Vol.2019-EIP-84 No.20
2019/6/4

各国の個人情報保護における遺伝子情報の扱いと MPEG 遺伝子符号化のセキュリティ機能

金子 格†

ISO/IEC では遺伝子情報符号化の国際標準化を進め現在 ISO/IEC 23092-1～5 の 5 パートがほぼ完成し，現在パート 6 の標準化を進めている．一方各国は遺伝子情報の個人情報保護を規定している．これらの要求項目を満たすためには MPEG は個人情報保護のための API 拡張を行うべきだと考えられる．具体的には個人毎の遺伝子情報保護以外に，属性毎の制限や，特定 read の制限が望ましいと考えられる．

On the genomic information processing in the privacy regulation of various countries and security API of MPEG genomic coding

ISO / IEC promotes international standardization of genomic information coding. Five parts of ISO / IEC 23092-1 to 5 are almost complete. Standardization of Part 6 is currently in progress. Meanwhile, each country stipulates the protection of personal information including genomic information. In order to meet these requirements, it is possible to consider an API extension for personal information protection in MPEG standardization Specifically, in addition to protection of genomic information for each individual, selective protection of genomic information depending on the sensitivity or criticalness of the region or the element of genomic data is possible to be desirable.

ITARU KANEKO†

1. 概要

ISO/IEC JTC 1/SC 29/WG 11(MPEG)では遺伝子情報符号化表現の国際標準策定を進めている．現在 ISO/IEC 23092(MPEG-G Genomic Information Representation 以下 MPEG-G)の標準策定作業が進行しておりこれまでにパート 1～5 の 5 パートがほぼ完成した，現在パート 6 の策定が進んでいる．

しかし MPEG-G の標準化作業中にもセキュリティとプライバシーに関する社会的要求が高まり，各国および国際間の法整備も進んだ．プライバシー保護に利用可能な，いくつかの興味深い技術も議論され始めている．

本報告ではさまざまな国，地域における遺伝子情報処理における，セキュリティとプライバシーに関する要求の調査と比較の結果について説明する．このような調査により MPEG-G をより広く適用可能な標準とすることを目的としている．

また, MPEG-G においてセキュリティとプライバシーのために導入された仕様についても調査した結果を述べる. MPEG-G をプライバシー対応とする技術として，報告者はすでに MPEG-G において匿名化技術の導入を検討することを提案した[2]．そこで匿名化等についての調査も紹介する．これらのテクノロジを MPEG-G と調和させることは MPEG-G を利用する際に大きな利点となると考えられる．

さらにプライバシーとセキュリティのツールを MPEG-G に導入するため，提案募集を行うことが望ましい理由を説明する．これは難しい技術的課題であるが，そうした困難な技術課題に対し提案募集を行うことで MPEG-G をより付加価値の高い標準とすることができると考えている．

1.1. GIR と二次利用の定義

まず対象となる遺伝子情報の符号化表現とはシーケンサと呼ばれる装置で取得した遺伝子情報の符号化表現とし，以下では GIR と表記する．GIR の二次利用とは本人の同意をえた治療，検査目的以外の利用を表すものとする．本研究報告は GIR の二次利用に関するものである．

† 名古屋市立大学
Nagoya City University

プライバシーの概念は今日十分に確立されており，さまざまな規則が合意され共有され，さらに発展している分野である．しかし,「医療情報の二次利用」は依然として複雑な分野である．治療のための検査は患者が自ら進んで行うものであり問題がない．これは医療情報の一次利用といえる．医療情報の主な用途である自分のための医療および健康データの使用には問題はない．他の健康情報との相互相関分析のための GIR の蓄積およびそれらの利用は，検査を受けた者自体がその利益を受けるわけではないから，二次利用と認識されている．したがって関連する規制を遵守しない限り許されない．規制を満たすためには，一般的には二次利用の許諾を得るか，必要な情報のみを得ながらプライバシーを保つための，いくつかの技術的手段が有用である場合がある．

