

医療等情報の二次利用に関する技術作業班 構成員提出資料の要約（1）

東京大学情報基盤センター 清水 央子 構成員

※構成員による御発表を事務局の責任でまとめたもの。

項目	内容
様々なデータベースの違い、利活用上の課題	<ul style="list-style-type: none"> ● 次々と新しいデータベースが構築されている中、個々のデータベースから得られる情報は限定的であるものの、それぞれの性質を理解しうまく組み合わせることで、有益な知見が得られる可能性が高まる。 ● データを集積するだけでは利活用は進まず、データの標準化、インプット情報の質の向上、マスターの整備、同一患者に関する情報の連結性といった、解析前の下準備段階にかかる負荷を軽減したり、過度になり過ぎない適正なセキュリティ要件に基づく利用環境を整備したりすることが重要である。 ● 申請手続きの簡素化と審査の迅速化についても検討する必要がある。
マスターの整備	<ul style="list-style-type: none"> ● データ解析という二次利用を想定し、解析基盤で共有するグランドマスターの構築が必要である。マスターとしては傷病名、医薬品、臨床検査項目等が挙げられているが、医療機関の全国版も対象に加えるべきと考える。 ● 特に、移転、経営母体の変更等に伴う医療機関コードの変更、販売会社、名称の変更等に伴う医薬品コードの変更等に対応するために、各マスターについては時系列での管理も必要である。
解析環境の整備	<ul style="list-style-type: none"> ● 情報の価値を最大限保ちつつ、過度になり過ぎないセキュリティ要件を検討すべきである。 ● セキュリティ管理の対象を考える上で、悪意ある第三者による情報不正取得、悪意ある利用者による目的外使用、データから個人が特定できてしまう可能性等を考慮する必要があるが、これらを一律に考えるのではなく、特に「悪意ある」行為については必ずしもシステムのセキュリティ対策ではなく、審査プロセス等の運用で対応することも考える必要がある。 ● 解析環境の準備に係るデータ提供者と利用者双方の負担が大きくなり過ぎないように、共通の解析基盤を構築して運用していくことは有用と考える。 ● 必要かつ十分なネットワークのセキュリティ要件を定義し提供することで、共通の解析基盤へのアクセス性が高まり、全体のコスト・負荷が軽減し、データ利用のハードルが下がる。これにより二次利用が活性化することが、本グループの目的に合致するものと理解している。



NDB解析に必要な「マスターデータ」

第1回医療等情報の二次利用に関する技術作業班
令和6年2月8日
資料4-1

例示

主要マスターデータ	一般公開されているマスターの種類
傷病名マスター (約26,000レセ電処理用傷病名)	「診療報酬情報提供サービス(厚労省保険局)」1) 「社会保険診療報酬支払基金」1) 2) 「医療情報システム開発センター[MEDIS]」2) 3)
医薬品マスター (約21,000の医療用医薬品)	
診療報酬マスター (医科診療行為: 約7,500)	
医療機関マスター (全国約18万医療施設)	全国版のマスターは存在しない 研究者各自が各地方厚生局のサイトより47都道府県分の医療機関リストをダウンロードし統合

