

研究データ基盤整備と国際展開

ワーキング・グループ

第2フェーズ報告書

図表集

令和3年3月

研究データ基盤整備と国際展開ワーキング・グループ

1 . 本報告書で紹介した事例

本報告書で紹介した事例一覧

事例名称等	内容	URL
IUGONET	分散管理されている太陽・地球物理に関する研究データの異分野間での流通を進め、新たな課題研究を推進することを目的として、メタデータのデータベースをネットワーク上で共有し、異分野の研究者でも容易に各研究データの概要を知ることができ、かつ、それら研究データへアクセスできるシステムを構築。	http://search.iugonet.org/list.jsp
COCN「エネルギー革新に向けたMI基盤の構築」	機能性高分子材料のデータ駆動型科学による開発加速という目標において、まず、拠点となる大学と企業が1対1で連携し、互いのデータベースによるデータ連携の有用性を確認した後、大学を核としたコンソーシアムを形成し、各社の協調領域を見極めた上で、大学のデータベースやインフラの共用化、また、新たなデータ取得のための材料開発を実施することを提言。	http://www.cocn.jp/report/5bd594907a344229192c6132546a52143c31bde8.pdf
NIMSを中核としたオープンプラットフォーム (MOP : Materials Open Platform)	化学メーカーの保有する高分子材料について、NIMSがマテリアルズ・インフォマティクスなどを用いて評価する。その結果、得られた基礎的な研究データは各化学メーカーへと共有され、新たな材料開発等へと展開される(化学MOP)。NIMSは鉄鋼メーカーとも、鉄鋼業界におけるオープンプラットフォーム(鉄鋼MOP)の取組を実施している。	https://www.nims.go.jp/news/press/2017/06/201706190.html
JOINUS (Joint Open Innovation of drug repositioning)	アステラス製薬(株)、田辺三菱製薬(株)、第一三共(株)が作成したドラッグ・リポジショニング化合物ライブラリーを、公募によって選ばれた国内研究機関(大学、公的研究機関、企業等)の研究者が、独自の評価系で評価する。得られた研究データは各社へとフィードバックされ、これにより、ライブラリー化合物の新たな適応症が発掘され、新薬開発につながると期待される。	https://open-innovation-joinus.jp/joinus/
武田薬品工業(株)と東北大学東北メディカル・メガバンク (ToMMo)による共同研究	一般住民の全ゲノムリファレンスパネルの充実と日本人集団に特徴的な遺伝子型の発見、および全ゲノム情報と脳MRI画像を含む健康情報・医療情報の統合的解析による新薬・治療法の研究開発を目的とした共同研究。武田薬品工業は、資金提供を行う代わりに、同共同研究を通じて取得される全ゲノム情報への1年間の優先的アクセス権を獲得。その後、武田薬品工業とToMMoは、我が国のゲノム情報の医療分野での活用を推進するため、全ゲノム解析とこの解析結果の利活用を目的としたコンソーシアムの実現を目指している。	https://www.tohoku.ac.jp/japanese/2020/04/press20200414-03-tommo.html
リジェネロン社が主導するバイオ医薬品会社のコンソーシアム	米国のバイオテクノロジー企業であるリジェネロン社が主導するバイオ医薬品会社のコンソーシアム。同コンソーシアムでは、参画企業が優先的に研究データにアクセスできる期間が設定されている。	https://www.ukbiobank.ac.uk/2019/03/new-data-available-exome-sequence-data-on-50000-participants/

本報告書で紹介した事例一覧

事例名称等	内容	URL
Clinical Trial Data Sharing (CTDS)	臨床試験データは高い透明性が求められるべきという考えと、臨床試験データは公共の財産であり、公衆衛生の向上をもたらす研究に積極的に活用すべきとの考えに基づいて、主として欧米の製薬企業により、臨床試験で得られた被験者レベルの詳細な個別被験者データ (Individual Participant/Patient Data : IPD) の公開・共有が行われるようになった。我が国でも、臨床試験データの共有 (Clinical Trial Data Sharing : CTDS) という言葉が広く知られるようになり、活発に議論が行われるようになってきた。	http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/pdf/ctds.pdf
AMEDによる「産学連携による次世代創薬AI開発」プロジェクト	産学が保有する創薬研究における多面的で膨大なデータを広く集約することで、産学で利用可能な化合物設計AIを開発することを目的としている。同プロジェクトでは、federated learning技術を活用することで秘匿のままでの研究データ提供を可能にしており、開発されたAIは、その開発に貢献した参加企業に提供することで企業内での研究開発に活用される予定。	https://www.amed.go.jp/program/list/11/02/001_02-04.html
MELLODDY (Machine Learning Ledger Orchestration for Drug Discovery)	federated learning技術を用いて、各製薬企業が自社の研究データを秘匿のまま提供して、それらのデータを基に、正確かつ効率的な創薬を実現するためのAI技術の開発を推進している。	https://www.melloddy.eu/
農業データ連携基盤 (WAGRI)	WAGRIを通じて提供された気象や農地、地図情報等のデータ・システムを基に、データ・システム利用者 (農機メーカー、ICTベンダー等) が、農業関連サービス (Webアプリケーション) を開発。農業者等はそれらのサービスを選択・活用し、データ駆動型農業を実施することが可能になる。	https://www.naro.affrc.go.jp/project/research_activities/laboratory/rcait/121064.html
FANTOM (Functional Annotation Of Mammalian genome) プロジェクト	哺乳類を対象として網羅的に遺伝子の機能を解析することを目的とした、理化学研究所主導の国際共同研究プロジェクト。同プロジェクトでは、研究成果および解析に用いたデータやその解析結果は、理化学研究所のFANTOMウェブサイト等を通じて、全世界に無制限で公開されている。	https://fantom.gsc.riken.jp/
COVID-19データポータル	COVID-19に関する研究データやサービスを共有する欧州を中心とした国際的な取り組み。日本も欧州の呼びかけに賛同し、国立情報学研究所と国立遺伝学研究所が中心となって、COVID-19データポータルJAPANを構築。	https://covid19dataportal.jp/

