to physical and mental health conditions) from third parties.

- If a genomic leakage occurs, it is not possible to revoke or replace individual's DNA sequence. Moreover, other individuals (relatives) are also affected by the leakage forevermore (future generations share part of their DNA with their ancestors).
- Current anonymization and aggregation techniques for privacy protection do not apply to genomics, as a DNA sequence always identifies the individual.
- Genomic data protection policies must recognize the need for informed consent. This is needed to ensure that an individual's DNA is not sequenced without the individual's consent. Nevertheless, this poses a serious problem for genomic researchers, who will need informed consent from individuals and relatives. In addition, they will be forced to destroy the samples after the study.

In addition to the previous issues, some guidelines on genomic data protection and use can be found in [4] [5], such as:

- Storage and long term protection (also identified in MPEG). Encryption or genome split are proposed.
- Accessibility (again considered in MPEG). For example, define standardized operations to search for specific genetic diseases. In this case, only an operation, which accesses to a specific well-defined part of the genomic information, is permitted, and the rest of genomic information remains secure (it is not accessed).
- Use of an open-consent form, which states that data are deposited in an open-access database and may be re-identified, and that participants can choose to withdraw from the study at any time, as the Personal Genome Project [6] proposes.
- Participant-centered initiatives, which make patients active participants in the study and adopt a more "adaptive" informed consent model, like the ones proposed in [7].

As a third point, we should take into account that there are other techniques for sharing genomic information without sharing the genome, like the use of genomic data-sharing beacons [8] [9]. Beacons provide a web service for responding questions like "Do you have any genomes with an 'A' at position 100,735 on chromosome 3?" or other questions over equivalent data. The response to this question is simply "Yes" or "No".

The objective of this kind of services is to share genomic information without sharing the genome. Nevertheless, although the complete genome is not exposed, there are some privacy risks in these services, as described in [10]. The main risks come from the number of genomes inside the beacon, the characteristics of the genomic information stored (some of them are devoted to a specific medical condition, like autism) or the origin of the people included in it (regions of origin of the participants).

We would like just to highlight that genomic information can give more information than expected from an individual and her relatives even if the genome is not exposed. This is why it is important to consider the use of the Privacy-by-design concept [11] when defining a new format for representing genomic information, which would also take into account the system giving access to such information.

3 Examples of usage and access Rules and scenarios

Section 3 of [2] identifies some example scenarios, defined in several genomic-related papers and surveys, where rules for accessing or using genomic information would facilitate its privacy.

Example possible applicable rules include (see section 3 of [2] for details):

- Not for further use.
- Not for commercial use.
- Not for relatives search use.
- Not for forensics use.
- Forensics use.
- Open access.
- Only for scientific purposes.
- Inform of study result.

On the other hand, specific example scenarios include (see section 3 of [2] for details):

- Scenario 1: Genomic information donation for a specific scientific study.
- Scenario 2: Request of genomic sequencing to discard susceptibility to develop some diseases.
- Scenario 3: Genomic sequencing of crime scene biological evidences.
- Scenario 4: Paternity test or search for biological parents.
- Scenario 5: Public genomic information individuals' database.
- Scenario 6: Animal genomic information sequencing.

Table 1 summarizes a possible combination of scenarios and applicable rules.

Rules / Scenarios	**1**	**2**	**3**	**4**	**5**	**6**
Not for further use	X	X		X		
Not for commercial use	X	X		X		
Not for relatives search use	X	X				
Not for forensics use	X	X		X		
Forensics use			X			
Open access					X	X
Only for scientific purposes					X	
Inform of study result					X	

Table 1. Examples of scenarios and applicable rules

Taking into account the current and proposed requirements and the examples of scenarios and applicable rules, some additional issues should be considered:

- Could the expression of rules be included in the genome information file? The size of the rules is negligible when compared to the size of genomic information. In this way, the

inclusion of rules inside the genome file format should not be a problem, even if several rules are defined.

- Expression of security and privacy restrictions. Rules like the ones proposed could be used to describe security (encryption techniques, access from external systems, etc.) and privacy (who can access, when, for what purpose, who has to be informed, etc.) features associated to the genome. In this way, the security and privacy requirements already identified could be fulfilled in different ways. One example could be by using an application/system accessing to the genome file, which should be aware of the restrictions. It is worth noting that the rules would permit the description of allowed usages of the genome, which could be checked prior to performing any scientific test to avoid misuses from scientific community as reported in [12].

- Provision of genome search services based on access and usage rules. With the inclusion of rules in genomic information, one can imagine search services based on the permitted uses of the genomes. In this way, the scientific community may have a wider genome base for their studies whilst individuals may provide access to their genome with more confidence.

4 Conclusions

The discussion on the provision of security and privacy protection on the genomic information should continue. The scenarios, rules and additional requirements presented are based on real world situations already described in books, research papers and regular newspapers. See [15] to [29] for additional references.

The scenarios show rules and conditions that illustrate the variety of usages of genome information under different situations: user request, donation, crime scene biological evidence, paternity tests or even animal genome sequencing.

The main objective is to show that security and privacy information can be included into a genomic information format considering at least the following aspects:

- The rules for defining access and usage may allow usage of genomic information for some purposes whilst protecting it against other purposes.
- The rules' size could be considered negligible when compared to the total size of the genomic information.
- The addition of rules does not impact on the genomic information processing, but it may govern its usage.

