

3. データ連携・利活用に向けたデータの管理・運用方法

2. では、データ創出・利活用に携わるあらゆる関係者に対して、実際にデータ創出・利活用を進めて効果的に成果創出に至るために理解しておくべき基本概念を示した。3. では、共同研究に携わる関係者に対して、データ連携・利活用に資する形でデータを整備し、管理運用するための方法や、関係者間での事前の調整内容について示す。

特に、共同研究のデータ実務者においては、データ連携・利活用に向けたデータ整備・管理運営等の実務を行う際の参考情報として活用していただきたい。また、共同研究体の代表者や研究資金配分機関においては、組織内でデータ共有のあり方を示す際等に参照していただきたい。

そもそもデータ連携・利活用の成果創出を促進するためには、データの取扱いに関する共通認識が関係者間で醸成されていることが重要である一方、現状、具体的な取組内容までは十分に共有されていない場合も多い。それゆえ、データ連携・利活用の取組に向けた目線合わせさえ、十分に出来ていない場合もある。そこで本章では、データ創出の関係者が、データ連携・利活用を進める際に求められる取組に関して具体的に理解し、データ連携・利活用に向けた共通認識を醸成しながら取組が進められるようになることを目的に、取組や留意点を説明する。

以下では、大きく 3 つの取組に関して、関係者が直面し得る課題とその対応方策に関して具体的に説明する。

● データの整理と可視化

データ連携・利活用に向けて、まずは研究組織として管理しているデータを棚卸しすることが重要となる。具体的には、データに関する各種メタデータを整理するとともに、構造的に整理することで、可視化するといった取組が求められる。本章では、メタデータの整理や、データの構造化といった各種取組に関して、対応方策(3.1)を提示する。

- データ連携・利活用を行う際の方針の策定

データ連携・利活用を組織として円滑に進めていくためには、関係者間でデータの共有のあり方に関する共通する方針(以下「データシェアリングポリシー」という。)、個々の研究データの管理運用計画(以下「データマネジメントプラン」という。)、契約手続き等の策定を進めることが重要となる。そのため、データシェアリングポリシーの策定方法(3.2)、データマネジメントプランの策定方法(3.3)、研究データ全般の公開・共有を行う際の具体的な契約・手続きを検討するにあたっての課題やその対応方策(3.4)などを提示する。

- データ連携・利活用を進める際の関係者間での調整

データ連携・利活用を通じた成果創出を持続的に行うためには、関係者間での様々な調整が必要となる。そこで、データ連携・利活用を進めるにあたって留意すべき関係者間での調整点として、データの成果化(3.5)、利害関係の調整(3.6)、セキュリティやトラストの確保(3.7)、研究対象者とのコミュニケーション(3.8)を行うにあたっての課題やその対応方策などを提示する。

なお、各取組は必ずしも時系列的に実施されるものではなく、個々の研究開発プロジェクトの状況に応じて適宜実施されるものであることに留意されたい。

3.1. データの整理と可視化

<現状と課題>

- バイオ分野は領域の広さから取組の中で得られる情報が広範囲に及ぶため、データの精度や、付属するメタデータの詳細さ、オントロジー²⁸の差異が大きい。
- 上記認識を一致させておかなければ、実証におけるデータマネジメントプランに不備が生じ、データが公開不可能となるリスクや、参加機関の営業秘密・ノウハウが流出するリスクがある。
- 個別の調整の対象となりうるデータの範囲と粒度については、「AI・データの利用に関する契約ガイドライン」や「農業分野における AI・データに関する契約ガイドライン」に整理されている(ガイドラインの対象は研究データのみではない。)。

<対応方策>

- データの整理

同一の組織内でも各研究ユニットによってデータを管理するためのファイルの形式やフォーマットが異なることから、組織内でデータを収集したとしても整理することが出来ないケースが見受けられる。そこで、組織内の各ユニットに向けて、データの属性や特徴を表す付随的な情報(メタデータ)²⁹が記載された同一のフォーマットを共有し、各自同一のフォーマットで整理してもらうことも一案である。図表 15 に記載しておくべき情報の一例を示した。

28 語彙や概念の関係性等を意味する。

29 本章で示すのは個々のデータの属性や特徴を示すメタデータではなく、プロジェクトや実験を通じて得られた一連のデータに対するメタデータのこと。

メタデータの項目例		内容
プロジェクトに関する情報	ID	該当するプロジェクトを特定するための ID ※ID 付与が行われていない場合は、プロジェクトの特徴を示すキーワード等を記載
	概要	プロジェクトの目的や取組概要
	実験 ID	概要する実験情報を特定するための ID ※ID 付与が行われていない場合は、実験情報の特徴を示すキーワード等を記載
	実験概要	実験の目的や取組概要 ※プロジェクトの目的と紐づけて記載
データに関する情報	項目	データセットに含まれる個々の情報 (対象作物・成分単位の情報等)
	概要	データセットに含まれる個々の情報の概説
	粒度	生データ/加工データ/解析データ 等
	取得方法	生データ/加工データであれば測定等を行った実験方法, 解析データであれば解析方法の概説
	取得条件	データを取得した実験条件や環境条件
	ファイル形式	管理されているファイルの形式 (RDF, Excel, txt, 紙媒体等)
	作成者	データの作成主体となった機関・個人
データベースに関する情報	格納先	データセットが格納されているデータベース
	管理者	データベースの管理者 ※データオーナーと同じ場合もある
利用条件等	アクセスレベル	データアクセスに係る方針
	利用条件	利用に際して満たすべき条件等
	データオーナー	データの所有者 ※規約等におけるデータの権利帰属関係を確認の上、記載