共通課題: いずれもある時点でのマスターであり
過去に遡った累積情報を収載していない

データ解析という“二次利用”を想定したマスターデータが整備されていない

- 1) 別サイトで運用されているがソースは同じ
- 2) 「傷病名」においては共同で更新
- 3) 「診療報酬」においては、手術・検査など一部の項目のみ

医療等情報の二次利用に関する技術作業班 構成員提出資料の要約（2）

独立行政法人医薬品医療機器総合機構 山口 光峰 構成員

※構成員による御発表を事務局の責任でまとめたもの。

項目	内容
MID-NET®の概要	<ul style="list-style-type: none"> ● 薬機法の基準に基づき高い信頼性が確保された医療情報データベースであり、医薬品安全対策等に活用されている。 ● 各拠点病院にデータベースが設置されている点で公的データベースとは異なるものの、MID-NET®の構築・運用時に得られたデータの品質管理や標準化等に関する技術・知見を参考にすべきである。
データの品質管理（信頼性確保）	<ul style="list-style-type: none"> ● MID-NET® Real-time Data-quality Assurance：GPSPに準拠した管理・運営を実現する信頼性確保のための仕組みを運用している。 ● MID-NET®品質管理計画（信頼性確保の3要素） <ol style="list-style-type: none"> ①システムの品質管理：データ送受信プロセスの各機能を検証している <ul style="list-style-type: none"> ・データ取り込み機能、標準化機能、抽出機能、送信機能、SAS変換機能 ②データの品質管理（オリジナルデータとの一致性確認）：全データ種別を対象に、定期的に元データとの一致性を確認している ③日常的品質管理：各プロセスの稼働状況を日々確認している <ul style="list-style-type: none"> ・データ受信件数、データファイル格納状況、システム稼働状況、スクリプト処理状況 等
データの標準化	<ul style="list-style-type: none"> ● MID-NET®では、医療機関で独自に用いられているローカルコード・標準コード変換表を継続的に維持管理し、統合データベースのデータ取込み時にデータ標準化を実施している。 ● 必要に応じて、医療機関への調査も実施し、適切な標準コードを決定している。 ● 検体検査について、検体検査実施頻度及び安全対策の観点を考慮し、優先度の高い検査を標準化対象として選定し、結果値換算や単位の統一化を実施している。

(1) 品質管理

- **データの信頼性が確保されていなければ、データの二次利用は困難である。**
- 電子カルテは、全国統一されておらず、同一ベンダー製であっても医療機関の事情によりカスタマイズされている。電子カルテ更新、部門システムの変更、運用方針変更等があるため、データ送信状況は常に変化する。→**医療DXによる統一に期待！**
- 現在の状況において、データベースの信頼性を確保するためには、システム稼働前に一度確認するだけでは不十分であり、継続的な確認が必要である。
- MID-NET[®]ではMRDA手法により**継続的な品質管理作業を実施**している。

(2) 標準化

- **データが標準化されていなければ、データの二次利用は困難である。**
- 医薬品、検体検査等のコードは、多くの医療機関で、標準コードで管理されておらず、ローカルコードで管理されている。→**医療DXによる統一に期待！**
- MID-NET[®]では実態に沿った**継続的な標準化作業を実施**している。

- 今回の検討に際し、MID-NETの構築・運営時に得られた標準化、品質管理等に関する技術・知見について情報共有が可能。
- 今回の検討は厚生労働省所管DBが対象であり、MID-NETとデータ収集経路が異なる部分もありますが、MID-NETの手法（考え方）を参考に、それぞれのデータ収集方法に適した手法を検討することが重要。

医療等情報の二次利用に関する技術作業班 構成員提出資料の要約（3）

一般社団法人医療データ活用基盤整備機構 岡田 美保子 構成員

※構成員による御発表を事務局の責任でまとめたもの。

項目	内容
診療データの収集に関する課題	<ul style="list-style-type: none"> ● 各医療機関では院内の業務目的に独自マスターが作成されており、多施設利用には標準化されたマスターの整備が必要である。 ● 電子カルテ製品ごとにデータ形式が異なるため、二次利用のためのデータ入力やシステム整備に関して様々な課題が生じている。
解析に関する課題	<ul style="list-style-type: none"> ● データベースごとに標準規格及び標準コードが統一化されていないため、解析が困難である。 ● 公的データベースと民間データベースを連結させる「キー情報」の整備が重要である。 ● 死亡情報等の長期のアウトカムの評価や治療成績の研究が困難である。
医薬品コードの標準化	<ul style="list-style-type: none"> ● 厚生労働省標準規格であるHOTコードマスターは院内業務用ではないため十分普及しておらず、添付文書に記載されているYJコードを推奨してはどうか。 ● 医薬品の用法・用量・単位の標準の確立、普及を図ることが検討される。
傷病名コードの標準化	<ul style="list-style-type: none"> ● 医療等情報の一次利用の議論を踏まえつつ、ICD-10/11を推奨する方針としてはどうか。 ● 癌のステージ分類など、版が変わるコード体系については併せてバージョンも管理する必要がある。
検体検査コードの標準化	<ul style="list-style-type: none"> ● 厚生労働省標準規格であるJLAC10コードは医療機関内ではほとんど使用されていない。 ● 新規の検体検査にはJLAC11コードを割り当てることとするが、既存の検査に対してはJLAC10コードが割り当てられている場合があることから、その対応表を整備していくこととしてはどうか。
標準化全体について	<ul style="list-style-type: none"> ● 継続的な標準コードの維持・管理が困難であることから、データ標準及び標準コードのマスター整備・管理等を一元的に行う組織・団体等を設置する必要がある。 ● 院内システムの上流工程で、二次利用に用いる標準化コードが入力される仕組みを作る必要がある。 ● 国際的な共同研究での利用も見据え、異なるアプリケーションやデータソース間で情報を交換できるコモンデータモデル（OMOP等の国際標準規格）を検討してはどうか。