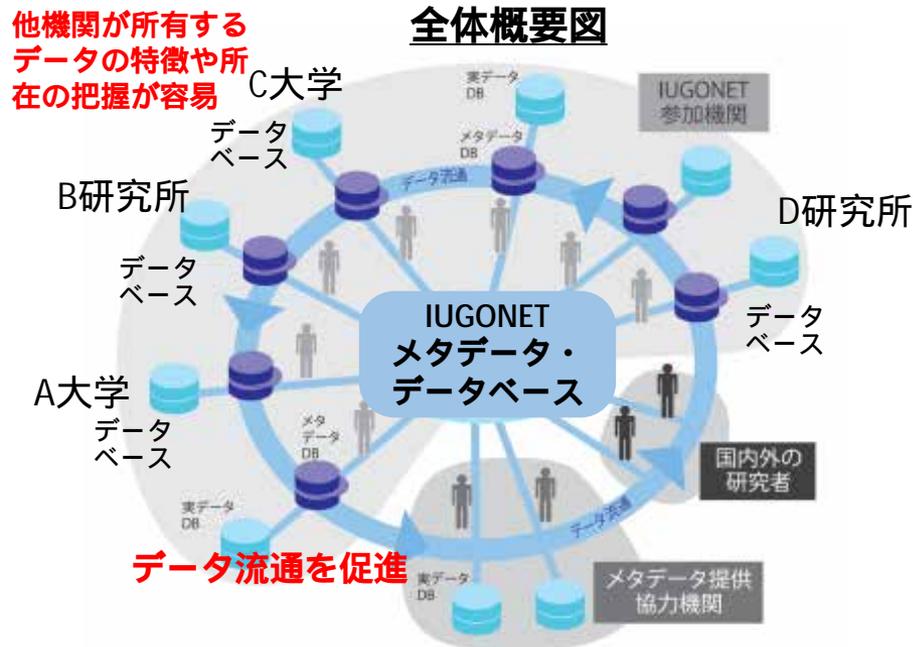
IUGONET

(情報・システム研究機構、東北大学、名古屋大学、京都大学、九州大学の連携事業)

目的：分散管理されている研究データの異分野間での流通を進め、新たな課題研究を推進

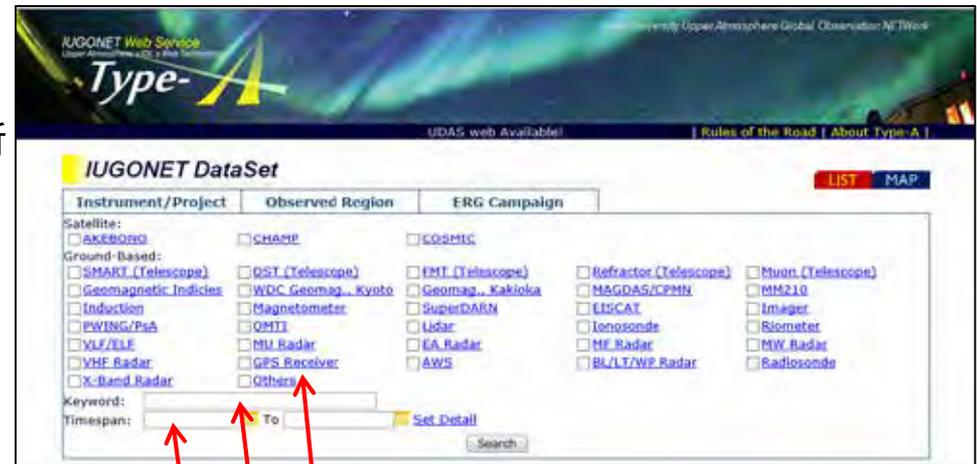
手法：メタデータのDBをネットワーク上で共有し、異分野の研究者でも容易に各研究データの概要を知ることができ、かつ、それら研究データのDBへアクセスできるシステムを構築