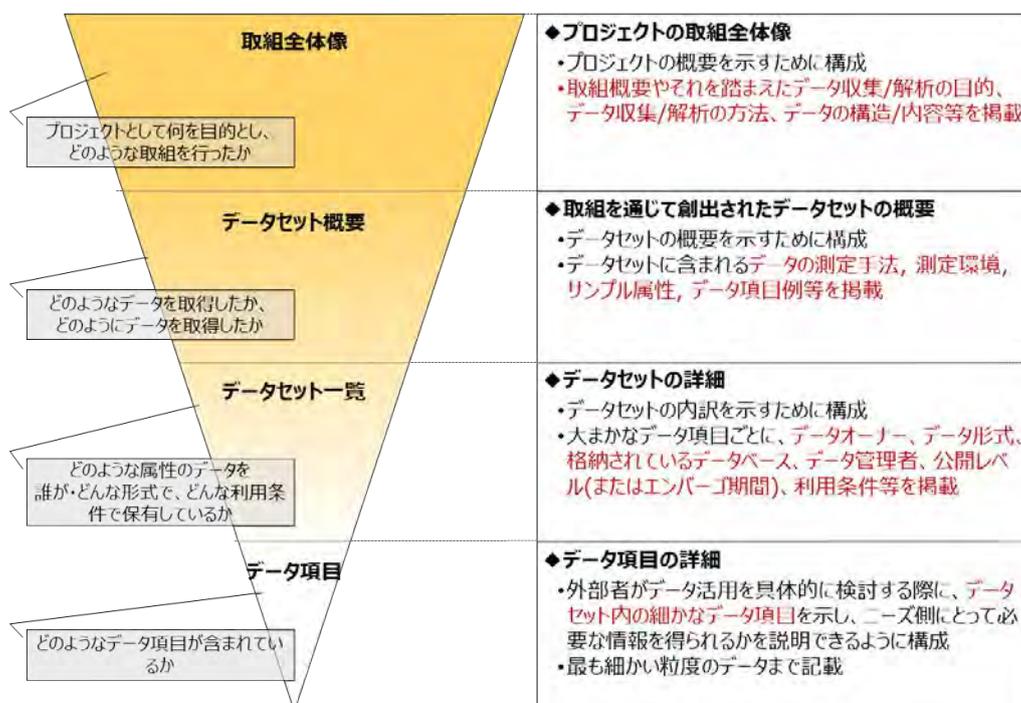
図表 15 メタデータの項目例

<留意点>

● データを整理する際の工夫：データの構造化

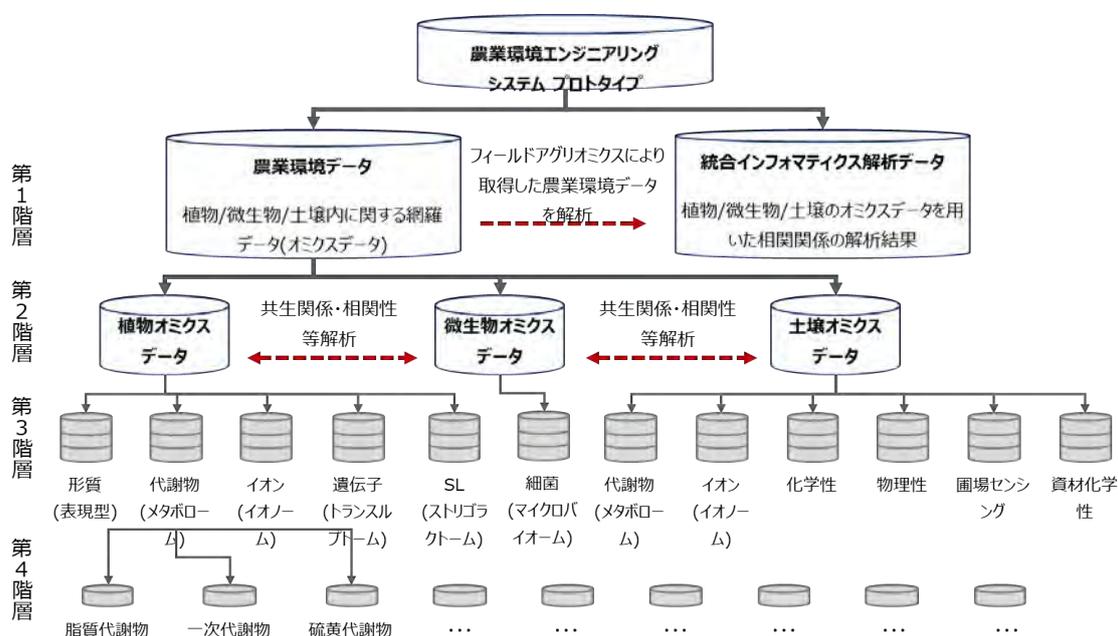
データを項目別に羅列するだけではデータの全体像が見えず、個々のデータの関連性や位置づけが理解できない場合がある。そのため、研究データを整理するにあたっては、データを階層別に構造化し、視覚化して示すことが大切である。

構造化のアプローチとしては二つの段階で整理することも一案である。まず第1段階は、プロジェクトの取組全体像と、その取組において取得した一連のデータの概要、収集したデータを一覧形式で整理することである。「プロジェクトの取組全体像」では、プロジェクトの取組の目的や取組内容が提示されることで、取組の全体感を示すことができる。「データセット概要」では、取組の中でデータを取得した経緯や、取得した一連のデータの位置づけを提示することで、取得したデータの意義や重要性を示すことができる。「データセット一覧」「データ項目」では、個々のデータに関して、「誰が」「どのようなアプローチで」「どのような環境条件で」「何を対象として」取得したデータであるのかが網羅的に整理されることで、どのようなデータであるのか具体的にイメージしてもらいやすくなるといったメリットがある。(図表 16)。



図表 16 データの構造化①：研究組織における取組全体像とデータセットの位置づけの可視化

第2段階は、プロジェクトを通して得られた個々のデータを階層化して整理することである。例えば、「データの取得方法」「データの取得の際の環境条件」「データ取得対象」など適当なカテゴリを定め、カテゴリ別に分類することで、取得した個々のデータ間の関係性が明確になる。図表 17 では一例として、SIP 第2期 バイオ・農業「農業環境エンジニアリングシステム(1C)」にて取得されたデータの階層構造を示しているが、植物・微生物・土壌に関するオミクス解析のデータを対象に、第1階層では「測定したデータか解析後のデータか」を整理し、第2階層では「データの取得対象」、第3階層・第4階層では「測定項目」別に整理を行っている。



図表 17 データの構造化②：各研究データの階層構造化(例)
 (下記資料は SIP 第2期 バイオ・農業「農業環境エンジニアリングシステム(1C)」
 のデータベースの階層構造図)

3.2. データシェアリングポリシーの策定

<現状と課題>

データ連携・利活用におけるデータの取扱いに関して、組織の統率が取れておらず、プレーヤが各々データ連携の取組を行っている、組織で定められた共通のルール・規約に則ったデータの運用が成されない懸念がある。そのため、個々の研究者間での認識違いによる契約関係のトラブルや、具体的な契約等に関する内容の抜け漏れが発生するリスクも考えられる。そこで、3.2では、組織としてデータ連携・利活用におけるデータの取扱いに関する共通の方針を示すための進め方を示す。

<対応方策>

データ連携・利活用を行う際に参照すべきデータの取扱いに関する方針(データシェアリングポリシー³⁰)は、組織内での議論を踏まえながら作成し、適宜アップデートしていくことが重要である³¹。データシェアリングポリシーにおいて、研究成果を連携・利活用していくことの重要性を示しつつ、組織内で個々の研究者を巻き込んで策定していくことで、データ連携・利活用に向けた共通の方針・指針として個々の研究者の間にも浸透させることができる。また、上記の議論を経て策定されたデータシェアリングポリシーを踏まえて、研究組織内で、個々の研究データのメタデータや公開範囲等の取り扱い方針(データマネジメントプラン)、契約手続きの方法等に関して議論していくことが可能となる。

- データマネジメントプランの策定

データシェアリングポリシーで定められたデータの取扱いの方針に則り、個々のデータ別に詳細な公開範囲・共有範囲、公開・共有する際のデータの形式や手続き方法等を定めた資料。

- データ連携・利活用における契約等手続きの策定

契約等手続きについては、通年的に使用している既存の契約書にとらわれず、研究開始前・研究実施中に議論を行い、関係者間で必要となる記載事項を調整することが必要である。

30 AMED「ゲノム医療実現のためのデータシェアリングポリシー」

31 作成主体としては、研究資金配分機関、出資先の代表者・代表機関等が考えられる。また、複数機関によるデータ連携・利活用を行うコンソーシアム等組織では、コンソーシアムとして作成する場合も考えられる。

<留意点>

- コンソーシアム型では意思決定に全参加機関の合意が必要となり時間を要することが少なくないため、参加機関数や成果の活用方針を勘案すると、個別型の契約が適している場合は多い。
- 研究計画書作成、データ利用申請、成果の知財化といった取組内の手続において、参加機関の秘密情報が他機関へ共有される可能性が一定程度存在する。研究事業開始後の参加機関の離脱を未然に防止するためにも、参加機関に共有される可能性がある情報についてはその内容を事前に調整し、契約書や覚書に盛り込むことを検討すべきである。
- 長期間にわたる取組の場合、研究の内容やその進捗によっては研究開始前に作成した文書が更新される可能性がある事を予め参加者と合意しておく必要がある。
- ゲノム関連プロジェクトにおいては、AMED の取組に則りデータシェアリングポリシー³²を設けることが望ましい。
- データ連携・利活用に関する取組に特有の内容は「AI・データの利用に関する契約ガイドライン」及び「農業分野における AI・データに関する契約ガイドライン」に整理されているが、個別案件において特殊性も考えられることや今後想定される分散型連携におけるデータ連携に関しては個別の検討が必要である。
- 研究インフラ、社会実装する仕組みが欧米で整備されつつある中で、日本の研究者も国際的な動向をチェックしながら、産業実装化を行う上での準備に取り組む必要がある。欧米においては、競争的資金制度への参加条件として、データシェアリングポリシーに関してルールを課すことで、データ連携・利活用の取組を加速させようとしている。
- 海外では、例えば NIH のデータの運用・管理方針を定めた「Data Management and Sharing Policy」³³、欧州の H2020 Programme Guidelines on FAIR Data Management in Horizon 2020³⁴が公開されている。グラントと紐づいた研究プロジェクトのオープンサイエンスの取組や各国のデータ連携・利活用に向けた制度・仕組みを踏まえて、国内のルールとの整合性を確保しつつ国内外でのデータ連携に向けた取組を進めていくことが重要である。

32 <https://www.amed.go.jp/koubo/datasharing.html>

33 <https://sharing.nih.gov/data-management-and-sharing-policy>

34 https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

3.3 データマネジメントプランの策定

<現状と課題>

政府研究開発プロジェクト等の国費を用いて創出された研究データに関しては、「オープン・バイ・デフォルト」³⁵の原則から、個々のオープン・クローズ戦略に応じた制限公開・公開に至るまでのロードマップや、公開のための仕組み作りが求められる。そこで、研究データを公開するためのデータ基盤や座組を作るにあたって、個々の研究データの公開レベルや権利帰属関係が整理されたデータマネジメントプランを定めておくことが不可欠となる。しかし、少なくともバイオ分野においては、データマネジメントプラン策定のための取組は、依然として十分ではない。その要因としては様々考えられるが、その一つとして、研究データの管理者が各データの公開レベルや権利帰属関係を明らかにすることの重要性を理解できておらず、現場研究者も管理しているデータを棚卸し整理するためのインセンティブを感じることができていないということも考えられる。3.3. では、データマネジメントプランの重要性を踏まえ、データマネジメントプランを個々の研究者が効果的に運用するための対応方策を示す。

<対応方策>

データマネジメントプランの作成はデータ連携・利活用の取組を行う上で欠かせないプロセスであることに鑑み、データ連携・利活用を専門とする部署あるいは専門人材による総合的な監督及び個々の研究者へのレビューを行う必要がある。また、研究チームの設立当初から、データ連携・利活用を行うことや、それを実現するためにデータマネジメントプランの作成を行うことを研究チームのメンバーに共有することで、研究代表者だけでなく現場の研究者にも意識するよう促していくことも重要である。また、各研究チームにおけるデータマネジメントプランの作成者にとって、現状管理・保有しているデータを棚卸し、各データの公開レベルや権利帰属関係を整理することは負担がかかる取組である。特に政府研究開発プロジェクトに参画している民間企業の場合、整理に時間がかかることだけでなく、管理・保有するデータを公開することが競争力を毀損するのではないかという懸念から、心理的なハードルも存在し得る。