1.2. EU における規制

1.2.1. GDPR

GDPR (General Data Protection Regulation) [1] は，2018 年 5 月 25 日に施行された．GDPR には GIR のプライバシーとセキュリティに関連する項目が含まれると考えられる．Shabani [15] は GDPR の GIR に対する影響を次のように説明している.「新しい規則では，匿名化されたデータを個人の（識別可能な）データとして認識し，特別なカテゴリのデータのカタログに機密データを含めることで個人データの範囲を明らかにしようとしている．さらに，個人データを処理するための規則に，科学的研究のための免除の項目において一連の新しい規則が定められている.」

1.2.2. 匿名化と仮名化

GDPR では，規制の免除の条件として匿名化と仮名化をあげている．匿名化後のデータは GDPR の適用外となる可能性がある．仮名化後のデータの場合, データは依然として GDPR によって規制される可能性があるが，規制のいくつかが緩和される．

仮名化は ISO 25237 [8] に記述されている．匿名化と仮名化は ISO / IEC 20889 [7] に記述されている．

1.3. 日本における規制

1.3.1. 個人情報保護条例等の規制

日本では 2003 年(平成 15 年)に個人情報保護法が成立した．この法律では,「地方公共団体は，この法律の趣旨にのっとり，その地方公共団体の区域の特性に応じて，個人情報の適正な取り扱いを確保するために必要な施策を策定し，及びこれを実施する責務を有する(平成 15 年 個人情報保護法第 5 条)」とした．これにもとづいて各地方自治体がそれぞれにプライバシー保護「条例」を策定したため日本における規制は団体により異なることになった．約 2000 のバリエーションがありこれは「個人情報保護条例 2000 個問題」として知られている．一方 2019 年に EU と日本の間で十分性認定が発行した．したがって日本において事業者や公共団体はそれぞれの地域や事業内容に応じて，これら多様なプライバシー規制に柔軟に対応できる必要があると考えられる．

1.3.2. 日本と欧州の GDPR 十分性認定

前述のように日本と欧州の間にはすでに十分性認定が発効し，日本と EU の間で、相互に「個人データの移転を行うことができるだけの十分なデータ保護の水準を持つ」ことが認定された．GDPR は前述のように遺伝子情報の扱いにも影響を及ぼすと考えられている．

板倉, 寺田等は日本における GDPR の十分性認定における課題と今後の展望について報告している．[5] [6]

1.3.3. 健康情報の二次利用

日本では健康情報の二次利用が計画されている[14] このようなシステムには匿名化処理が想定されている．

1.4. 米国

1.4.1. NIH の遺伝子プライバシー方針

米国 NIH(National Institute of Health) は Genomics に関する情報をまとめているが，その中の”Privacy in Genomics”[9] で以下のように説明している．

現状研究用の DNA サンプルはや解読データは氏名を除くという形で個人が特定ができないようにして交換されている．しかしこれで匿名化がされているかというと Gemrek 等の報告[11] によれば遺伝子から個人を特定できる可能性があることが示されている．名前を除いただけではプライバシーが保たれず，プライバシーと研究のバランスが困難な課題であるとしている．

NIH は研究機関に Certificate of Confidentiality を付与し, 遺伝子サンプルやデータのアクセスを，適格性を持つ研究機関制限している. 遺伝子情報利用に関して, Common Rule と HIPAA (Health Insurance Portability and Accountability Act) を定め，これらによって研究や事業とプライバシーのバランスを図っている．GINA(Genetic Information Nondiscrimination Act)は保険事業と雇用者による遺伝情報アクセスを制限し HIPAA は健康情報の共有を制限する．

一方犯罪捜査のための遺伝子検索は CODIS と NDIS(National DNA Index System)により可能である．また最高裁は 2013 年 Maryland 対 King で捜査機関は容疑者の遺伝子を指紋と同様採取できるとした．

1.4.2. Common Rule

1991 年に制定された連邦政府の人体保護政策である[7]