RWDの二次利用 — 医薬品（薬剤）・検体検査の標準化に係る課題

医薬品（薬剤）	課題	解決策/改善策の例
厚生労働省標準規格	<ul style="list-style-type: none"> 厚生労働省標準規格では「HOTコードマスター」（HOT基準番号を他のコードと対応付けたもの）を指定しており、厚労省事業、厚労科研等では厚生労働省標準規格に準じる必要がある。 「HOTコード」は院内業務用ではなく、院内で活用されていない。 厚生労働省DB事業では「HOTコード」に変換して保存する必要がある。 医療従事者の負担、エラー、継続的な標準コードの維持管理に課題がある。 	<ul style="list-style-type: none"> 医療施設内で確実に導入される標準コードを推奨するため、厚生労働省事業で「YJコード」を利用できるようにしてはどうか。 薬機法改正で添付文書と「GTINコード」が紐付けられたので、GTINコードと他の標準コード（YJコード、HOTコード）を紐付けたマスターの整備を進めてはどうか。
用法・用量・単位	<ul style="list-style-type: none"> RWDでは一般に医薬品（薬剤）の投与量を統一的に取得できない。 電子処方箋の医薬品の用法・用量等について議論がなされている。 	<ul style="list-style-type: none"> 医薬品の用法・用量・単位の標準の確立及び普及を図ってはどうか。
検体検査	課題	解決策/改善策の例
JLAC10 JLAC11	<ul style="list-style-type: none"> 検体検査結果項目コードは施設内でローカルコードが使用されている。 検体検査項目コードの厚生労働省標準規格はJLAC10コードが採用されているが、医療機関の施設内ではほとんど導入されていない。 既存のRWD/臨床DBではJLAC10コードの割が多い。 医療機関の臨床検査部ではJLAC10コードの正確な割当が相当困難である。 JLAC10コードからJLAC11コードに移行しつつあるが、既存DBの検査項目にJLAC11コードを全て割当てることは相当困難である。 	<ul style="list-style-type: none"> JLAC11コードはJLAC10コードの課題を解決したコードであり、新規の体外診断用医薬品の場合はJLAC11コードを付番する方針してはどうか。 対応策として、「マスター対応表」の構築・公開を行ってはどうか。 <ul style="list-style-type: none"> - 複数DB間で共通する検査項目にJLAC10とJLAC11の両方を割当てる。 - 既存DBのJLAC10コードに全てJLAC11を採番する。 <ul style="list-style-type: none"> ※付番: 新規の対外検査診断医薬品にJLACコードを当てること。 ※採番: 既存の院内検査項目等にJLACコードを割り当てること。
JLAC11の現状	<ul style="list-style-type: none"> JLAC11コードは現時点で280項目のみが公開されている。 	<ul style="list-style-type: none"> JLAC11で付番されていて、公開されていないものは早急に公開する。 JLAC11が付番されていない項目は、付番委員会でただちに付番する。
JLAC10/11コードの付番・採番・認証に係る組織体制	<ul style="list-style-type: none"> JLAC10/11コードに係る全体的な判断、質問・要望する先の機関・組織が確立していない。 MID-NETではJAC10コードが空欄で差分抽出されたものをガバナンスセンターで割当・返却しているが、この継続は費用対効果から困難である。 	<ul style="list-style-type: none"> 日本全体でデータが使えるためには管理・付番、採番し、日本全体に普及させ、オーサライズできる非営利の組織及び仕組みを確立する。
院内システム対応	<ul style="list-style-type: none"> DB登録時にJLAC10/11コードを割り当てる方法では他の用途に利用できず、維持管理が困難である。 	<ul style="list-style-type: none"> 院内システムの上流工程で、JLAC10/11コードが入る仕組みを作り、一次利用、二次利用ともに標準化を図る。
過去データ	<ul style="list-style-type: none"> 長期のデータに関しては、方法、機器、試薬等が変更される場合に値が影響するが、現状のDB上にこれらの情報はない。 過去に遡って院内システムから取得したデータの信頼性の問題もある。 	<ul style="list-style-type: none"> 検査の仕様等を変更する場合に、医療機関等のマスターで登録するとともに、この情報をDBに反映させる必要がある。 具体的な目的ごとに対応するのが現実的である。