こうした作成負担や心理的なハードル等を少しでも解消していくためには、日頃から個々の研究者とコミュニケーションを密に行うことが大切である。また、その際には、データマネジメントプランを整理することにより、①自身が関

35 https://read.oecd-ilibrary.org/governance/open-government-data-review-of-poland_9789264241787-en#page1

与した研究データを保護することが出来ること、②他の研究組織の研究データを活用することが出来るといったメリットを提示することが望ましい。なお、データの管理運用計画については、プロジェクトの開始前から予め検討しておくことが重要であり、個々の研究者には、プロジェクトの期間中、適宜管理運用計画をアップデートしていくことが期待される。

<留意点>

- データマネジメントプラン策定の際には、以下を参照のこと。
 - ・「研究データ基盤整備と国際展開ワーキング・グループ報告書」
 - ・JST「データマネジメントプラン作成の手引き」³⁶
 - ・AMED「AMED 研究データ利活用に係るガイドライン及びデータマネジメントプランの提出の義務化について」³⁷
 - ・SIP 第2期 バイオ・農業「データマネジメントプラン記載要領」³⁸
- データマネジメントプランの参考例については、新エネルギー・産業技術総合開発機構(NEDO)「NEDO プロジェクトにおけるデータマネジメントについて」内「データマネジメントプラン兼簡略型データマネジメントプラン」を参照のこと。
- 上記のほか、国際連携を見据え本ガイドブックの他の箇所で参照している、諸外国のルールも参照しつつ国際的な整合性にも留意すること。

<参考情報>

データ連携・利活用やデータマネジメントプランに関する国際動向に関しては次ページに整理している。データマネジメントプランの策定の際には参照されたい。データ連携において、データマネジメントプランを作成することの重要性は海外でも強く謳われており、欧米をはじめとする各国の政府研究開発プロジェクトの一部では、データマネジメントプランの作成が義務付けられている。

36 https://www.jst.go.jp/kisoken/presto/manual/data_houshin.pdf

37 <https://www.amed.go.jp/koubo/datamanagement.html>

38 http://www.naro.affrc.go.jp/brain/sip/files/sip2_besshi4.docx

国	データマネジメントプランの作成方針例
日本	内閣府「研究データ基盤整備と国際展開 ワーキング・グループ 報告書 図表集」 ³⁹ NEDO「データマネジメントプラン 兼 簡略型データマネジメントプラン」 ⁴⁰ ・ 研究組織間でのデータ共有を円滑にするための方針やデータマネジメントプランの作成方針を提示。
米国	USGS 「USGS Data Management Plan Checklist」 ⁴¹ ・ データマネジメントプラン作成の際に検討すべき項目として、チェックリストを提示。
英国	RCUK 「Guidance on best practice in the management of research data」 ⁴² ・ データマネジメントプランの作成の際の原則について提示。
EU	European Research Council 「Open Research Data and Data Management Plans」 ⁴³ ・ 研究データを公開することで、他組織によるデータ利活用を増やすとともに、透明性と説明責任を促進することを方針として提示。

図表 18 諸外国におけるデータマネジメントプランの作成方針例

39 <https://www8.cao.go.jp/cstp/tyousakai/kokusaiopen/zuhyou2.pdf>

40 <https://www.nedo.go.jp/content/100897759.pdf>

41 <https://www.usgs.gov/data-management/data-management-plans>

42 <https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-GuidanceBestPracticeManagementResearchData.pdf>

43 https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf

3.4. データ連携・利活用における契約等手続き方法の策定

<現状と課題>

データ連携・利活用において、データの作成者やデータオーナー等の権利帰属関係を明確にしておかなければ、取得したデータの公開性が危ぶまれる場合や、データの権利を主張することができなくなる恐れもある。ここでは、取得したデータの公開性を保護し、将来のデータ連携・利活用につなげるための対応方策について示す。

<対応方策>

- データを成果物とするかの判断を含め取組の成果物を定義した上で、データの取扱方針や活用主体を明確にする。公開範囲・共有範囲の策定の際には、以下の点について、個別間での調整や認識共有を行ったうえで、関係者間での認識を一致させることが必要である。
 - ・ データオーナーシップ
 - ・ 公開範囲や公開時期
 - ・ 研究目的か商用目的かを含めた参加機関の利用目的等の個別調整が必要である事項について関係者の認識を一致させる。
- 産官学連携の共同研究に関するモデル契約書については、文部科学省「オープン&クローズ戦略時代の共同研究における成果取扱いの在り方に関する調査」結果であるさくらツールに類型(共同研究組織の形態や成果の活用主体、活用方針等)とともに示されているため、契約書を作成する際には、参照されたい。ただし、今後複雑化する分散型でデータ基盤が連携モデルを想定した場合、モデル契約が想定している一般的な事例に当てはまらない場合は、むしろ当該調査の記載に拘らずに、専門家と協議しながら、オープンクローズに関連する知的財産関連法、不正競争防止法、個人情報保護法、人権問題、倫理、安全保障の観点などの様々な視点から契約書の設計を行う必要がある。

<留意点>

- データ利活用の際の条件やコストが不明確であることが企業による利活用を妨げている側面もあることから、例えばデータを非公開とする場合、データの存在についての公開可能性や、公開可能な場合の利用条件及び利用に係る諸経費も定義しておくことが望ましい。

- バイ・ドール制度において寄託者に帰属させうる知的財産権は、
 - ・ 特許権、特許を受ける権利(特許法)
 - ・ 実用新案権、実用新案登録を受ける権利(実用新案法)
 - ・ 意匠権、意匠登録を受ける権利(意匠法)
 - ・ 著作権(著作権法)
 - ・ 回路配置利用権、回路配置利用権の設定の登録を受ける権利(半導体集積回路の回路配置に関する法律)
 - ・ 育成者権(種苗法)と定義されていることを認識する必要がある。

- 協調領域として公益性が高い知見や(例えばデータ基盤でのデータ蓄積の在り方に対する知見等)、企業を含む参加機関が共同で実施した解析結果や開発したアルゴリズムについては、その成果物としての特許を参加機関で共有する場合がある。こういった特許は幅広い活用が期待できる一方、取組終了後の活用の際し調整に時間を要する場合がある。そこで、例えば成果創出を主導した研究者の特許としつつ、実施権の運用で参加企業のインセンティブを担保し、成果を幅広く活用するといった仕組みも検討すべきである。

3.5. データの成果化

<現状と課題>

データを成果とする場合でも、得られたデータをそのまま成果とすることは困難であり、利用者が簡便に利用できる状態にする作業が必要である。そのために満たすべき条件は FAIR 原則⁴⁴として知られているが、研究事業においてその実現に必要な資源が確保されないまま取組が行われ、連携・利活用を妨げる要因となっている場合がある。これらを解消するために必要なタスクを 3.5.1 以下に記載する。

<対応方策>

利用者が簡便にデータを利用可能な状態にするために必要な人的・経済的資源を確保する。

<留意点>

以下のような、適切な資源確保を妨げている要因が考えられ、中長期的にはこれらを解消する仕組みを構築する必要がある。

- 作業に一定程度の資源を確保することの必要性が認識されていないこと
- 作業を担当した結果得られる、取組の成果となるデータが研究者としての成果とはみなされない場合があること

3.5.1. 標準化、データクレンジングとデータ形式の設計

<現状と課題>

- バイオ分野における研究データは手法により異なることが多く、その解決策として、例えばマイクロバイオームに関しては NEDO 先導研究プログラム「ヒトマイクロバイオームの産業利用に向けた、解析技術および革新的制御技術の開発」にて推奨分析法が提案されている。ほかにも日本人のゲノム情報に関しては JST センター・オブ・イノベーション(COI)プログラム及びAMED ゲノム医療実現バイオバンク利活用プログラムにおいて、東北メディカル・メガバンク機構(ToMMo)がジャポニカアレイ⁴⁵を開発し、SIP 第2期 バイオ・農業においても日本食を構成する農産物含有成分の一斉分析法の開発が行われている。

44 <https://biosciencedbc.jp/about-us/report/fair-principle/>

45 <https://www.megabank.tohoku.ac.jp/news/11110>

- 高度分析機器であっても機器トラブル等で一部のデータが得られない事態は起こりうる。また、現在のコホート研究では拠点間で調査票の全質問文や結果の記録方法が統一されている状況にはない。これらにより、得られるデータに重複や欠損値が存在する可能性や、データ連携の際のフォーマットに差が存在する可能性は常にある。
- 従来はデータが研究成果である論文と紐付いていたため、論文から取得方法やデータの再現性等の質の違いを読み取ることが可能であった。しかしながら、論文に紐付けずデータそのものを成果とすると、研究者のインセンティブが不明確になることが多いため、適切なメタデータが作成されない場合がある。