1.4.3. GINA

遺伝情報非差別法：GINA は 2008 年に行われた米国の法律であり遺伝的プライバシーを保護する．

1.5. その他の国

その他の国について調査中である．経済規模としては中国およびその他の TPP 加盟諸国が大きく，これらのエリアとの整合性が得られることは有益であると考えられる．

2. プライバシー保護技術

2.1. 匿名化

次にプライバシー保護技術について述べる．まず、匿名化は医療情報を他のデータ形式に変換する方法である．変換の後，全ての有用な医学的統計的属性を保存しながら，個人データのどれも識別されないことを目指している．例として ALLSTAR 心電図解析などの試みがある[2]．

2.2. 仮名化

仮名化は「消費者データレコード内の個人を特定できる情報フィールドを，後でレコードを再識別するために呼び出すことができる 1 つまたは複数の人為的な識別子，または偽名に置き換えるデータ管理手順.」である．仮名化でもある程度の利用が可能となる．

2.3. 準同型暗号化

完全準同型暗号化は，1978 年に Rivest，Adleman, および Dertouzos によって提案された暗号化手法である．準同型暗号化によれば，秘密鍵へのアクセスを必要とせずに暗号化データに対して直接計算を実行できる．そのような計算の結果は暗号化された形で残り，後で秘密鍵の所有者によって解読できる．

2.4. 部分暗号化

部分暗号化は，データの一部だけを暗号化する方法である．この機能はすでに MPEG-G でサポートされている．

3. 適用分野（ユースケース）

3.1. 大規模ゲノムデータベース

次に遺伝子情報のセキュリティが実現可能な場合に，どのような効用が得られるかを検討する．最初に大規模ゲノムデータベースについて述べる．匿名化が実行可能であるかは現状では明らかではないが．もし可能であれば匿名化したデータは許諾を得ることなく利用できるため，原理的には分析したすべての遺伝子情報をデータベース化することが可能となる．これにより大規模なゲノム情報データベースの作成や利用が容易になる．

もちろん，匿名化の際に一部のデータが欠損することは避けられない．だとしてもデータベースの大きさはその欠点を補う利点となる可能性がある．

3.2. ゲノムデータベースの偏りを修正するための中規模ゲノムデータベース

匿名化されたデータは，たとえそこに含まれる遺伝情報が不完全で，かつ中規模なものであっても，統計的な分析に役立つ可能性がある．

匿名化がなければ，遺伝子サンプルは自発的にしか提供されないため，母集団のデータは自分のゲノムデータを提供する動機を持つ遺伝子サンプルに偏ってしまう．多くの場合，そのグループはおそらく平均的なグループと比較して健康により強い関心がある集団であると予想される．このような偏りが分析に影響を及ぼす可能性がある．

統計的に限られて，中規模のデータベースであっても，匿名化により本人の選択を排したデータがあれば，偏りのないデータを得ることができる．このような「制御されたデータセット」があることは匿名化データベースから直接データを得るのみでなく，非匿名化遺伝子データベースの分析結果の信頼度を向上するためにも，利用することができる．

3.3. 電子健康情報と組み合わせたゲノムデータベース

EHR は電子カルテの略で，医療情報と健康情報のデータベースである．全国 EHR はすでにいくつかの国で始まっている．カナダ，アメリカ，イギリス，フィンランド，エストニア，シンガポール，オーストラリア，ニュージーランド，ロシア（モスクワ）．日本でも全国 EHR が検討あるいは開始されている[13] [14]．日本の全国 EHR の場合，医療情報のプライバシーは非常に重要であり，厳重に保護されている．しかし，臨床研究などの

医療情報の二次利用もまた重要な問題として認識されている．そのために，匿名化はすでに基本システム構造に含まれている．

GIR の匿名化により，GIR をデータベースに含め，健康データを GIR と相互分析できれば，大きなメリットが得られる可能性がある．そのようなデータベースは，ゲノム特性と様々な健康特性との間の関係を分析するのに非常に有益だろう．さらに重要なことは，そのようなデータベースは，匿名化なしの場合と比較して，「健康な集団」に関するより多くのデータを集積できることである．つまり病気の分析よりも，大部分の健康な人のさらなる健康改善に役立つと考えられる．