RWDの二次利用 — 傷病名コードの標準化に係る課題

傷病名	課題	解決策/改善策の例
厚生労働省標準	<ul style="list-style-type: none"> 厚生労働省標準として、医療情報システム開発センター（MEDIS）の「標準病名マスター」の病名及びICD10コード（国際疾病分類）が国内で普及している。 ICD-11コードは、分類コードだけでなく、URI（Uniform Resource Identifier）を用いてより詳細な情報管理が可能であり、ICDコード、ファウンデーションID、用語IDを使用できる。 	<ul style="list-style-type: none"> 厚生労働省標準として、医療等情報の一次利用の観点も踏まえつつ、ICD10/11コードの使用を推奨する方針としてはどうか。
真の病名	<ul style="list-style-type: none"> 電子カルテ情報には保険傷病名（保険請求目的）はあるが、「真の病名」がない。 	<ul style="list-style-type: none"> ICD10の他に医師が考える「臨床病名」を電子カルテで入力できるようにすることが考えられ、実施している医療機関もある。 米国では電子カルテからの表現型抽出（EHR Phenotyping）による「真の病名」あるいは記載されていない病名の推測方法について研究されており（※）、大規模DBへの適用も考えられる。 ※Electronic health record phenotyping improves detection and screening of type 2 diabetes in the general United States population: A cross-sectional, unselected, retrospective study. J Biomed Inform. 2016 Apr;60:162-8
その他	課題	解決策/改善策の例
標準コードのバージョン更新及びデータ蓄積	<ul style="list-style-type: none"> 癌ステージ分類UCCなどは版が変わるが、蓄積されているデータには、その情報は通常は反映されていない。 	<ul style="list-style-type: none"> ICD10からICD11、JLAC10からJLAC11のように大きな改訂では、変わるコード体系の名称自体が変わり、新旧が混在することはない。 途中で、マイナーに変更があり、版が変わるコード体系については、コード体系とバージョンをあわせて記録する。
施設ごとのオーダー運用等の違い	<ul style="list-style-type: none"> 施設により薬剤や検査のオーダーの運用の違いがあり、医療等情報の二次利用に影響する可能性がある。（例：COVID-19でワクチンが注射オーダーでなく処置オーダーとなっている場合がある。PCR検査結果が検査オーダーの場合と細菌検査オーダーの場合もある。） 処置オーダーと処方オーダーで機能が異なり、データ内容も異なる場合もある。 病院の運用は簡単には変更できない。 ベンダー依存とは限らずローカルルールで動いている場合もある。 	<ul style="list-style-type: none"> 困難な課題ではあるが、RWDを生成する病院情報システムの機能として、医療等情報の二次利用の観点から課題を整理し、推奨を公表してはどうか。