IUGONET (Inter-university Upper atmosphere Global Observation Network)



出典：IUGONETパンフレット (<http://www.iugonet.org/product/>)

IUGONETのUI



出典：「IUGONET Web Service Type-A」、IUGONET (<http://search.iugonet.org/list.jsp>)

データ種別検索

キーワード検索

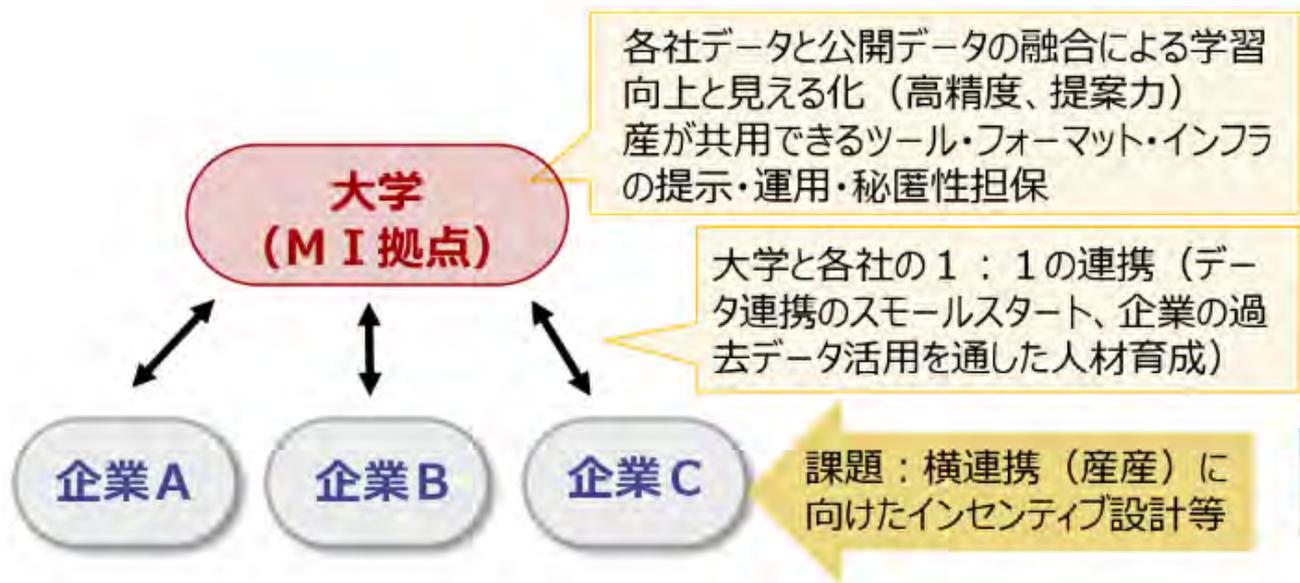
時間検索

メタデータ：データの説明、実データの所在 (URL)、利用規約、コンタクト先等。さらに、現象のサマリーや解析手順も掲載。

大学の人材・インフラを核としたMI拠点: コンソーシアム体制で構築

産学がunder one roofで**エネルギー機能材料のMI研究開発と人材育成を両輪で推進**

コンソーシアム体制（大学のデータベース、インフラを共用）



ハイスツールプット開発設備の導入



新研究開発センター-産学連携施設など

出典：「COCNフォーラム2019 エネルギー革新に向けたMI基盤の構築」（平成29年7月16日 早稲田大学 特任研究教授 西出 宏之）、一般社団法人産業競争力懇談会（COCN）、
<http://www.cocn.jp/forum/2019/>