<対応方策>

複数の研究者から得られたデータを連携させ成果とする場合や他のデータとの相互運用性を確保する場合には、必要に応じ以下の作業を実施する。

- データ取得方法の標準化又は標準化を目指して提案されているデータ取得方法との差異の明確化
- 表記揺れ、重複や欠損値への対応、半角・全角や日付・時間といったデータのフォーマットの統一等のデータクレンジング
- 分散型データベースの連携⁴⁶を前提に、格納するデータベースや連携するデータ流通プラットフォームに応じたスキーマやデータ形式の設計、データ取得者・取得方法・品質・利用許諾条件等を整理したメタデータの作成

<留意点>

- データ取得方法は研究目的により統一することが適さない場合もあるため、全ての場合に対してデータ取得方法の標準化を推奨しているわけではないが、将来価値のあるオープンサイエンスに貢献の観点から標準化する品質管理の意識は必要である。
- 調査票の質問文について、連携や標準化の観点から第三者の作成した質問文を用いることも多い。一方で、そういった質問文は産業利用する際に有償と

47 分散型連携の具体的な事例として、薬剤疫学評価分野における世界規模での Real World Data (RWD) 活用のためのオープン規格である OMOP Common Data Model (OMOP CDM) 及びその維持発展を担うコミュニティ OHDSI の取組が参考になる。

<https://fnih.org/what-we-do/major-completed-programs/observational-medical-outcomes-partnership-omop>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9810199/>

https://www.jstage.jst.go.jp/article/jjpe/27/1/27_27.34/_pdf

なる場合があるため、留意する必要がある。

- 標準化には、我が国が世界をリードする場合と、我が国が遅れを取り戻す場合の2種類があることを念頭に、国際連携も意識した戦略的なアプローチが必要である。すなわち、既に国際的な標準やデファクトスタンダードがある領域⁴⁷では、それに準拠してガラパゴス化を避ける一方で、まだそれらが存在しない領域を見出し、我が国の取組の成果として国際標準の取得を検討することが望ましい。
- メタデータ等は、アプリケーションプログラミングインタフェース(API)によるデータ抽出や機械判読によるオンライン統合利用が可能な形式であることが望ましい。
- 将来の利活用を踏まえ、データの品質の管理の枠組み⁴⁸や、継続的にデータが拡充されるようなプロジェクトにおいては、品質維持のためのデータ形式やプロトコルの変更についても科学的に分析と議論が継続できる仕組みの取組についても参考にされたい。

3.5.2. 事業・分野間連携

<現状と課題>

事業間や分野間の連携を想定するときには、3.5.1の内容に加えて更なる検討が必要となる。それらについては、包括的データ戦略や国際的動向を踏まえたオープンサイエンスの推進に関する検討会で議論が進められている。

<対応方策>

事業・分野間連携のあり方は多様であり、一概に対応方策を定めることは困難

47 データの相互運用性の議論に関しては諸外国の取組も参考になる。臨床データに関しては、データの標準化に向けての整理を議論した欧州の TEHDAS の「Recommendations to enhance interoperability within HealthData@EU」における整理の成果なども参照されたい。

<https://tehdas.eu/app/uploads/2022/12/tehdas-recommendations-to-enhance-interoperability-within-healthdata-at-eu.pdf>

また、日本での現場実装の取組が遅れている OMOP CDM の議論も参考になる。

<https://www.ohdsi.org/data-standardization/>

48 欧州の TEHDAS の「European Health Data Space Data Quality Framework」なども健康医療情報の関連では参考になるものと思われる。またゲノムのバイオバンクとの関係では、NIH の ClinVar と ClinGen の Partnership の取組についての考え方についても参考になるものと思われる。

<https://clinicalgenome.org/about/clingen-clinvar-partnership/>

だが、SIP 第2期 バイオ・農業では関連する議論を踏まえ、分野間データ連携基盤等も活用し、以下の取組を行った。一例として参照されたい。

- 分野別に定められているコード群の類型化・関係整理・共有化や、アノテーション形式の共通化を含めてオントロジーの考え方の整理を実施。
- 必要に応じ、バイオ分野におけるベースレジストリの検討、NoSQL-RDB 間のミスマッチ解決の検討を実施。

<留意点>

- 異なる分野のデータを統合し集中型データベースで管理することは困難であり、分野間連携においては分散型データベースの連携を前提とする必要がある。

3.5.3. 事業終了後のデータの取扱い

<現状と課題>

成果としてのデータが整備できた場合も、予算等の確保ができず、事業期間内だけの利活用に限定される場合がある。しかしながら、本来であれば、事業終了後に幅広い機関が再利用も含めてデータを利活用できることが望ましい。

<対応方策>

例えば以下のような、事業終了後もデータを継続的に利活用可能な状態を維持するための方策を策定する。

- 制限公開を含むオープンデータ化
- データ流通プラットフォームを活用した自走
- 運用に係る経費の恒常的な確保(ライフインフラと同等)

<留意点>

- 課金の有無にかかわらず利活用を促進するためには、利用者のアクセス性や発見可能性を高める必要がある。そこで、各種インターネット検索エンジンから到達可能であることや、同じ領域のデータが登録されているカタログサイト^{49, 50}をはじめ農業データ連携基盤(WAGRI)、NBDC等のデータ流通基盤への登録やAPIの作成に取り組むことが必須である。
- 国費以外で取得された関連データの連携・利活用を促進するためのインセンティブ確保がなされていないため、その手段としても、データ流通プラット

49 Integbio : <https://integbio.jp/dbcatalog/>

50 バイオバンク横断検索システム : <https://biobank-search.megabank.tohoku.ac.jp/v2/>

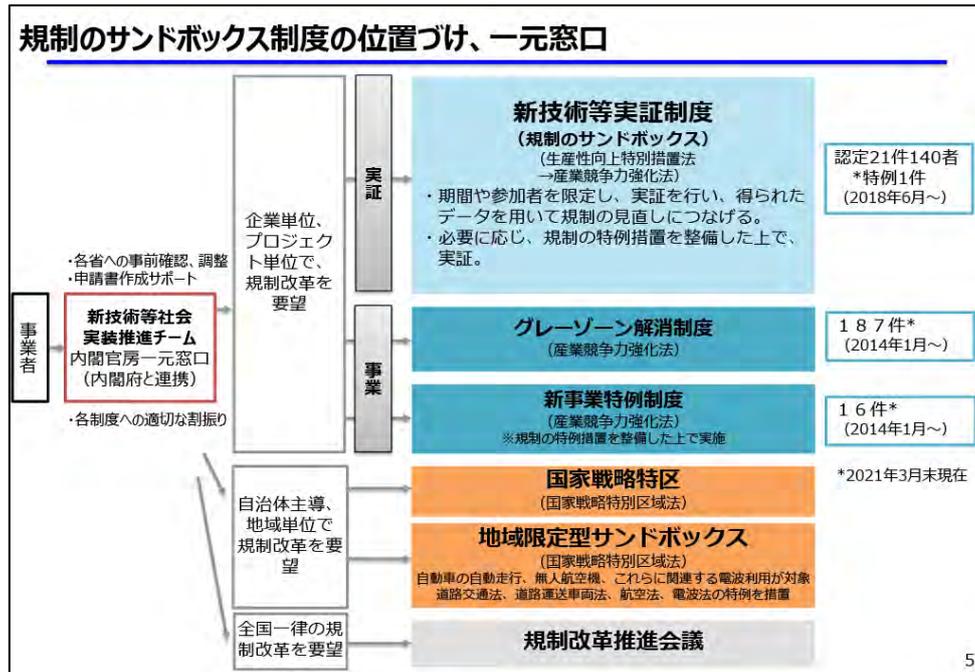
フォーム等を活用した課金について検討すべきである。その際、適切な課金システムの設計といった課題があるが、すでに運用されているデータプラットフォーム等を参考にしながら検討する必要がある。

<参考資料>

- 得られたデータがそのまま価値となる場合、データを取得する取組に留まらず、データ利活用基盤(WAGRI や NBDC 等が該当する)との接続を通じ、計画終了後のオープンデータ化⁵¹を含めデータ取引市場での流通を検討する(図表 19)。
- データを処理・分析することにより得られた知見を活用して取組の中で市場を獲得する場合、それが食品であれば、機能性表示食品制度や特定保健用食品制度といった制度や、民間認証制度の活用を検討する。もし既存の規制がその妨げになるのであれば、国家戦略特区や規制のサンドボックス制度の活用も検討する(図表 20)。