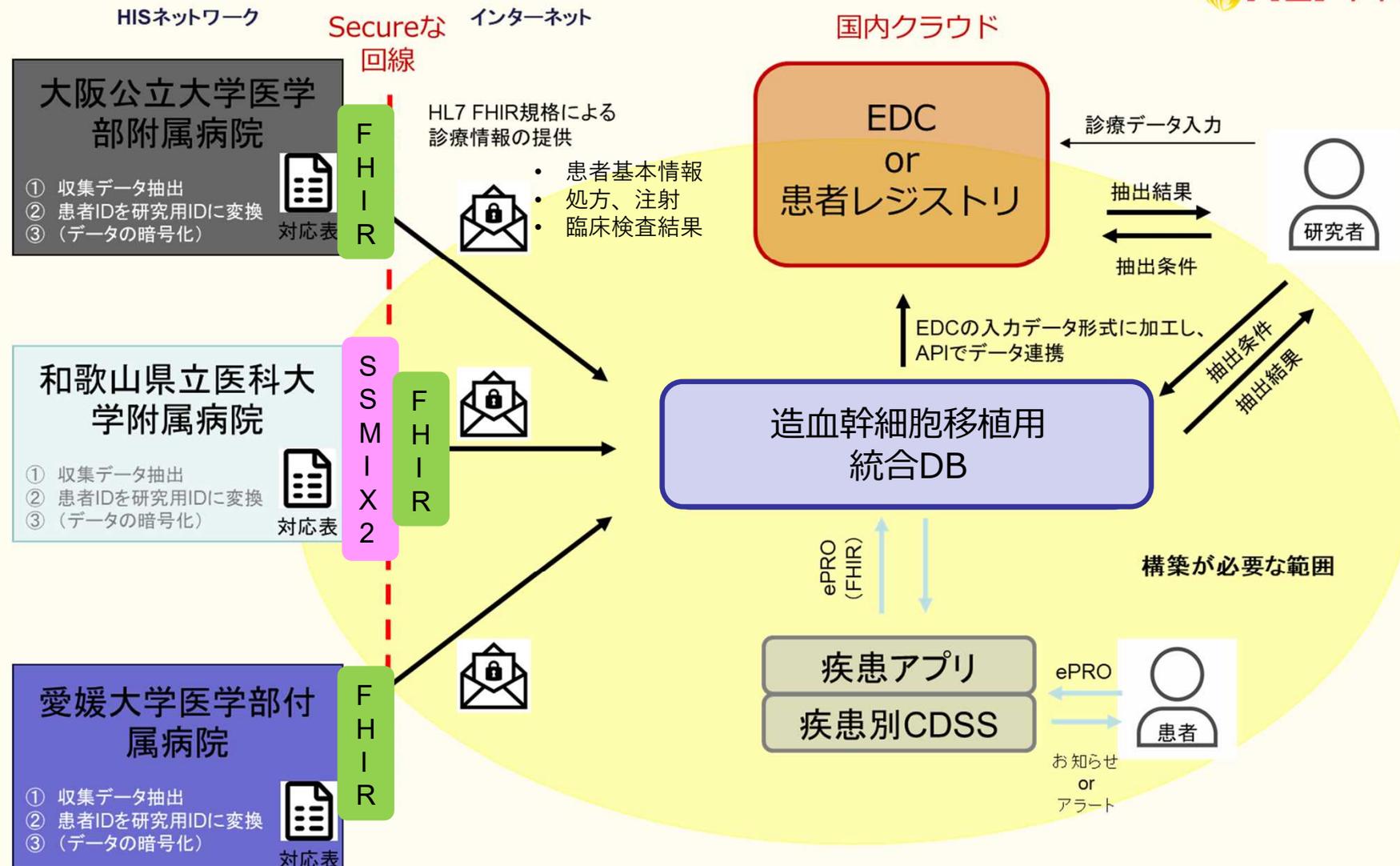
医療等情報の二次利用に関する技術作業班 構成員提出資料の要約（４）

大阪公立大学医学研究科臨床検査・医療情報医学 岡村 浩史 構成員

※構成員による御発表を事務局の責任でまとめたもの。

項目	内容
情報連携基盤の構築、体制と実用例	<ul style="list-style-type: none"> ● 全ゲノム解析等実行計画：がん患者と難病患者の全ゲノムデータと診療情報を統合した情報活用基盤の構築を目指している。 ● Genomics England：希少疾患患者とその家族、がん患者の全ゲノム情報と一次臨床データ、HES（入院、外来、救急情報等）、NCRAS（がん登録情報）等が情報連携基盤上で統合して解析できる。 ● 造血幹細胞移植患者の多施設電子カルテデータと電子的な患者報告アウトカムとを統合したデータベースを構築し、集積したデータを機械学習等で解析する臨床判断支援システムを開発中である。
JASPEHRプロジェクト	<ul style="list-style-type: none"> ● 疾患領域や臨床研究プロトコルごとに必須の入力項目を定めてテンプレートを定義しておき、各社の電子カルテにインストールすることで、電子カルテが異なっても統一形式でデータ収集ができる仕組みが開発されている。医療機関や電子カルテベンダー間の垣根を越え、様々な診療業務の運用と情報の標準化が実現し得る。記述仕様として、HL7-FHIRを採用している。
コードの標準化、OMOPの紹介	<ul style="list-style-type: none"> ● 検査コードとして、試薬の承認時等にJLACコード対応を行うべきであるが、普及が進んでいない。日本の全体最適として、誰がどのタイミングでどの範囲を付番することが適切なのか、ワークフローを整備しなければならない。マッピング作業に必要な人材、予算等を確保する必要がある。 ● OMOP：厳格に定義された国際的なコモンデータモデルであり、複数のデータベースに対して共通の解析プログラムが適用でき、解析を支援する多くの分析ツールが提供されている。国際共同研究を見据えた二次利用のためには、国際標準コード（SNOMED、RxNorm、LOINC等）へのマッピングも必要である。Network study方式では、各施設の個票を外部に出す必要がなく、各データベースの分析結果を集めて統合することができる。 ● Genomics EnglandもOMOPへの対応を進めている。
その他	<ul style="list-style-type: none"> ● 医療DXの推進によって大規模なデータが収集され、リアルタイム性も高まる。ある臨床研究に参加している患者データが、他の研究目的で二次利用されてしまう可能性があることから、進行中の臨床試験の被験者であることが判別できる情報等を管理してデータベースに加える必要があるのではないかと。