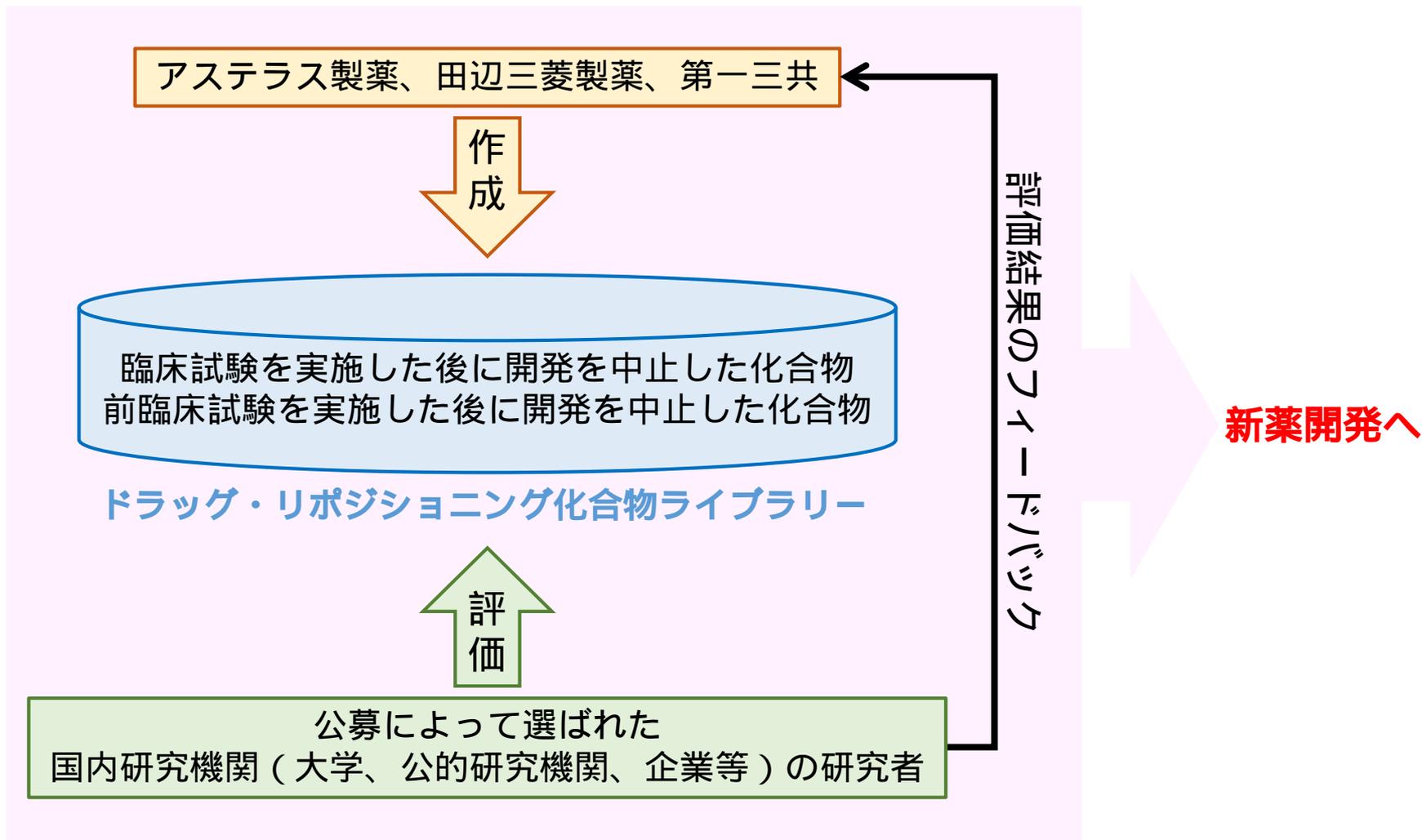
マテリアルズ・オープンプラットフォーム（MOP）

～基礎研究と民間ニーズの融合による未来を見据えた非連続な革新材料創出の場の構築～

- ・ NIMSにはこれまで国家により多大な投資がなされており、我が国の国際競争力確保に向け**国家資産を有効活用**すべき。
- ・ 中長期的かつ非連続な研究開発課題について、NIMS（特定国立研究開発法人）を起点として**産業界と大学を結ぶ業界別のオープンプラットフォームを形成**。
- ・ オープンプラットフォーム**成功の鍵**は、産官学総掛かりで将来の我が国産業力強化に資する各産業界の将来を見据えた**魅力的なテーマが設定できるか否か**。
⇒NIMSは化学業界の各企業との間で27回議論を行い、化学分野におけるMOPのテーマを設定。