51 制限公開による公開を含む。

図表 19 データ利活用基盤の代表例：
農業データ連携基盤(WAGRI)の構造とメリット⁵²



52 <https://www.maff.go.jp/j/kanbo/smart/attach/pdf/index-43.pdf> より引用

サンドボックス制度活用の流れ（簡略版）

- まずは、内閣官房の一元窓口に相談（リモートも可）。
- 実証計画の内容を（人数、金額、場所、内容等）を工夫し、まずは、既存の規制の適用を受けることなく実証を実施できる環境をつくる。
必要があれば、規制の特例措置を求めることも可能。
- 実証計画を主務大臣（規制所管省庁、事業所管省庁）へ申請。
内閣官房の一元窓口が、一貫してサポート。
- 主務大臣は、実証計画が、既存の規制法令に違反しない場合には認定。主務大臣の見解（認定の可否、しない場合の理由、等）は、内閣府に設置した新技術等効果評価委員会でも審議。
- 実証後、規制所管省庁は、検討結果に基づき、必要な規制の撤廃又は緩和のための法制上の措置その他の措置を講じる。

図表 20 新規事業設立に向けた実証等を進める上で活用し得る施策：
サンドボックス制度の概要⁵³

申請書の記載事項

1. 新技術等実証の目標
2. 新技術等実証の内容
 - (1)新技術等の内容、新技術等を用いて実施しようとする事業活動の内容
 - (2)新技術等の実用化の可能性について行う実証の内容及びその実施方法
 - (3)新技術等に関する規制についての分析の内容及びその実施方法
3. 新技術等実証の実施期間及び実施場所
4. 参加者等の具体的な範囲及び同意の取得方法
5. 新技術等実証の実施に必要な資金の額及びその調達方法
6. 規制について規定する法律及び法律に基づく命令の規定
7. 規制の特例措置の内容（適用を受けようとする場合）
8. その他新技術等実証の実施に関し必要な事項

図表 20 新規事業設立に向けた実証等を進める上で活用し得る施策：

⁵³ <https://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/underlyinglaw/sandboximage.pdf> より引用

3.6. プレーヤ間の利害関係調整機能の整備

<現状と課題>

社会実装には特に企業の果たすべき役割が大きいにもかかわらず、データ連携・利活用を実現できていない企業は、自社保有ではないデータを取り扱った経験に乏しい。その結果、企業が必要なノウハウや能力を有していないことがデータ連携・利活用の促進を阻害する要因となっている場合がある。以下にその例を示す。

- 多数の機関が参加する研究事業における、大規模なデータを用いた仮説構築を目的とする研究計画書の立案経験が不足している。
- 複数の機関でデータを取得し、複数の企業がそれらを活用するという経験が不足しており、効率的に調整することが困難である。
- 研究計画書作成、データ利用申請、成果の知財化といった取組内の手続において参加機関の秘密情報が他機関へ共有される可能性が一定程度存在するものの、経験不足から情報のオープン・クローズを適切に判断することが困難である。

<対応方策>

54 <https://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/underlyinglaw/sandboximage.pdf> より引用

企業の研究デザインや要望に対応でき、統計学・データサイエンス・法務・知財・研究領域の知見を有するチームを組織し、データを取得する機関と利用する機関の間での調整を取りまとめる。また、研究代表者による参加機関間の利害関係の調整をサポートする。

<留意点>

オープン・クローズだけでなく、権利者が開示を了承する一定の関係者にのみ開示する考え方も導入すべきである。

また、研究機関においても、複数の研究機関に分散して保有されているバイオデータを連携させた研究基盤整備を行うようなことも考えられ、この場合にも利害調整が求められることになる。利害関係の調整は必ずしも企業間にものみ限定される観点ではないことに留意する必要がある。

3.7. セキュリティやトラストの確保

<現状と課題>

- データ利活用・連携を進めるに当たり、価値を最終消費者に届けるバリューチェーンの各段階において、利用機関や最終消費者が安心してデータや成果物を利用できるようにする必要がある。これは、2019年1月開催のダボス会議にて、安倍首相(当時)から発信された「データ・フリーフロー・ウィズ・トラスト(DFFT)」の概念に整理されている^{55, 56}。
- DFFTの推進は、各国の社会的・政治的・文化的背景と密接に関わるものであり、我が国でも包括的データ戦略やサイバーセキュリティ戦略⁵⁷等を踏まえつつ、今後更に検討が進むと想定されるバイオ分野においても、上記観点から、トラストの確保とデータの利活用のバランスをとっていく必要がある。

<対応方策>

55 安倍総理大臣による世界経済フォーラム年次総会演説 「希望が生み出す経済」の新しい時代に向かって(仮訳) https://www.mofa.go.jp/mofaj/ecm/ec/page4_004675.html

56 <https://www.mofa.go.jp/mofaj/files/100167362.pdf>

57 <https://www.nisc.go.jp/active/kihon/pdf/cs-senryaku2018.pdf>

- 正当な公共政策目的による例外に該当するかの検討も含め、DFFT の概念を
実現する検討を行う。
- プライバシー、知財、安全保障といった前提を踏まえ、取組内で活用する IT
について講ずべきセキュリティ対策を検討する⁵⁸。
- 必要に応じ、ゼロトラスト⁵⁹に則り、攻撃されることを前提としたセキュ
リティ体制の構築について検討する。

<留意点>

- 例えば、AMED「利活用を目的とした日本疾患バイオバンクの運営・管理」
では、品質・情報セキュリティの向上を目指す取組の一環として、国際標
準化機構 (ISO) の定める二つの国際規格 (ISO9001 : 2015 (品質)、
ISO/IEC27001 : 2013 (情報セキュリティ)) マネジメントシステムの認証をバ
イオバンク・ジャパンにて取得している。また、ISO20387 として制定され
たバイオバンキングの一般要求事項への対応を含め、関係する議論を踏ま
え、バイオ分野においても国際連携を視野に入れた実証を進めることが望
ましい⁶⁰。
- セキュリティの確保において必要な要件については米国国立標準研究所
(National Institute of Standards and Technology, NIST) が提案してい
る Cybersecurity Framework⁶¹におけるコア要件 (図表 23) などを参考とし検
討を始めることが望ましい。なお、NIST のプライバシーフレームワークと
の相互に補完関係⁶²にも留意されたい。
- 上記のほか、インフラとしてシステムを構築する際は、ソフトウェアサブ

58 例えば米国では、あるシステムを構成する複数のソフトウェアを一覧化し SBOM (Software Bill of Material) と呼ばれる構成表を作成した上で、それを基にソフトウェア単位に必要なセキュリティ対策を検討するといった手法が提案されている。

<https://www.mitre.org/sites/default/files/publications/pr-21-1440-horizon-strategy-framework-science-technology-policy.pdf> (P. 14)

59 「内部であっても信頼しない、外部も内部も区別なく疑ってかかる」という「性悪説」に基づいた考え方。利用者を疑い、デバイス (機器) を疑い、許されたアクセス権でも、なりすまし等の可能性が高い場合には動的にアクセス権を停止する。

https://cio.go.jp/sites/default/files/uploads/documents/dp2020_03.pdf より引用

60 <https://biobankjp.org/plan/iso.html>

61 <https://www.ipa.go.jp/files/000071204.pdf>

62 NIST プライバシーフレームワーク

<https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.01162020.pdf>

さらに、2023年6月にゲノムデータに特化した補完文書として Cybersecurity Framework Profile for Genomic Data のドラフト版が公表されており、こちらにも必要に応じて参照いただきたい

<https://csrc.nist.gov/publications/detail/nistir/8467/draft>

ライチェーンの関連標準などの関係ルール⁶³や欧州の仮名加工に関するベストプラクティスや技術的対応の動向⁶⁴なども理解しておく必要がある。例えば、PHR等の外部のシステムの連携を深化させる際に国際標準との差にも留意すべきである。

- 国際連携を踏まえ安全保障とバイオデータの管理の観点についても諸外国の考え方を理解しておく必要がある⁶⁵。

国際連携を見据え、分散するデータを外部連携する可能性がある場合は、今後クラウドを活用したインフラの整備も想定され、連携するインフラによっては異なる環境を連携させる必要も発生しうる。そこで、ハイブリッドクラウドなど異なる環境を管理する仕組み⁶⁶の議論の動向も理解しておくことが望ましい。

63 2022年9月14日米国大統領令 14028

<https://www.whitehouse.gov/wp-content/uploads/2022/09/M-22-18.pdf>

・The Secure Software Development Framework NIST 800-218

・<https://healthsectorcouncil.org/hic-scrim-v2/>

なども必要に応じて参照

また、バイオデータに限るものではないが、データ連携からさらに踏み込んでAI化を志向する際のリスクマネジメントの議論としては一例として以下を参照のこと。

<https://www.nist.gov/itl/ai-risk-management-framework>

64 欧州サイバーセキュリティ庁 (ENISA) が策定した「Pseudonymisation techniques and best practices」<https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices>

ENISA 「Deploying Pseudonymisation Techniques」

(<https://www.enisa.europa.eu/publications/deploying-pseudonymisation-techniques>)

ENISA 「Engineering Personal Data Sharing」

(<https://www.enisa.europa.eu/publications/engineering-personal-data-sharing>)

65 バイオエコノミーと安全保障についての各種取組の全体からのグランドデザインに関する背景的な考えについては米国大統領令「Executive Order on Advancing Biotechnology and Biomanufacturing Innovation for a Sustainable, Safe, and Secure American Bioeconomy」の記述が参考になる。また、データとの関係では、近い時期に出された2022年9月15日大統領令「Executive Order on Ensuring Robust Consideration of Evolving National Security Risks by the Committee on Foreign Investment in the United States」においてもデータに対する投資規制の観点を言及している。