造血幹細胞移植患者の臨床判断支援システム開発



医療等情報の二次利用に関する技術作業班 構成員提出資料の要約（5）

情報処理推進機構デジタル改革推進部 田辺 里美 構成員

※構成員による御発表を事務局の責任でまとめたもの。

項目	内容
情報システムの利用環境の変化	<ul style="list-style-type: none"> ● かつては情報システムの利用が院内で完結していたが、現在は外部との連携が加速化している。一方で、セキュリティ対策が従前の「外部接続なし」文化から「外部接続あり」環境を守る文化へ切り替えられていない。
監視対象について	<ul style="list-style-type: none"> ● これまでの外部環境との境界線の監視のみならず、攻撃者に侵入されることを前提として、内部環境も情報資産ごとに監視する体制が必要がある（EDR等）。
ゼロトラスト・セキュリティモデルの考え方と必要性	<ul style="list-style-type: none"> ● クラウド利用により環境や接続先が多様化し、攻撃者の手法・経路も多様になってきたことから、ネットワーク上に信頼できる環境はないというゼロトラスト・セキュリティモデルの考え方が必要となる。 ● 情報連携基盤では、利用者ごとの本人確認（認証）、アクセスできる範囲の確認（認可）が必要である。 ● クラウドサービスはマネージドサービスであるため、環境の設定が変化し得ることに留意が必要である。
二次利用基盤のセキュリティ管理	<ul style="list-style-type: none"> ● 認証方式：基盤へのアクセス者の洗い出しと個別の設定を行い、本人の二要素認証、認可コントロールが必要。 ● 通信暗号化：仮想デスクトップの利用では、インターネット上に仮想でプライベート空間を構築するIPSec VPNを用いることが想定される。閉域網や暗号化の過信は禁物である。 ● ログの活用と緊急対応：システムの全体構成図/情報資産のリスト化、収集するログの種類（正規化のため）と発報ルール設定（SIEMの活用）、発報を受けた後の対応フロー設定、緊急対応マニュアル・フローの作成（CSIRT立ち上げ基準の設定、立ち上げ後の対応手順等）を行うことが重要である。 ● Visiting環境でのデータ保護のルールの例としてGenomics EnglandのTRE環境の紹介：Airlock（研究環境からエクスポートできるデータの制限、事前承認の必須化、ウイルスチェック等）の考え方を利用してファイルのインポートとエクスポートを行う。 ● いずれも、技術とコストとのバランス、利便性による比較・検証を行う。また、システムだけに頼るのではなく、規程・ルールの整備や利用者の教育・啓発を行うことも必要である。
安全管理措置	<ul style="list-style-type: none"> ● 利用者環境：技術的、物理的、人的な安全管理措置が必要 ● 提供者環境：医療情報安全管理ガイド、政府機関等の対策基準策定のためのガイドライン等の各種ガイドラインに準じた対策と対策実行後のモニタリングを行う。データ漏えい事案等においては、罰則を設ける。

二次利用基盤セキュリティ管理の概観

【ログの活用】

- ログの活用による予防と早期発見
- 極力自動化、人手の削減
- 分析ログの取捨選択と相関分析の実施（リアルタイム）
- 証跡として保管するものの管理（アーカイブ）