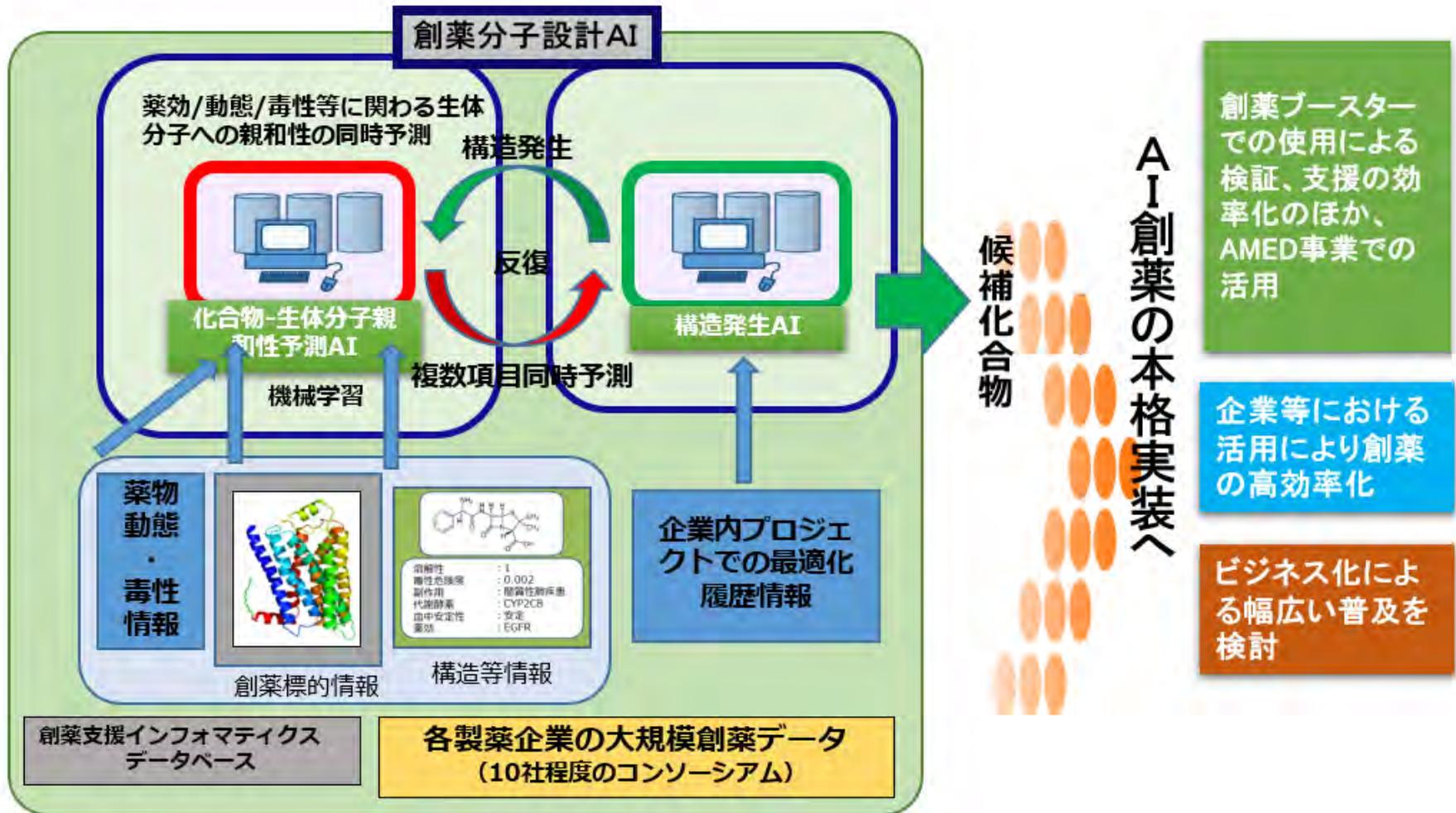


JOINUS (Joint Open Innovation of drug repositioning)



(内閣府作成)