66 <https://nvlpubs.nist.gov/nistpubs/specialpublications/nist.sp.800-190.pdf>

<https://csrc.nist.gov/publications/detail/sp/800-204/final>

<https://csrc.nist.gov/publications/detail/sp/800-204b/final>

<https://csrc.nist.gov/publications/detail/sp/800-204c/final>

DFFT: Data Free Flow with Trustとは

(※以下「デジタル時代の新たなIT政策大綱」(IT総合戦略本部, 2019年6月7日決定)より抜粋)

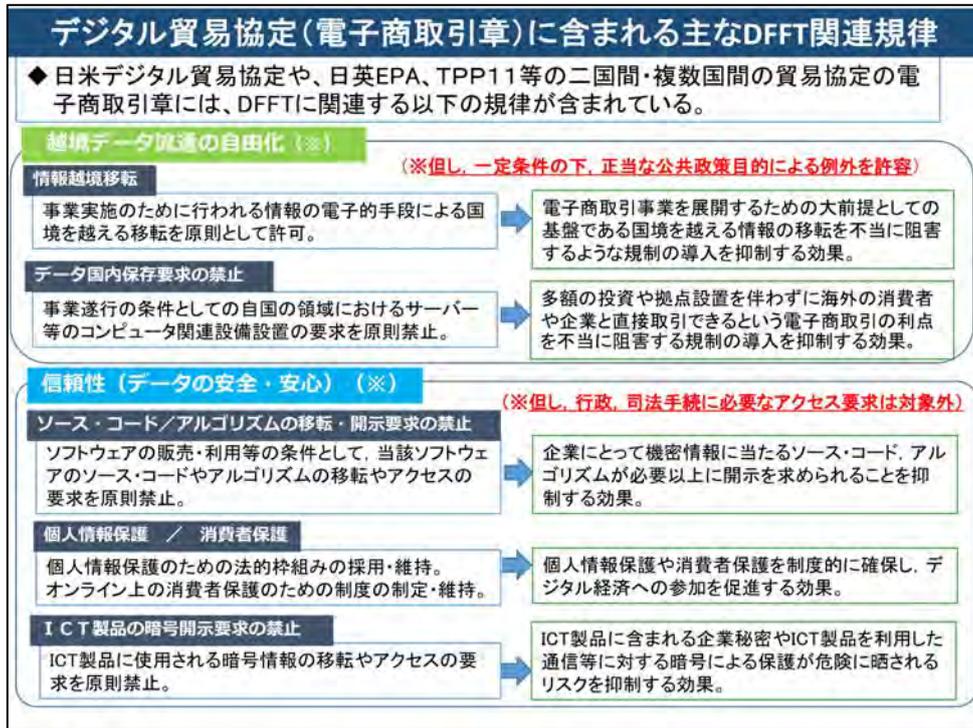
- 「**プライバシーやセキュリティ・知的財産権に関する信頼を確保しながら**, ビジネスや社会課題の解決に有益なデータが国境を意識することなく自由に行き来する, **国際的に自由なデータ流通の促進**を目指す, というコンセプト」。(パラ41)
- 「**DFFT(信頼性のある自由なデータ流通)のコンセプトに基づく「国際データ流通網」を広げていくことを目的として, より多くの国との間で, デジタル貿易ルールの形成(※)等を促進することが求められる**」(パラ42)

(※)世界最先端デジタル国家創造宣言・官民データ活用推進基本計画(2019.6.14閣議決定): DFFTのコンセプトに基づく「国際データ流通網」を広げていくことを目的として, より多くの国との間で, デジタル貿易ルールの形成等を促進するため, 電子商取引の貿易関連の側面に関するWTO交渉について, 早期の交渉妥結に向け, 日本がWTOにおける議論の加速化に積極的に貢献する。(脚注66)

Data Free Flow with Trust (DFFT)

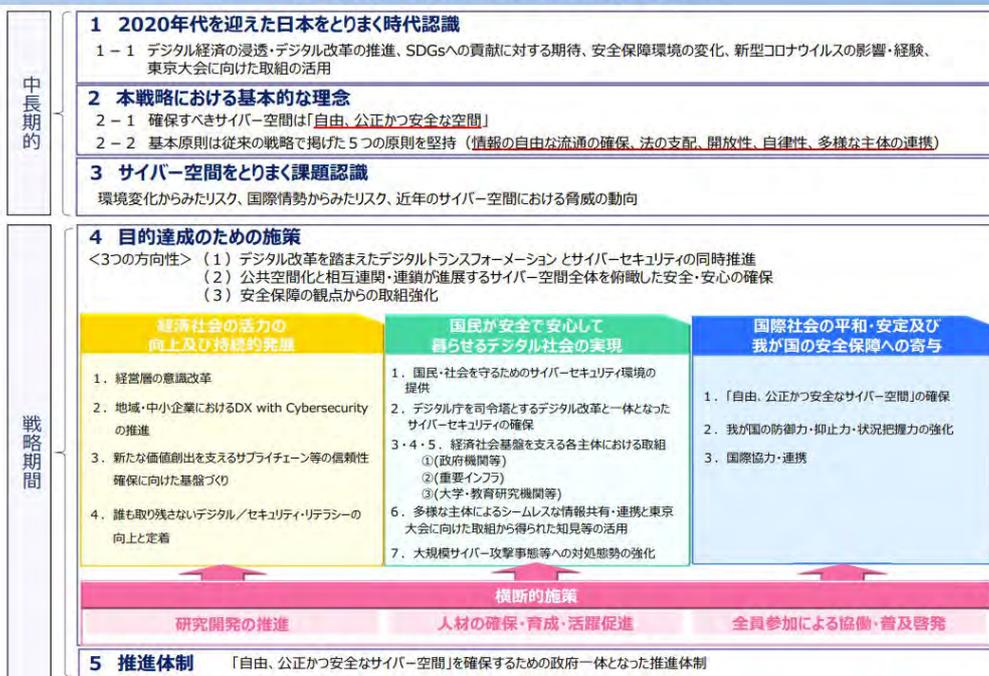
自由で開かれたデータ流通

データの安全・安心



図表 21 データ・フリーフロー・ウィズ・トラスト(DFFT)について 67

サイバーセキュリティ戦略の構成



67 <https://www.mofa.go.jp/mofaj/files/100167362.pdf> より引用

「Cybersecurity for All」を踏まえた対応の強化



10

図表 22 サイバーセキュリティ戦略⁶⁸

68 <https://www.nisc.go.jp/pdf/policy/kihon-s/cs-senryaku2021-gaiyou.pdf> より引用

機能	カテゴリー
識別 (ID)	資産管理 (ID.AM) : 自組織が事業目的を達成することを可能にするデータ、人員、デバイス、システム、施設が、識別され、組織の目的と自組織のリスク戦略における相対的な重要性に応じて管理されている。
	ビジネス環境 (ID.BE) : 自組織のミッション、目標、利害関係者、活動が、理解され、優先順位付けが行われている。この情報は、サイバーセキュリティ上の役割、責任、リスクマネジメント上の意思決定を伝えるために使用されている。
	ガバナンス (ID.GV) : 自組織に対する規制、法律、リスク、環境、運用上の要求事項を、管理し、モニタリングするためのポリシー、手順、プロセスが理解されており、経営層にサイバーセキュリティリスクについて伝えている。
	リスクアセスメント (ID.RA) : 自組織は、(ミッション、機能、イメージ、評判を含む) 組織の業務、組織の資産、個人に対するサイバーセキュリティリスクを把握している。
	リスクマネジメント戦略 (ID.RM) : 自組織の優先順位、制約、リスク許容度、想定が、定められ、運用リスクに対する意思決定を支援するために利用されている。
	サプライチェーンリスクマネジメント (ID.SC) : 自組織の優先順位、制約、リスク許容度、想定が、定められ、サプライチェーンリスクマネジメントに関連するリスクに対する意思決定を支援するために利用されている。自組織は、サプライチェーンリスクを識別し、分析・評価し、管理するためのプロセスを定め、実装している。
防御 (PR)	アイデンティティ管理、認証/アクセス制御 (PR.AC) : 物理的・論理的資産および関連施設へのアクセスが、認可されたユーザ、プロセス、デバイスに限定されている。また、これらのアクセスは、認可された活動およびトランザクションに対する不正アクセスのリスクアセスメントと一致して、管理されている。
	意識向上およびトレーニング (PR.AT) : 自組織の人員およびパートナーは、関連するポリシー、手順、契約に基づいた、サイバーセキュリティに関する義務と責任を果たせるようにするために、サイバーセキュリティ意識向上教育とトレーニングが実施されている。
	データセキュリティ (PR.DS) : 情報と記録(データ)が、情報の機密性、完全性、可用性を保護するための自組織のリスク戦略に従って管理されている。
	情報を保護するためのプロセスおよび手順 (PR.IP) : (目的、範囲、役割、責任、経営コミットメント、組織間の調整について記した) セキュリティポリシー、プロセス、手順が、維持され、情報システムと資産の防御の管理に使用されている。
	保守 (PR.MA) : 産業用制御システムと情報システムのコンポーネントの保守と修理が、ポリシーと手順に従って実施されている。
	保護技術 (PR.PT) : 技術的なセキュリティソリューションが、関連するポリシー、手順、契約に基づいて、システムと資産のセキュリティとレジリエンスを確保するために管理されている。
検知 (DE)	異常とイベント (DE.AE) : 異常な活動は、検知されており、イベントがもたらす潜在的な影響が、把握されている。
	セキュリティの継続的なモニタリング (DE.CM) : 情報システムと資産は、サイバーセキュリティイベントを識別し、保護対策の有効性を検証するために、モニタリングされている。
	検知プロセス (DE.DP) : 検知プロセスおよび手順が、異常なイベントに確実に気付くために維持され、テストされている。
対応 (RS)	対応計画 (RS.RP) : 対応プロセスおよび手順が、検知したサイバーセキュリティインシデントに対応できるように実施され、維持されている。
	コミュニケーション (RS.CO) : 対応活動が、内外の利害関係者との間で調整されている(例: 法執行機関からの支援)。
	分析 (RS.AN) : 分析は、効果的な対応を確実にし、復旧活動を支援するために実施されている。
	低減 (RS.MI) : 活動は、イベントの拡大を防ぎ、その影響を緩和し、インシデントを解決するために実施されている。
復旧 (RC)	改善 (RS.IM) : 組織の対応活動は、現在と過去の検知/対応活動から学んだ教訓を取り入れることで改善されている。
	復旧計画 (RC.RP) : 復旧プロセスおよび手順は、サイバーセキュリティインシデントによる影響を受けたシステムや資産を復旧できるよう実行され、維持されている。
	改善 (RC.IM) : 復旧計画およびプロセスが、学んだ教訓を将来の活動に取り入れることで改善されている。
	コミュニケーション (RC.CO) : 復旧活動は、内外の関係者(例: コーディネーティングセンター、インターネットサービスプロバイダ、攻撃システムのオーナー、被害者、他組織のCSIRT、ベンダ)との間で調整されている。

