［安全なデータ活用を実現する秘密計算技術］

⑥組織間データ結合における海外制度の動向

美馬正司 （株）日立コンサルティング

データ結合の課題

ビジネスだけでなく，社会におけるあらゆる場面においてデータの活用が進められているが，パーソナルデータの活用においては，データの結合が大きな課題となる．個々人がさまざまなサービスを利用したり，いろいろな活動をすることで，データが散在しているが，これらを結び付けた方が個人を深く理解し，より適切なサービスの提供や，社会全体としての最適化が図れると考えられる．しかしながら，このようなパーソナルデータの結合に関しては，基本的に本人の同意が求められ，将来的なデータ結合を想定した同意の取得は困難である．また，すでに収集されているパーソナルデータの再同意には大きなコストが必要になる．我が国では，「行政手続における特定の個人を識別するための番号の利用等に関する法律」が 2015 年 10 月に施行され，行政分野においては，個人番号（マイナンバー）を使って，規定された業務に限って本人の同意を必要とせず，個人の情報を連結することが可能になっている．

一方，2017 年 5 月に施行された改正個人情報保護法により，匿名加工情報が規定され，本人の同意なくしてパーソナルデータの流通が可能になったが，匿名加工情報において同一個人のデータの結合は認められていない．

このような課題から特にデータの結合が求められる医療分野においては，「医療分野の研究開発に資するための匿名加工医療情報に関する法律」（以下，次世代医療基盤法）が 2018 年 5 月に施行され，本人の同意なく医療情報（個人情報を含む）の収集，結合を行い，匿名加工することが可能になっている．本人への通知，オプトアウト機会の提供は求められるものの，主務大臣が認定した認定匿名加工医療情報作成事業者（以下，認定事業者）に限って，本人の同意を必要としないデータの結合が可能となり，匿名加工医療情報を作成し，研究開発等へ供することができる．

認定事業者のように，パーソナルデータの結合を可能とする組織は信頼できる第三者機関（Trusted Third Party，以下，TTP）と呼ばれており，海外でも同様の取り組みが見られる．本稿では，我が国におけるパーソナルデータ結合の今後の可能性検討の参考とするため，海外における TTP やそれ以外のデータ結合の事例を紹介する．

英国の CPRD

CPRD の概要

英国の Clinical Practice Research Datalink（以下，CPRD）は 2012 年 3 月から開始したプログラムで，研究用データを提供する政府の非営利サービスである．CPRD は，1987 年からデータを収集していた保健省医薬品・医療製品規制庁（Medicines and Healthcare products Regulatory Agency）の General Practice Research Database と国立衛生研究所（National Institute for Health Research）の Research Capability Programe を統合したデータベースで，双方が共同で運営を行っている．CPRD の利用は，英

国内外の大学や企業等の研究者が利用できるが，基本的には研究目的は学術的なものに限られている．ただし，その利用は広く行われており，これまでに1,800以上のCPRDを用いた研究が発表されている．

CPRDは英国のかかりつけ医（General Practitioner，以下，GP）から治療データを収集しており，500万人以上の患者のデータが登録されている．CPRDの大きな特徴はこのようなGPのデータを他の関連データと結合した形で提供できることであり，具体的には表-1に示すようなデータと結合されている．

CPRDのデータ結合の仕組み

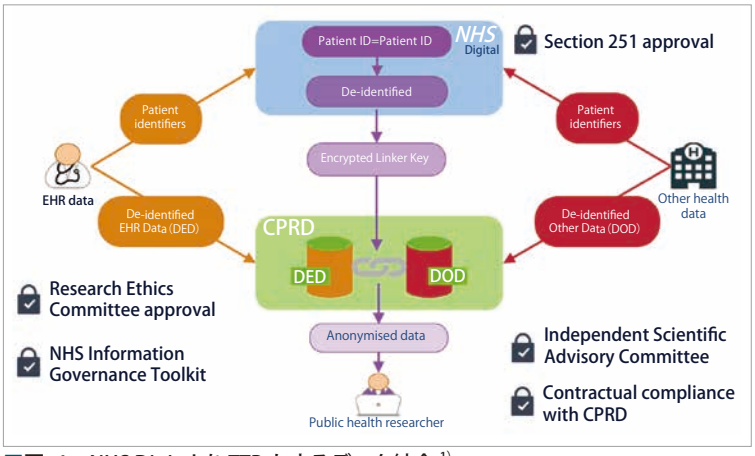
CPRDにおけるGPのデータの他のデータの結合は国民保健サービス（National Health Service，以下，NHS）の情報部門であるNHS DigitalがTTPとして実施している．NHS DigitalをTTPとしたデータ結合の仕組みを図-1に示す．