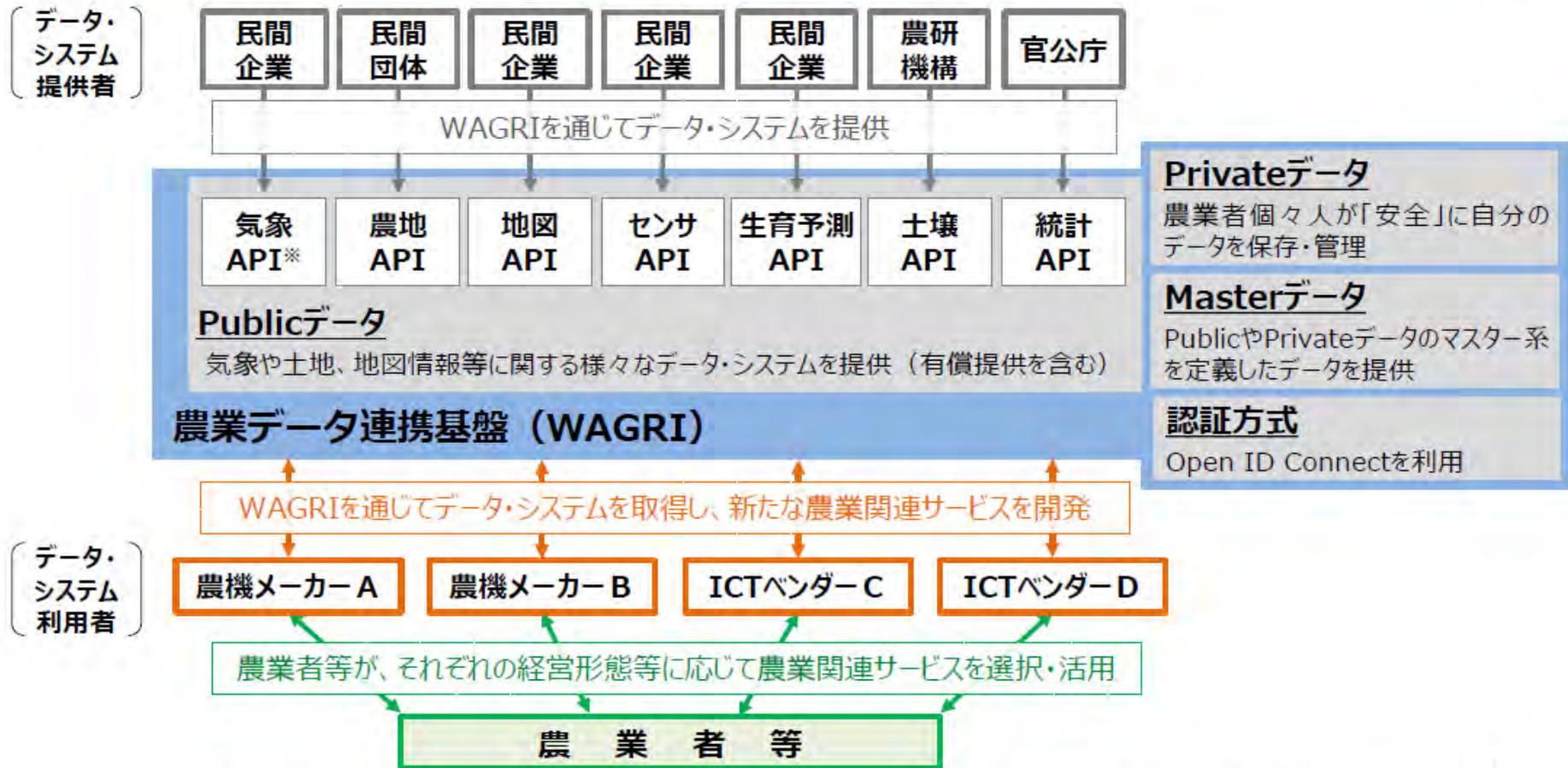
AMED「産学連携による次世代創薬AI開発」プロジェクト



出典：「産学連携による次世代創薬AI開発」、AMED
(https://www.amed.go.jp/program/list/11/02/001_02-04.html)