4. MPEG-G のセキュリティ機能

次に MPEG-G のセキュリティ機能を説明する．遺伝子情報の基本構造を示し，MPEG-G のデータ構造を示し，API のセキュリティ機能を説明する．

遺伝子は塩基の配列であるが，現在の分析手法では配列全体が直接得られるわけではなく，Read とよばれる断片の検出結果が多数得られる．この多数の Read から配列全体を推定するのに利用される．遺伝子情報の保存ではこの Read の検出結果をそのまま記録する．

図 1 に MPEG-G のファイル構造を示す．ファイルは先頭に File Header があり複数の Dataset Group を含んでいる．Dataset Group は複数の Dataset からなり Dataset はアクセスユニットと Descriptor Stream からなる配列構造を持つ．

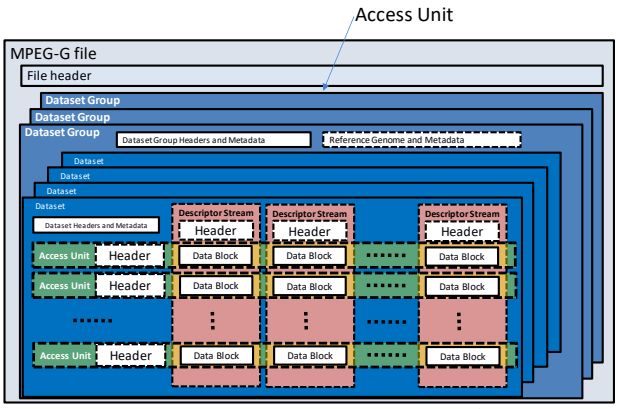

図 1 MPEG-G ファイル構造

図 2 に Descriptor Stream の構造を示す．Descriptor Stream は多数の Read Descriptor の連続からなり，Read Descriptor は，前述の遺伝子分析によって得られた Read の検出情報を表す．

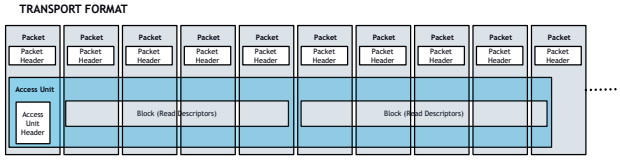

図 2 Descriptor Stream

表 1 に MPEG-G の API(アプリケーション・プログラム・インタフェース)を示す．ここでセキュリティに関する API を黄色でマークしている．

先に示したように MPEG-G のデータにはファイル全体，メタデータ，ヘッダ等に暗号化による保護を行うことができる．メタデータには複数国の制約に対する条件なども，格納することが可能であると考えられる．制約を満たすデータのみが取り出されるようにデータの読み出しを制限することが可能な API となっている．

このようなセキュリティ機能を各国のプライバシー規制と整合させるためにどう利用するかについては，現在技術情報としてドキュメント化を進めている．

氏名などの情報を隠蔽することも可能であるし，個人の特定につながるような遺伝子の特定の場所を保護し，その部分については適切な権限を持っている場合にのみデータの取得を可能とすることもできる．

表 1 MPEG-G API

errorCode_t GetHierarchy
errorCode_t GetDataBySimpleFilter
errorCode_t GetDataByAdvancedFilter
errorCode_t GetDataBySignature
errorCode_t GetDataByLabel
errorCode_t GetMetadataFields
errorCode_t GetMetadataContent
errorCode_t GetDatasetGroupProtection
errorCode_t GetDatasetProtection
errorCode_t GetDatasetRegionProtection
errorCode_t GetDatasetReference
errorCode_t GetSimpleStatistics
errorCode_t GetAdvancedStatistics