【認証方法（二要素認証）】

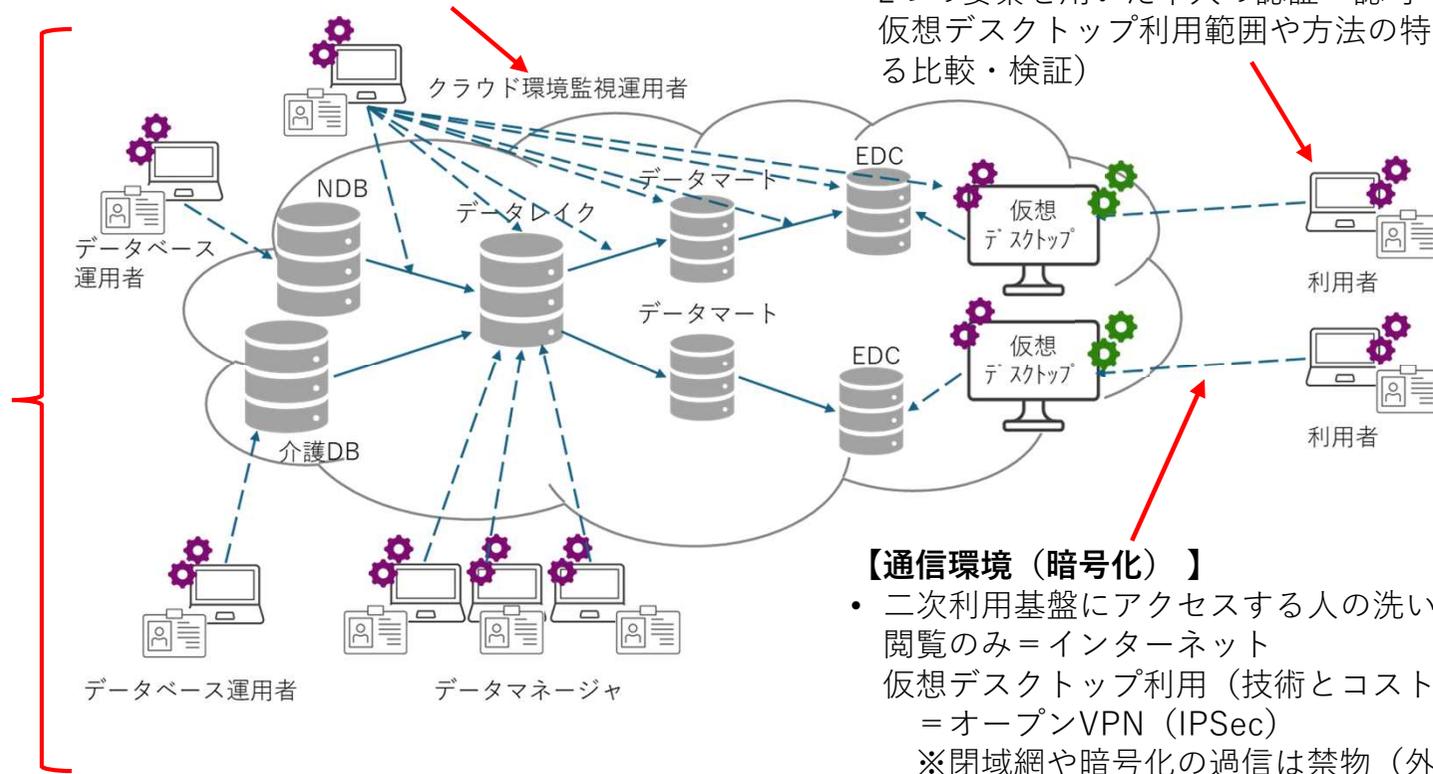
- 二次利用基盤にアクセスする人の洗い出しと個別の設定
情報閲覧のみ、仮想デスクトップ利用（利用方法の分類）等
- 2つの要素を用いた本人の認証+認可コントロールも必要
仮想デスクトップ利用範囲や方法の特定（技術とコストによる比較・検証）

【安全管理措置】

- 技術的
ゼロトラスト
- 物理的
利用場所の制限（公共の
場所はNGなど）
責任範囲の明確化（利用
環境、提供環境）
- 人的
利用前研修
誓約書の提出
アクセス権限別の管理
ルール周知徹底

【緊急対応】

- アラート発報ルール設定
- 緊急度に合わせた対応
ルール設定
- 体制整備



【通信環境（暗号化）】

- 二次利用基盤にアクセスする人の洗い出しと個別の設定
閲覧のみ=インターネット
仮想デスクトップ利用（技術とコストによる比較・検証）
=オープンVPN (IPSec)
※閉域網や暗号化の過信は禁物（外務省事案）