データの結合を行うため，2つのデータベースから患者の識別子（NHS番号，生年月日，郵便番号と性別）と各データベース独自IDがNHS Digitalに送付される．その際，患者の識別子以外の治療データ等はNHS Digitalへは送付されない．

NHS Digitalは，2つのデータベースからの患者識別子を照合して，患者識別子を含まない暗号結合鍵を生成し，各データベースの独自IDと合わせてCPRDへ送信する．CPRDは2つのデータベースから患者の識別子を含まないが独自IDを含むデータセットを受け取り，NHS Digitalから受け取った暗号結合鍵を用いてデータセットを結合する．

■表-1　CPRDでGPのデータと結合可能なデータ[1)]

• 病院エピソード統計
• 国家統計局の死亡登録
• イングランド公衆衛生サービスのがんデータ
• メンタルヘルスデータセット
• 貧困等の社会経済状況のデータ



■図-1　NHS DigitalをTTPとするデータ結合[1)]

CPRDの制度的な位置付け

CPRDはNHS法の251条に基づきConfidentiality Advisory Groupの審査を毎年受けており，その承認を基に運営されている．

また，TTPであるNHS Digitalは患者を識別できるデータを受け取れる法的機関としてイングランドで許可されている．

さらに，EU離脱が決まっている英国ではあるが，EUの一般データ保護規則（General Data Protection Regulation，以下，GDPR）に基づく整理もされている．CPRDおよびNHS Digitalによるデータ結合は，GDPRの第9条「特別な種類の個人データの取り扱い」の適用除外である第9条（2）(i)「公衆衛生の分野における公共の利益」，第9条（2）(j)「科学的研究目的」のために行われていると，位置付けられている．

CPRDから研究者へのデータ提供については，2012年11月に英国のプライバシー・コミッショナー（ICO，Information Commissioner's Office）で公開した匿名化に関する行動規範（code of practice）に基づいて行われることになっている．

このような法的な位置付けに加えて，結合データの利用については，CPRD内の独立科学諮問委員会（Independent Scientific Advisory Committee）の承認が必要になるほか，患者にはGPにおいてオプトアウトの機会が提供されている．

韓国の専門機関

匿名化のガイドライン

韓国では「個人情報保護法」が 2011 年 3 月に公布され，同年 9 月に施行されている．それに伴いプライバシー・コミッショナーに当たる個人情報保護委員会が設立され，同委員会が「個人情報保護法」にかかわる事項を管轄している．その後，ビッグデータ活用に向けた検討が進められ，「個人情報保護法」の補足として「ビッグデータのプライバシーガイドライン」が 2014 年 12 月に放送通信委員会において策定された．ただし，詳細なガイドラインにはなっておらず，引き続き具体的な匿名化の在り方が検討されていたと推察される．結果，匿名化データの活用に必要な個人データの匿名化の標準や範囲について明らかにすることを目的とした「個人データの匿名化のためのガイドライン」[2] が 2016 年 6 月 30 日に国務調整室，行政自治部，放送通信委員会，金融委員会，未来創造科学部，保健福祉部など韓国政府機関共同の取り組みとして公表された．

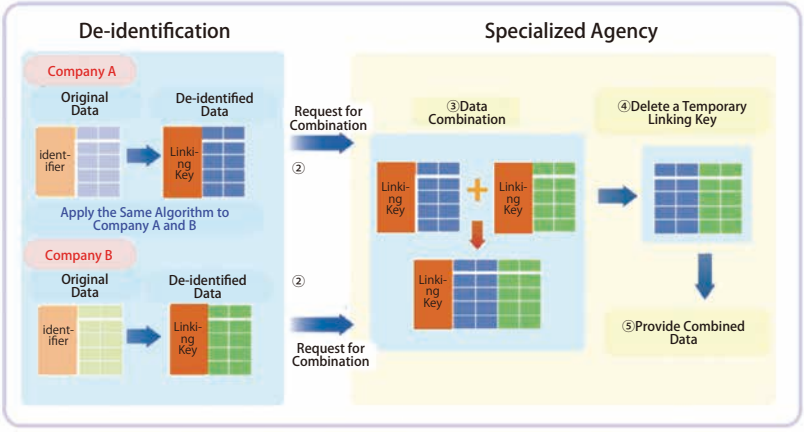
このガイドラインの特徴は，大きく 2 点ある．1 つは匿名化の指標として k-匿名性を活用することが明示されていることである．リスクを評価して，リスクが十分に小さければ匿名化を行うという標準的なプロセスを示し，リスクの評価指標として k-匿名性による評価を必須としていることが他の諸外国と大きく異なる．もう 1 つの特徴は匿名化したデータの名寄せを行える TTP として専門機関の設置が示されたことである．