5. 議論: 遺伝子情報保護の特性

以上のように遺伝子情報保護に関して現状の運用と法規制，MPEG-G の機能を概観してきた．ここでは遺伝子の個人情報保護はどのような規範で実現すべきかを論じる．

遺伝子のプライバシー，個人情報保護法制との整合性，遺伝子情報の医療利用による公益とのバランスをとることがいうまでもなく必要であると考えられるが，そのためにはいかなるシステムが適当なのだろう．

現状の遺伝子情報の管理は遺伝子サンプルがだれのものであるかとう情報をまず保護し，また遺伝子サンプル自体も個人を特定される可能性があるから保護する，という方法をとっている．しかしこの方法では大量の遺伝子データを共有し分析することが非常に困難になり情報漏洩による被害の可能性も高まるように思われる．

遺伝子情報そのものに本人を特定できる情報が含まれている状態では，遺伝子サンプルが漏洩しただけで個人情報の漏洩と考えることもできる．プライバシー法制度の面でも，個人を特定できる医療情報が自動的に個人情報とみなされるならば DNA サンプルは部分的であっても指紋と同様に個人を特定するのに十分な情報であるから個人情報としての取り扱いが必要になるのではないかと考えられる．

それでは DNA サンプル自体を個人情報とすべきかと考えると，現状の DNA 利用においてそれではデメリットが大きすぎるのではないかと考えられる．

さらに DNA から得られる個人情報は事実上「公知の」情報である，という側面も冷静に見るべきではないかと思われる．法的には DNA サンプルの入手はむろん本人の同意を得なければ行えない．しかし DNA サンプルが適切に取得されたものかを確認する手段がなく，分析自体は規制されていないから，実際上だれの遺伝子でも容易に取得分析をして結果を参照することは容易である．望ましくない行為ではあるが現実的にそれを防ぐことは難しい．

DNA のプライバシーはむろん適切に保護されるべきであるが, DNA の特徴と DNA 分析により得られる情報の有用性を考慮して，規制の実施面での最適化がされる必要があると考える．

6. MPEG-G 拡張の提案

各国の個人情報保護における遺伝子情報の扱いと MPEG 遺伝子符号化のセキュリティ機能を概観し, MPEG-G のセキュリティ機能を確認した．

DNA のプライバシー保護においては，秘匿性や利用価値において特徴があり，その特性にあったセキュリティ機能の最適化が望ましいと考えられる. MPEG-G にはすでにそのために有用な機能が含まれているが，これをさらに拡張していけば GIR のセキュリティと効用のバランスをとることができるのではないかと考える．

特に匿名化，偽名化，準同型暗号化は現状含まれていない．これらを MPEG-G に適用可能かは確定できないが，もし可能であれば有用な機能となると考えられる．この分野での提案募集を行い拡張が可能となれば，今後も加速度的に拡大する GIR の利用において必要になるセキュリティ機能の提供と GIR 利用による効用の拡大に資するのではないだろうか．

以下に拡張をする場合の要求条件(提案募集を行う際にその基本となる要求項目)の素案を示す．

[要件条件(案)]

(1) プライバシー保護に信頼できるゲノムデータのセキュリティ技術(匿名化、仮名化、暗号化、同型暗号など).
(2) プライバシー保護のために信頼できる匿名化を可能にするための API.
(3) ゲノムデータと他の HER との相互相関分析を可能にするための API.
(4) (1)を有効にするためのコード化表現.
(5) (2)有効にするためのコード化表現.
(6) (3)を有効にするためのコード化表現.

国際標準化においては匿名化技術他のプライバシー保護技術を現時点で固定化しないような仕組みが望ましい．したがって MPEG-DASH のセキュリティ機能のように現在最も適切な方式の他に将来他方式に交換が可能な構造とすべきだろう．標準としての利便性のためには現状で利用すべき方式を選択する必要はあるがその技術でロックするべきではない．

[実装方法(案)]