図表 23 NIST Cybersecurity Framework におけるコア事項⁶⁹

69 <https://www.ipa.go.jp/files/000071205.xlsx> より引用

3.8. 研究対象者とのコミュニケーション

<現状と課題>

- 健康・医療寄りの領域において海外連携を行うためにオープンサイエンスを実現するためには、研究参加者の取組に関する十分な理解を担保するインフォームド・コンセント(IC)の質を向上させることや、研究参加者がいつでも同意撤回可能であることが重要である。
- 「人を対象とする生命科学・医学系研究に関する倫理指針」(令和3年文部科学省・厚生労働省・経済産業省告示第1号)⁷⁰において、研究者等が研究対象者等からインフォームド・コンセントを受ける際に、電磁的方法(デジタルデバイスやオンライン等)を用いることが可能である旨とともに、その際に留意すべき事項が明記された。
- 健康・医療データ利活用基盤協議会においても、AMED事業で得られたデータは国民への還元の観点から有用な目的において幅広く活用されるべきであるとして、AMED及びAMED事業実施機関において、データの第三者利用に関する研究参加者への説明と同意取得の取組を進めることとしている。
- また、AMED研究データ利活用に係るガイドライン1.1版においては、同意取得の方法のひとつとして、ダイナミック・コンセントについての記載がある。

<対応方策>

必要に応じ、以下の検討を行う。

- 電磁的方法を含む適切なインフォームド・コンセントを受ける手続き
- ダイナミック・コンセントの活用可能性⁷¹
- 企業を含めた多様な主体が幅広い目的でデータを利活用するための研究対象者の権利保護のあり方

<留意点>

70 <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/hokabunya/kenkyujigyou/i-kenkyu/index.html>

71 研究対象者の保護の観点から、丁寧な説明を行うためにデータ収集時に、当該データに基づき新たな研究テーマが発生した場合に対象者となるかを検索させていただけるよう理解を得たうえで、改めてPHRインフラ等を用いて、2次利用のご理解を得るような取扱いの検討を行うことも考えられる。また、権利者保護の観点から世界医師会台北宣言の言及事項を参考にすることも考えられる。

健康・医療データ利活用基盤協議会では、得られたデータが幅広く活用されるために、民間企業による単独利用も含めた第三者利用が可能となるよう、共通的な同意事項の内容を定める方向で検討を進めている。今後の議論を踏まえ、バイオ分野でも連携しつつ検討することが望ましい。その際は、研究倫理の観点、個人情報保護の観点及び世界医師会台北宣言⁷²等を含む国際的整合性の観点から問題が無い内容とする事に留意する。なお、当初想定していなかった目的に広く利活用する場合には、利活用に基づき何らの選別が行われ特定の個人に対する措置に関連して不利益が生じないように、個人情報保護法等の観点から、取組の中にガバナンス体制の構築を検討する必要がある。

72 <https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/>

<コラム：データ連携・利活用におけるマネジメント体制の構築>

データ連携・利活用を進める場合、組織内での知財管理や、データの取扱いに関する認識合わせが十分に出来ていない場合には、データの帰属先や公開性が危ぶまれる場合がある。そのため、データ連携・利活用には、データの取扱いに関する専門知やデータ連携・利活用に関する関係者間での認識合わせが求められる。

そこで、関係する専門家が密に協力した組織内連携の構築を前提とした専門知を踏まえた方針策定を進めていくことが重要である。研究組織を統括する研究代表者のリーダーシップの下、データ管理・運用計画や、データベース構築の実務、法務・知財管理といったデータ連携・利活用における特定の対応課題を専門とする組織が、データベース構築の段階から、相互に密な連携を取りつつ進めていく必要がある。例えば、研究代表者をサポートする組織として、「データの管理・運用」「データベース構築」「法務・知財管理」「社会実装サポート」といった専門性を要する機能を持った委員会を組織して、それぞれの委員会が相互に連携して、データ連携・利活用の初期段階である組織内で創出されたデータの連携・利活用の方針策定から、実際の実務までを一貫して遂行するといった体制作りを行うことも一案である⁷³。なお、上で例示した委員会の機能としては例えば以下のように整理することができる。

● データ管理・運用

データマネジメントを担当する委員会。組織間でのデータの様式が異なる場合があることや、(特に民間企業等においては)データ提供が行われない場合があることを踏まえ、各研究組織を超えて、研究開発プロジェクト内で創出されたデータを一元して管理する。研究代表者と連携しつつ、データ管理を担当できる立場にある者が担当することが望ましい。またデータの品質に関しては利活用のニーズとの関係にも配慮しながら求められる品質に関して議論を行うべきで、科学的見地からの専門家の助言又は諮問機関の設置なども検討の余地がある。具体的な役割としては、研究計画当初から各研究組織において取得されるデータの構成や概要を把握・管理した上で、各研究組織に適合したデータマネジメントプランの様式を検討し、個々の研究者へ依頼を行いつつ、作成をサポートする。なお、この段階から、将来の利活用や実装における

73 例えば、SIP 第2期 バイオ・農業における一部の研究開発コンソーシアムでは、創出されたデータの連携・利活用を行うことをトップダウンのメッセージとしてプロジェクト開始当初から共有しており、研究代表者のリードの下、データ連携・利活用を進めるためのマネジメント体制を組んだ結果、データ連携・利活用の好事例を生み出すことができたといった事例もある。

課題の議論も並行して行うことが望ましく倫理や法律の専門家と連携しながら議論する必要がある。

- データベース構築

データベースの構築を担当する委員会。データ連携・利活用を進めるに当たり連携・利活用を可能とするようなシステム上での環境整備が必要であり、データベースの構築を専門して担う。データ管理やシステムに専門的な知見を有する者が担当することが望ましい。具体的な役割としては、データマネジメントプランを踏まえた研究組織におけるデータベースのプロトコル作成や構築を担当する。また、各データの公開レベルに沿って、オープンサイエンスの観点からの連携を踏まえた共有/制限共有が可能となる仕組みを整備する。

- 法務・知財管理

研究開発プロジェクトとして有するデータに関する法務・知財の方針検討・実務を担当する委員会。「データ」活用に関して専門的な知見を有する弁理士・弁護士等が担当する。但し、単純に IT 分野の専門や個人情報保護に関する法制の専門ではたりず、研究実務の課題・内容に関しても理解をしようする人材の育成が必要である。具体的な役割としては、各研究組織の管理するデータの権利・帰属関係を管理し、データ連携・利活用が行われる際には、法務手続き等を担当する。また、研究対象者の個人情報保護の観点からの利活用によって、不利益な影響を受ける決定がなされる恐れがある場合の取り扱いには十分注意し、最低限の利活用を心がける必要がある。また、参加機関間での利害関係の調整やトラストの確保、被試験者とのインフォームド・コンセント等の諸所のリスク対応についても担当する。