専門機関の概要

大規模なデータ分析のために，異なる個人データ管理者が所有するデータセットを組み合わせる必要がある．これを行うには，異なるデータセットを照合してデータセットを作成するためのキーが必要であるが，個人を特定できる直接識別子を使用すると，法律違反になる．したがって，データセットを結合して分析するためには，結合プロセスでのみ一時的にマッチングキーとなる「一時リンクキー」を使用する必要がある．

しかしながら，一時リンクキーによる組合せが許可されても，無差別な組合せによる個人情報の侵害の可能性を防止するための支援・管理体制が求められる．そこで，各分野に専門機関を置き，個人データ管理者間のデータセットの組合せを可能にすることとなった．

専門機関はすでに設置されている各分野の公的な期間が担うこととなっており，韓国インターネット振興院，韓国信用情報サービス，金融監督院，社会保障情報サービス，国家情報社会局等，各分野の専門機関が指定されている．韓国インターネット振興院（KISA）と韓国情報化振興院（NIA）は専門機関が存在しない産業を支援することになっている．



■図 -2　韓国の専門機関におけるデータ結合 [2]

専門機関におけるデータ結合

専門機関では図 -2 に示す手順に基づいてデータの結合を支援する．

各個人データ管理者は，他の個人データ管理者と同じアルゴリズムを適用して一時結合鍵（識別子で作られたデータ結合のためだけの一時的な仮の識別子）を作成し，そのデータセット

の匿名化措置および適切性評価を実施する．

次に個人データ管理者は匿名化データを専門機関に提供し，結合を要求する．

そして，専門機関は一時結合鍵を用いてデータセットを結合し，一時結合鍵を破棄する．

専門機関は，データセットの組合せを要求した個人データ管理者に結合データセットを提供する．専門機関は，それを提供した後，結合データセットを廃棄する．

なお，一時結合鍵を作成する場合，韓国の国民IDである住民登録番号の使用は禁止されている．また，*k*-匿名性の値は，一時結合鍵なしで計算されることになる．専門機関は，結合の過程で再特定性が検出された場合には，匿名化データを直ちに廃棄することになっており，受け取る側の利用者にいてもリスクの評価を行うことが求められている．

デンマークのBiobank

Biobankの概要

デンマークでは，出生時に新生児から遺伝子データを採取し，国民IDであるCPR番号と併せて保管し，研究用に提供するBiobankという事業が行われており，2012年3月に研究用途での遺伝子データの研究利用が開始されている．Biobankの運営を行っているのは，国立血清研究所（Statens Serum Institut）であるが，運営費用については，事業開始から10年間は製薬会社が設立したNovo Nordisk財団が拠出することになっている．

Biobankには約570万人分のデータが登録されており，これはデンマーク国民のほぼすべてに当たる．また，遺伝子データおよび試料の数は2,450万件に達している．表-2に示すように複数のデータベースから構成されているのが特徴であり，オンラインでも利用できるようになっている．

Biobankの利用は研究や統計に限定されているが，学術機関だけでなく民間企業も活用できる．Biobankのデータを利用した研究成果として，がん患者とその遺伝子データを併せて解析し，がん治療に関する特定の医薬品について効果がある集団とない集団の比較による要因分析等の事例がある．

■表-2　Biobankの主なデータベース[3)]

データベース名	内容
Patobank	1,700万人の試料に関するデータベース
Danish Cancer Society	がん協会に登録している57,000人の患者のデータベース
Rigshospitalet	コペンハーゲン大学病院の患者のデータベース
Danish Cancer Biobanks	がん患者のデータベース
Danish National Birth Cohort	国民出生コホートのうち，600,000人についてのデータベース
COSPAC	喘息を持つ子供のデータベース
DD2	2型糖尿病のデータベース
Danish Twin Registry	86,000件の双子登録データベース