実装方法は今後の課題である．しかし匿名化は有力な手法でありそのためにはMPEG-GのRead単位でのアクセス制限機能が有用であると考えられる．遺伝子には個人毎に違う部分，変異があるが，その中で非常に珍しい変異が含まれていれば Read 一つで個人を特定する大きな手掛かりとなる．したがってそのような変位を含む Read へのアクセスは選択的に制限する機能が必要ではないかと考えられる．Read へのアクセスを選択的に保護する機能はすでに MPEG-G に含まれているからその機能をベースに匿名化アクセスを実現することが想定される．

7. まとめ

本報告ではさまざまな国，地域における遺伝子情報処理におけるセキュリティとプライバシーに関する要求の調査と比較の結果について報告した．また現在パート 1～5 までが完成しパート 6 が策定中の MPEG-G の概要を示し，セキュリティ機能を有することを示した．DNA のプライバシー保護の現状から MPEG-G のセキュリティ機能の拡張の提案を示した．

最初のヒト・ゲノムの解析が 2003 年に達成さ

れて以来，遺伝子解読のスピードは驚異的に進歩を続け現在では個人でも遺伝子解読が委託できる時代になった．今後も遺伝子解読のコストは急速に下がると見込まれる．プライバシーに関する懸念がある一方で，適切な方法で遺伝子分析の結果と他の統計情報を効果的に組み合わせることができれば，その効用は計り知れない．もちろんプライバシー侵害や遺伝子差別などの社会的弊害は最大限防止する必要があることは言うまでもない．

MPEG-G をこれらのバランスがとれたより有用な国際標準とすることを目指して，今後も意見投入を継続する予定である．

References:

[1] Wiki - General Data Protection Regulation, https://en.wikipedia.org/wiki/General_Data_Protection_Regulation, 2019-03-18 20:35 (UTC).
[2] Itaru Kaneko, Emi Yuda, Consideration of anonymizations in MPEG-G and brief survey of personal medical data in Japan, 123th MPEG meeting, MPEG m43376 (2018)
[3] Itaru Kaneko, Emi Yuda, Requirements and Use cases on anonymization of Genomic Information Representation, 124th MPEG meeting, MPEG m45077 (2018)
[4] Yoichiro Itakura, Akiko Fujimura , Kumiko Kameishi , Fumihiko Magata, 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 (2019)
[5] Yoichiro Itakura, Maru Terada, 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 (2019)
[6] Maru Terada, Yoichiro Itakura, 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 (2019)
[7] ISO/IEC, ISO/IEC 20889 - Privacy enhancing data de-identification terminology and classification of techniques, https://www.iso.org/obp/ui/#iso:std:iso-iec:20889:ed-1:v1:en
[8] ISO, ISO25237, Health informatics — Pseudonymization
[9] NIH, Privacy in Genomics, https://www.genome.gov/27561246/privacy-in-genomics/, acquisited on March 19, 2019
[10] NIH, Privacy in Genomics, https://www.genome.gov/about-genomics/policy-issues/Privacy, acquisited on April 23, 2019.
[11] Melissa Gymrek, et al, 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
[12] Robinson Bradshow, Privacy and IP Law affecting Health Care , Research and IT, https://theprivacyreport.com/, acquisit 3/19/2019
[13] Mahsa Shabani, Pascal Borry, 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 (2018)
[14] Yuda E, Furukawa Y, Yoshida Y, Hayano J & ALLSTAR Research Group, Association between Regional Difference in Heart Rate Variability and Inter-prefecture Ranking of Healthy Life Expectancy: ALLSTAR Big Data Project in Japan, Proceedings of the 7th EAI International Conference on Big Data Technologies and Applications (BDTA), Chung-ang University, Seoul, South Korea, November 17-18 (2016)
[15] Samarati, Pierangela; Sweeney, Latanya (1998). "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression" (PDF). Harvard Data Privacy Lab. Retrieved April 12, 2017. https://dataprivacylab.org/dataprivacy/projects/kanonymity/paper3.pdf
[16] YOSHIHARA Hiroyuki, gEHR Project: Nation - wide EHR Implementation in JAPAN, Kyoto Smart city Expo, https://expo.smartcity.kyoto/2016/doc/ksce2016_doc_yoshihara.pdf (2016)