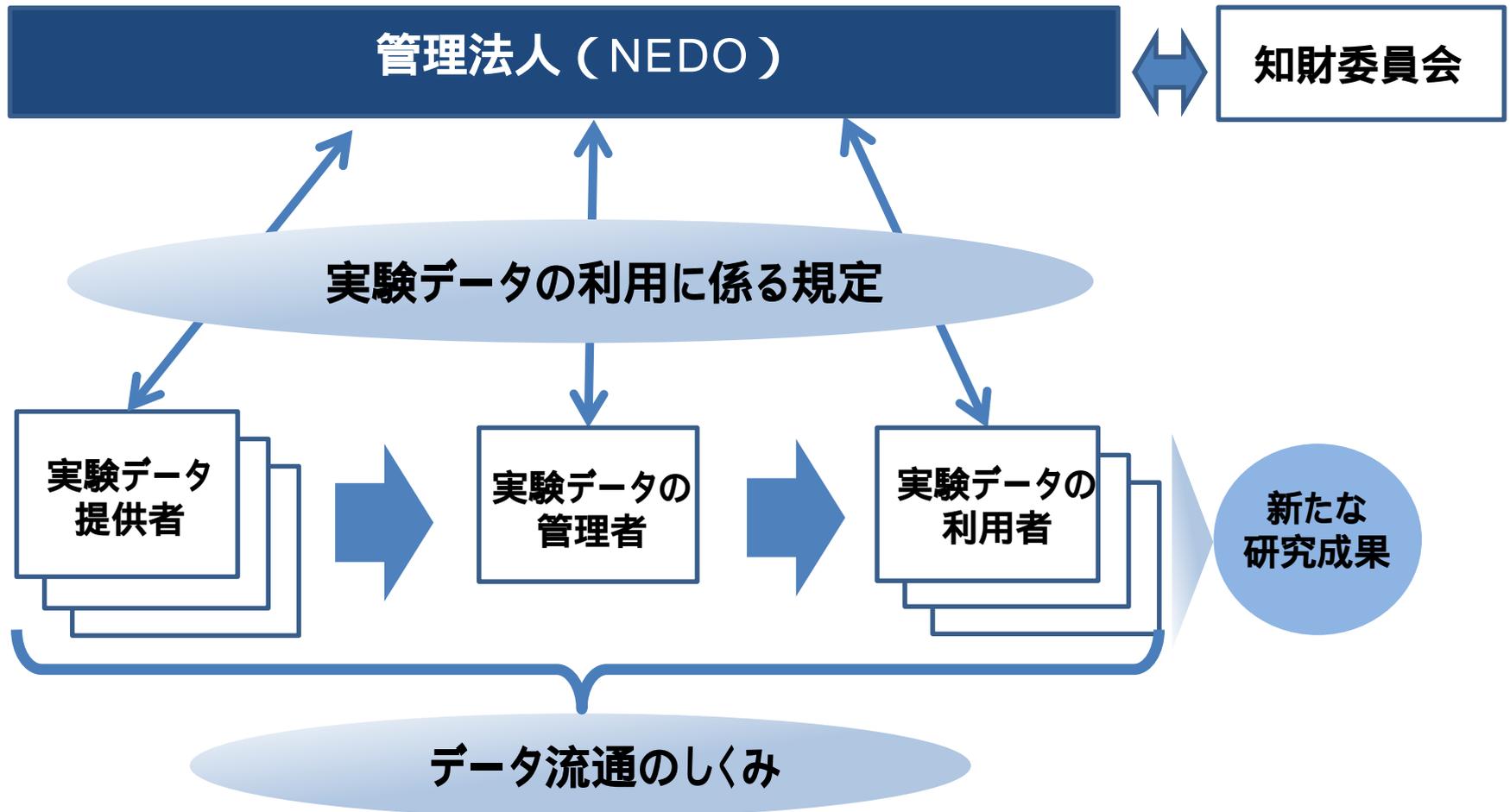
農業データ連携基盤（WAGRI）

- WAGRIは、農業ICTサービスを提供する民間企業の協調領域として構築。
- WAGRIを通じて気象や農地、地図情報等のデータ・システムの活用が可能となり、民間企業が行うサービスの充実や新たなサービスの創出を促すことで、農業者等が様々なサービスを選択・活用できるようになる。



※Application Programming Interface の略。複数のアプリケーション等を接続（連携）するために必要なプログラムを定めた規約のこと。

東京臨海部実証実験に係る実験データ活用のためのスキーム



出典：NEDO資料（SIP第2期自動運転(システムとサービスの拡張)、令和2年12月）

FANTOMプロジェクト

FANTOM = Functional ANnoTation Of Mammalian genome

理化学研究所の主導による国際共同研究プロジェクト

理化学研究所で配列決定したマウス完全長cDNAや独自技術であるCAGE法を用いた哺乳類のゲノム機能研究

FANTOM1 - マウス完全長cDNA(約2万個)の機能アノテーション

FANTOM2 - 約6万個のマウス完全長cDNAの機能アノテーション

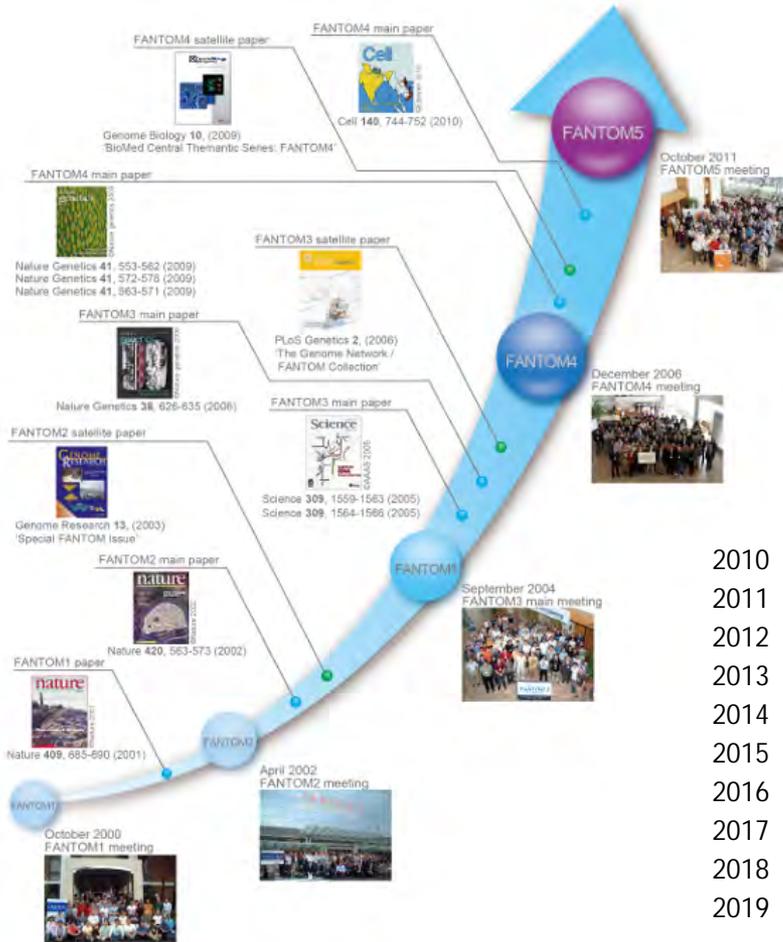
FANTOM3 - 哺乳類ゲノムの転写ランドスケープ

FANTOM4 - 発現制御ネットワークの理解

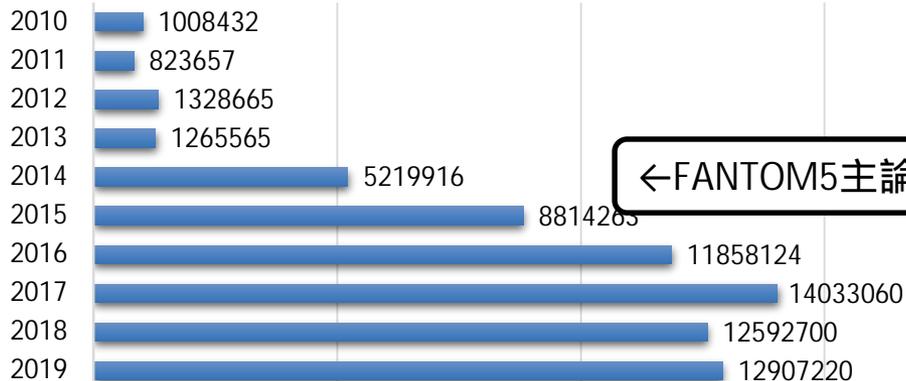
FANTOM5 - 哺乳類プロモーター・エンハンサー・長鎖ノンコーディングRNA(lncRNA)、マイクロRNA(miRNA)の発現地図