For these reasons, the security and privacy requirements on genome compression, transmission and storage should not be separated from the rest of requirements in the definition of the genome information file. With the inclusion of usage and access rules inside the genome information files, everyone will be able to easily determine what can be done over these files and act accordingly.

Two examples help to emphasize again that security and privacy issues in genomic information is a hot topic in the scientific and technical communities:

- A 2nd International Workshop on Genome Privacy and Security (GenoPri'15) was held in May 2015, where some of the topics described in this document have been discussed [13].
- A Conference on Genomics and Patient Privacy will be held at Stanford in a few weeks [14], where relevant speakers in the area of privacy in genomics will participate.

References

[1] J. Delgado, S. Llorente. ISO/IEC JTC1/SC29/WG11 m35838, "Initial survey on privacy and security requirements on genome compression, transmission and storage". February 2015.

[2] J. Delgado, S. Llorente. ISO/IEC JTC1/SC29/WG11 m36405, "Some application scenarios for privacy and security requirements on genome usage, compression, transmission and storage". June 2015.

[3] C. Alberti et al. ISO/IEC JTC1/SC29/WG11 N15738, "Requirements on Genome Compression and Storage". October 2015.

[4] E. Hayday, E. de Cristoforo, J.-P. Hubaux, G. Tsudik. "Whole Genome Sequencing: Revolutionary Medicine or Privacy Nightmare?". Computer, IEEE. February 2015.

[5] "Privacy and protection in the genomic era", Nature Medicine, 19, 1073, Nature publishing group, September 2013. http://dx.doi.org/10.1038/nm.3342

[6] Web site from where Personal genome project participation documents can be obtained, June 2015. http://www.personalgenomes.org/harvard/sign-up#documents

[7] J. Kaye et al. "From patients to partners: participant-centric initiatives in biomedical research", Nature Reviews Genetics 13, 371-376, Nature publishing group, May 2012. http://dx.doi.org/10.1038/nrg3218

[8] Global Alliance for Genomics and Health Data Working Group, http://ga4gh.org/#/beacon, 2016

[9] "Global search engine for genetic mutations", https://beacon-network.org//#/

[10] S. S. Shringarpure, C. D. Bustamante, "Privacy Risks from Genomic Data-Sharing Beacons", American Journal of Human Genetics, Volume 97, Issue 5, November 2015, http://www.sciencedirect.com/science/article/pii/S0002929715003742.

[11] A. Cavoukian, Information and Privacy Commissioner, Ontario, Canada. "Privacy-by-design. The 7 Foundational Principles". August 2009.

[12] S. Zielinski, “Henrietta Lacks’ ‘Immortal’ Cells”, report on Rebecca Skloot book “The Immortal Life of Henrietta Lacks”, http://www.smithsonianmag.com/science-nature/henrietta-lacks-immortal-cells-6421299/

[13] 2nd International Workshop on Genome Privacy and Security (GenoPri'15), Held in conjunction with the 36th IEEE Symposium on Security and Privacy (S&P), May 2015, http://www.genopri.org/

[14] Genomics and Patient Privacy Conference 2016, March 2016, https://med.stanford.edu/gapp/events/GAPPConference2016.html

[15] “Big money heading for consumer genomics”, http://www.biopoliticaltimes.org/article.php?id=8927

[16] “Genetic Surveillance: Consumer Genomics and DNA Forensics”, http://www.biopoliticaltimes.org/article.php?id=8960

[17] “Unlocking my genome: Was it worth it?”, http://www.cnbc.com/2015/12/10/unlocking-my-genome-was-it-worth-it.html

[18] “Tech Companies Are Not Trusted with Health Data”, https://www.technologyreview.com/s/543536/tech-companies-are-not-trusted-with-health-data/

[19] “Search of genetic mutations”, https://www.technologyreview.com/s/535016/internet-of-dna/

[20] “Modern genetics means you should say goodbye to privacy”, http://www.techinsider.io/modern-genetics-means-genomic-privacy-is-impossible-2015-10

[21] “D-N-Ain’t? Genome Sharing Vulnerability Spells Possible Privacy Issues”, https://securityintelligence.com/news/d-n-aint-genome-sharing-vulnerability-spells-possible-privacy-issues/

[22] “It’s 2015: Do you know where your genetic data are?”, https://www.geneticliteracyproject.org/2015/03/16/its-2015-do-you-know-where-your-genetic-data-are/

[23] “Gene testing framework ignores privacy and security concerns”, http://theconversation.com/gene-testing-framework-ignores-privacy-and-security-concerns-12650

[24] “Toward Protecting Participants’ Privacy”, http://www.the-scientist.com/?articles.view/articleNo/44369/title/Toward-Protecting-Participants--Privacy/

[25] “Your DNA, decoded”, http://www.medicaldaily.com/collection/your-dna-decoded

[26] “Precision Medicine Exposes Patients' Genetic Information; Sharing and Protecting Data with Google”, http://www.medicaldaily.com/precision-medicine-exposes-patients-genetic-information-sharing-and-protecting-data-353690

[27] “Your 23andMe DNA Can Be Used In Racist, Discriminatory Ways”, http://www.buzzfeed.com/stephaniemlee/your-23andme-dna-can-be-used-in-racist-discriminatory-ways#.sxKBljJaQ

[28] Australian Government, Office of the Australian Information Commissioner, https://www.oaic.gov.au/

[29] “Data Privacy Day 2016”, https://staysafeonline.org/data-privacy-day/