Biobankにおけるデータ結合

Biobankでは，収集しているデータ以外に他のデータと結合して研究を行えることが大きな特徴である．具体的には，表-3に示すデータとの結合が可能になっている．

これらのデータの結合は国民IDであるCPR番号によって行われており，出生状況や家庭環境，過去の病歴等を踏まえた多面的な疫学的研究を行うことが可能となっている．

Biobankの制度的な位置付け

Biobankで取り扱われる遺伝子データはデンマークにおいて機微情報であり，事業の開始にあたっては，医療研究倫理委員会（NATIONAL VIDENSKABSETISK KOMITÉ），データ保護

■表-3　Biobankと結合可能なデータベース[3)]

データベース名	内容
住民登録レジストリ	CPR番号，氏名，性別，生年月日，出生地，市民権，住所，両親の情報，配偶者の情報等
国民患者レジストリ	CPR番号，受診日，受診医療機関，診断結果，処置，検査結果等
病理レジストリ	CPR番号，病理学的検査結果（学者，検査の種類，検査結果等）等

庁（Datatilsynet）等による審査が行われた．医療研究倫理委員会は医療研究における研究倫理審査に関する法律（Lov om videnskabsetisk behandling af sundhedsvidenskabelige forskningsprojekter）に，データ保護庁は個人データの処理に関する法律（Persondataloven）に基づいて，Biobank が法律に違反するものでないかどうかを審査し，認可されている．結果，データのセキュリティの担保が必須であること，データの提供は研究目的または統計目的に限られること，データ保護庁が事前承認した対象に対してのみデータが提供可能であること等が定められている．

なお，GDPR における Biobank の扱いについては，定かではないが，第 89 条の公共の利益における保管目的，科学的もしくは歴史的研究の目的または統計目的のための取り扱いに関する保護措置および例外，そして CPRD と同様，第 9 条に該当すると考えられる．

また，国民は Biobank の登録に際してオプトアウト手続きを行うことが可能になっているが，その件数はきわめて少ないようである．

データ結合の今後

プライバシーを保護しつつ，異なる組織で収集したデータを結合するニーズは以前から存在していた．我が国の次世代医療基盤法を含め，特に医療分野では，データ結合に向けた制度を含めた仕組み作りが各国で進められている．表 -4 に前述した事例と我が国の次世代医療基盤法を整理する．公的な目的のための TTP は各国で検討が進められており，本稿では取り上げなかったが，米国では EBPM（Evidence-Based Policy Making）を推進するため，公的機関で保有するデータを突合する TTP として，National Secure Data Service（NSDS）が提案されている[4]．また，国民性に依拠する部分もあるが，デンマークのように，用途を限定し，TTP 等を介さず直接，データ結合を認めるということも，公的な分野では進んでいく可能性もある．

一方，データ結合を着実に行うためには，個人の識別子が非常に重要であり，英国，デンマークでは，NHS 番号，CPR 番号という国民 ID を利用しているものの，韓国では国民 ID である住民登録番号は利用できなくなっている．悉皆性，唯一無二性のある ID を使わない形でのデータ結合精度の担保等も今後のデータ結合において課題となる可能性もある．

このようなデータ結合技術や秘密計算等の新たな技術と法制度面，双方の対応により，プライバシー保護とデータ結合を両立し，より有用なデータ活用が，進むことが期待される．

■表 -4　組織間データ結合の事例

国	英国	韓国	デンマーク	日本
事例	CPRD	―	Biobank	―
TTP	NHS Digital	専門機関	不要	認定事業者
結合方法	NHS 番号等による結合鍵提供	住民登録番号は利用不可	CPR 番号による結合	被保険者番号等（想定）
法制度	NHS 法，ICO ガイドライン等	匿名化ガイドラインで規定	研究倫理，個人データ保護の法律に準拠	次世代医療基盤法

参考文献
1) https://www.cprd.com/intro.asp
2) Office for Government Policy Coordination, Ministry of Interior, Korea Communications Commission, Financial Services Commission, Ministry of Science, ICT and Future Planning, Ministry of Health and Welfare : Guidelines for De-identification of Personal Data.
3) http://www.biobankdenmark.dk/index.html
4) Commission on Evidence-Based Policymaking : The Promise of Evidence-Based Policymaking .

（2018 年 6 月 29 日受付）

美馬正司　tmima@hitachiconsulting.co.jp
大学卒業後，シンクタンク等を経て現職．総合研究大学院大学複合科学研究科情報学専攻（博士課程）単位取得退学．情報大航海プロジェクト等，国の大規模プロジェクトのプロジェクトマネジメントやプライバシー等，関連した制度面の検討に従事．