医療等情報の二次利用に関する技術作業班 構成員提出資料の要約（6）

一般社団法人医療データベース協会 足立 昌聰 構成員

※構成員による御発表を事務局の責任でまとめたもの。

項目	内容
データの品質管理に対する考え方	<ul style="list-style-type: none"> ● データの品質管理は、ISO/TS 8000で国際標準化されており、データ提供者からデータ利用者に至るプロセスについて、品質管理のPDCAサイクルを回すことが適切である。
データの品質特性	<ul style="list-style-type: none"> ● データ品質の国際標準規格であるISO/IEC 25012では、15種の品質特性が定義されており、正確性、完全性、一貫性からなる基礎的品質特性と、それ以外の付加的品質特性に大別される。 ● 基礎的品質特性、アクセシビリティ、追跡可能性は、データ入力段階から品質特性が担保されるべき協働領域である。付加的品質特性は需要者によって求められる水準が異なる競争領域でありプロセスも複雑になりやすい。
民間データベースでのクレンジング、コード化等の実際例	<ul style="list-style-type: none"> ● データベースへの取込データの最初のバリデーション（基礎的品質特性、精度の確認等）やクレンジングでは、検査値、投与単位等の学会公式基準等への変換や、いわゆる「その他のX」系データのコード化等を行う。 ● コードマッチしない項目のコード化は、各社が蓄積した辞書や自然言語処理に基づいて行われている。 ● 標準適合性のためのマスターのメンテナンスを頻繁に行い、過去分との統合的な解析を確保している。
品質確保における官民協調と標準化の考え方	<ul style="list-style-type: none"> ● マスターの整備等、民間事業者が相当の資本を投下している知的財産を官民の連携PFのために利用する必要がある場合には、官民構成員の当該マスターの利用機会の確保、相当の利益回収の機会が確保されるべきである。 ● 学術、製薬企業等の医療等データの二次利用者、医療データベース事業者を交えたコンソーシアム等で、協働領域・競争領域の合意形成、品質確保のための商業的に持続可能な仕組みの創出を官民で目指すべきである。
データプライバシー、本人関与に関する論点	<ul style="list-style-type: none"> ● 個人識別性の問題と本人の社会的実在の特定の問題を区別して議論する必要がある。 ● 匿名性の議論の際には、攻撃者の具体的な想定を置いて、統計的な定性的・定量的な評価手法を選択する。 ● データの提供における本人の関与について、オプトイン・オプトアウトで技術的な難しさに大差はないが、本人関与の機会の有無のフラグ管理をネットワーク全体で確保する必要がある。また、死亡情報の取り込みについても制度設計に関する議論が必要。
民間データベースとの連携に関する論点	<ul style="list-style-type: none"> ● データベースの連結解析では、その連結方法に加えて、安全管理措置を確保すべき主体、措置の内容、責任分界点等を検討する必要がある。

データ自体の品質特性

- データ品質の国際標準規格であるISO/IEC 25012では、**15種の品質特性**が定義されている。
- 正確性、完全性及び一貫性（基礎的品質特性）のほか、**最新性や効率性などの商業的な付加価値となり得る品質特性（付加的品質特性）**により市場競争が行われている。

正確性 (Accuracy)	データに誤りがなく、信頼性がある状態か
完全性 (Completeness)	必要な項目が網羅されているか、データに誤りや欠落がなく完全な状態であるか
一貫性 (Consistency)	データに矛盾はないか、データ間で整合性が取れている状態か
信憑性 (Credibility)	出所や更新日が明示されているか
最新性 (Correctness)	更新期間や最終更新日が明示されているか
アクセシビリティ (Accessibility)	環境依存文字等が使用されていないことが明示されているか
標準適合性 (Compliance)	標準に準拠したデータ形式であることが明示されているか
機密性 (Confidentiality)	アクセス制限等の対策が講じられていることが明示されているか
効率性 (Efficiency)	効率化のため、コードの割り当て等を行っていることが明示されているか
精度 (Precision)	データの精度について明示されているか
追跡可能性 (Traceability)	データの変更履歴（変更者／変更日等）が明示されているか
理解性 (Understandability)	データ全体や項目ごとにメタデータが明示されているか
可用性 (Availability)	利用時間帯等について明示されているか
移植性 (Portability)	標準的な形式でデータをエクスポートできることが明示されているか
回復性 (Recoverability)	バックアップ等の障害時の対応方針が明示されているか

クレンジング、コード化、標準化等の実際例（一部）



- 取込データの最初のバリデーション（基礎的品質特性、精度の確認等）やクレンジングでは、検査値、投与単位等の学会公式基準等への変換や、いわゆる「その他のX」系データのコード化などを行う。
- **コードマッチしない項目のコード化は、各社が蓄積した辞書や自然言語処理に基づいて行われている（競争領域）。**
- 標準適合性のためのマスタのメンテナンスは、各マスタについて月1回以上であり、**診療報酬情報提供サービスや薬価基準収載医薬品等の公的な情報については、改定の都度マスタを更新し、過去分との統合的な解析を確保している。**
- 医薬品毎の添付文書に含まれる情報等の付加や解析ニーズの高いコードの付加、多言語化等は、個別の需用者のニーズを踏まえ、匿名加工の前後で実施される。
- 全行程の工程数としては、3,000ステップを超えている* ところもある。

* 構成員個人の所属企業のユースケースに基づく