- 社会実装サポート

研究開発プロジェクトにおいて創出されたデータを活用し社会実装へと繋げることをサポートする委員会。データ利活用の専門家として実際に利活用するにあたっての方法をアドバイスできる組織や、外部的な視点から連携・利活用のアイデア創出を行うことができる諸外国の先行する事例などの情報を把握する専門家などが想定される。具体的な役割としては、例えば、創出されたデータの利活用に関するアドバイザリー・コンサルティングサービスの提供や、データの連携・利活用のアイデアソンなどを行い、シーズとニーズの仲介としてのサポートを実施するなど。なお、上記の組織を機能させるには、プログラムディレクターや研究代表者によるリーダーシップやデータ・連携・利活用の重要性に関するメッセージ出し、各組織の役職への十分なインセンテ

イブ(手当等)が不可欠である⁷⁴。

74 社会実装のためには、バイオ分野の特性を踏まえたサイバーセキュリティ確保に関する論点も重要である。当該論点については、例えば、バイオエコノミー特有のセキュリティ上の脅威への対処を目的として設立された国際的な非営利団体である BIO-ISAC の取組などを参照いただきたい。
<https://www.isac.bio/>

(別添) バイオ戦略におけるデータ関連取組一覧

<p>●SIP(第2期) スマートバイオ産業・農業基盤技術</p>
<p>[URL] https://www.naro.go.jp/laboratory/brain/sip/sip2/index.html</p>
<p>・国立研究機関等が提供するオープンデータベースを連携させた統合データベースシステムの構築を進めた。</p>
<p>・生物資源の情報を統合検索するウェブサイト(DBRPstanza)を開設し、NITEと理化学研究所が保有する微生物株情報のデータを登録、公開。</p>
<p>[URL] https://www.nite.go.jp/nbrc/dbrp/dbrpstanza/top.html</p>
<p>・微生物等の生物資源とその関連情報へワンストップでアクセスできる生物資源データプラットフォーム(DBRP)を公開。企業2社、自治体2県の保有微生物情報を登録、公開するとともに、経済産業省/NEDOの研究開発プロジェクトから得られたスマートセルに関するデータを登録。</p>
<p>[URL] https://www.nite.go.jp/nbrc/genome/db/dbrp.html</p>
<p>・民間企業等のスマート育種による品種開発を支援する「データ駆動型育種プラットフォーム」構築に向けた技術開発及び7作物による実証を実施中。</p>
<p>[URL]</p> <p>https://www.naro.go.jp/laboratory/brain/sip/sip2/theme/theme2102.html</p>
<p>・農水産物の生産・加工・流通・販売・消費の各段階を連携させるハブとなる情報共有システムである「スマートフードチェーン」を開発。</p>
<p>[URL] https://wagri.naro.go.jp/</p> <p>https://www.naro.go.jp/laboratory/brain/sip/sip2/about/sfs.html</p>
<p>・作物、微生物、土壌に関するデータを網羅的に収集、解析することにより、化学農薬・肥料の低減を可能とする植物-微生物共生を活用した営農法等の評価・開発を目的とし、マルチオミクス解析により農業生態系における作物-微生物-土壌の複雑な関係性の可視化を実施。</p>
<p>[URL]</p> <p>https://www.naro.go.jp/laboratory/brain/sip/sip2/result/20200623.html</p>
<p>・「食・マイクロバイオーム・健康情報統合データベース」の構築に向けて、調査研究を実施し、健常者の食・マイクロバイオーム・健康情報を集積中。</p>
<p>[URL]</p> <p>https://www.naro.go.jp/laboratory/brain/sip/sip2/theme/theme2104.html</p>

●PRISM 次世代栽培システムを用いたスマート育種技術開発の加速

[URL] <https://www8.cao.go.jp/cstp/gaiyo/sip/200312/sanko2-1.pdf>

- ・多様な環境条件の再現を可能にすることで品種開発を促進する人工気象器「栽培環境シミュレーター」を開発。
- ・複数の育種拠点を WAGRI を介して接続する「育種バーチャルラボ」の試行的運用(複数拠点からのデータ蓄積)を開始。

●スマート農業技術の開発・実証プロジェクト

[URL] https://www.affrc.maff.go.jp/docs/smart_agri_pro/smart_agri_pro.htm

- ・全国 148 地区でスマート農業技術の現場実証を実施。
- ・スマート農業実証プロジェクトにおいて、スマート農機のシェアリング等の新たなサービスモデルの現場実証を実施。

●人工知能未来農業創造プロジェクト

うち、AI を活用した病害虫早期診断技術の開発

[URL]

<https://www.affrc.maff.go.jp/docs/project/pdf/jisseki/2017/seika2017-22.pdf>

- ・AI を活用した土壌病害診断技術の開発、及び、診断結果・対策情報等を提供するシステムを構築中。AI を実装したアプリの開発を実施。

●スマート農業総合推進対策事業のうちデータ駆動型土づくり推進

[URL]

<https://www.affrc.maff.go.jp/docs/project/pdf/jisseki/2017/seika2017-22.pdf>

- ・スマート農業総合推進対策事業のうちデータ駆動型土づくり推進により土壌診断データベースを構築中(2022 年度まで)。

●林業イノベーションの推進

[URL]

https://www.jafta.or.jp/contents/jigyو_consulting/20_list_detail.htmlhttps://www.jafta.or.jp/contents/jigyو_consulting/20_list_detail.html

- ・森林・林業・木材産業の関係者におけるデータの共有・利活用を促進するため、森林クラウドと整合した ICT 生産管理システムの標準仕様を策定。

●マイクロバイオーム測定の精度向上のための「微生物カクテル」の提供

[URL] <https://www.nite.go.jp/nbrc/information/release/20190521.html>

- ・マイクロバイオーム分析技術を評価するためのレファレンスとなる「微生物カクテル」（複数の微生物を一定割合で混合したもの）を開発し、提供。試薬・装置等の製品開発や標準手順書のバリデーション、新規作業者のトレーニングなどにも広く活用。さらに、日本人の常在菌による微生物カクテルを開発し、ヒトマイクロバイオーム解析標準プロトコル作成に貢献。

●ゲノム・データ基盤プロジェクト

[URL] <https://www.amed.go.jp/program/list/index04.html>

- ・東北メディカル・メガバンク計画にて15万人規模のゲノム情報を含む大規模健常人コホートバイオバンクを構築。また、官民マッチングファンドによる10万人全ゲノム解析の実現に向けた取組を開始。

[URL] <https://www.megabank.tohoku.ac.jp/>

- ・ゲノム研究バイオバンク事業(バイオバンク・ジャパン)として疾患27万人規模の疾患バイオバンクを構築。

[URL] <https://www.amed.go.jp/program/list/14/01/003.html>

- ・ナショナルセンター・バイオバンクネットワーク(NCBN)として各ナショナルセンターの専門性を活かした生体試料等の収集に継続的に取り組む。

[URL] <https://ncbiobank.org/>

- ・がんや難病に関するゲノム医療の推進に向け、全ゲノム解析等実行計画およびロードマップ2021に基づき、全ゲノム解析を実施。

[URL] <https://www.mhlw.go.jp/content/10601000/000579051.pdf>

<https://www.mhlw.go.jp/content/10601000/000791111.pdf>

●バイオバンク横断検索システム

[URL] <https://biobank-search.megabank.tohoku.ac.jp/v2/>

- ・国立研究開発法人日本医療研究開発機構(AMED)による支援を受けて主要なバイオバンクの保有する試料・情報を一括で検索するバイオバンク横断検索システムを整備。

●産学官共同臨床情報利活用創薬プロジェクト(GAPFREE)

[URL] https://www.amed.go.jp/news/program/20200617_GAPFREE.html

- ・NCBN と日本製薬工業協会加盟企業により、疾患別情報統合データベース (NCBN が保有している高品質な臨床情報が付随した試料に対し、多層的オミックス解析を実施し、解析結果に検査値や画像データ等の臨床情報を突合せさせる仕組み) の構築及び利活用に係る産学官共同研究を開始。

● SIP(第2期)AI ホスピタルによる高度診断・治療システム

[URL] <https://www.nibiohn.go.jp/sip/>

- ・2022年度までに、秘密分散・秘密計算技術による多施設間分析や種々の統計解析等を通じ、医療情報データベースのシステム構築に係る検証を実施し、医療用語集の構築についても、ICD-10 への対応により一層の標準化、多言語対応に向けた改善を図りつつ、当該用語集を活用し、用語間の相関関係を統計処理することで、診断の補助に資するシステムの開発に取り組んだ。また、音声を用いた診療記録等の入力支援について、システム改修やその他課題の抽出及び対応により、入力精度向上を着実に進めた。

● 子どもの健康と環境に関する全国調査(エコチル調査)

[URL] <https://www.env.go.jp/chemi/ceh/>

- ・大規模かつ長期の出生コホート調査であるエコチル調査については、2020年度も約95%の参加者率を維持し、予定していた質問票調査、学童期検査、詳細調査を着実に実施。国立環境研究所においてデータ共有システムを構築中。