FANTOM6 - ノンコーディングRNAの機能解析

研究成果、解析に用いたデータやその解析結果はリソースとして全世界に無制限で公開(Creative Commons Attribution 4.0 International License)



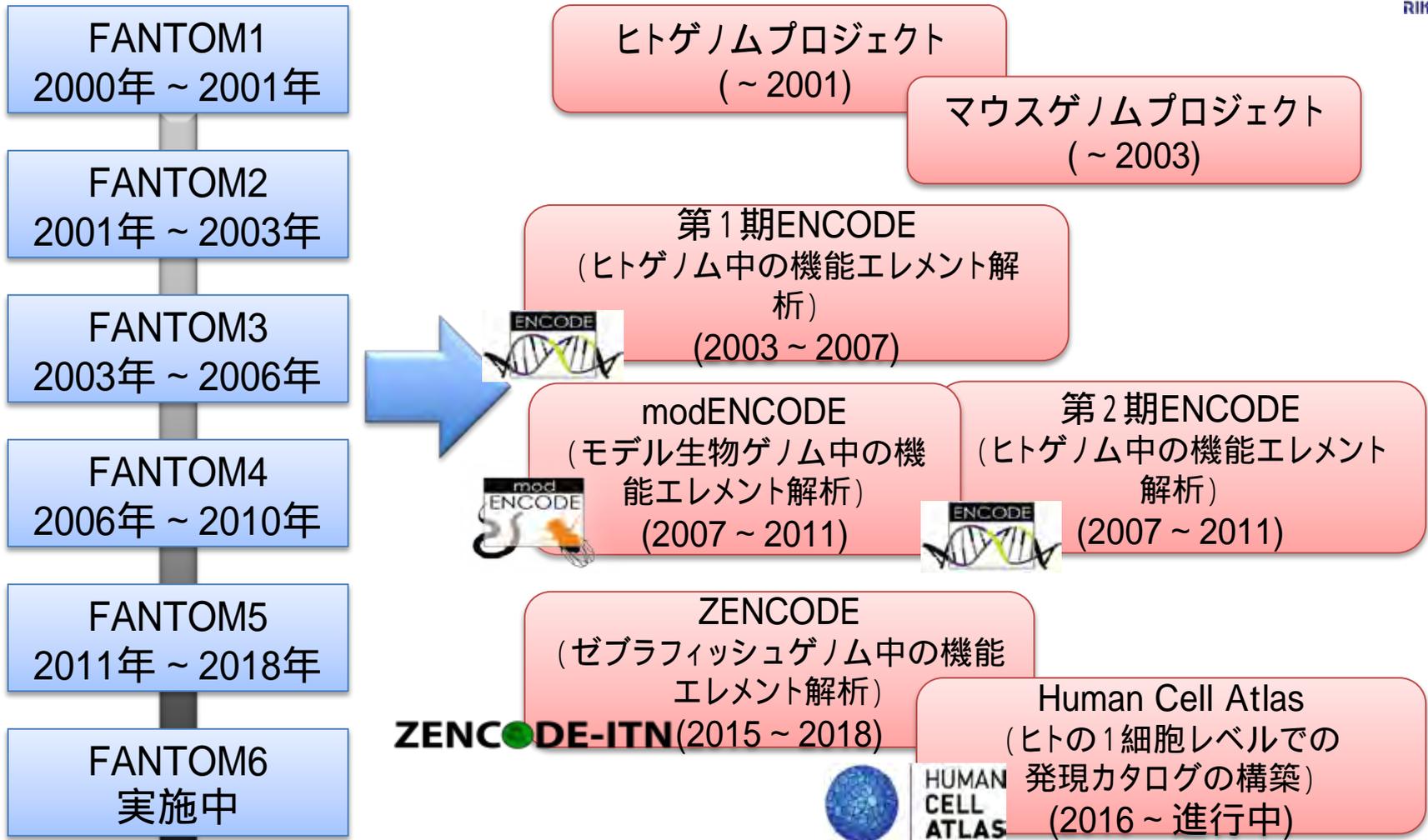
FANTOMウェブサイトのアクセス数



2019年は1~3月実績より推定

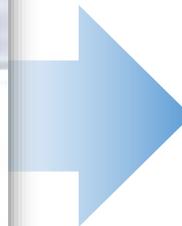
出典：理化学研究所

FANTOMプロジェクトの波及効果



様々な国際共同研究プロジェクトに招待、参画へ

COVID-19 データポータル JAPAN



トップページ <https://covid19dataportal.jp>

各詳細ページ

出典：国立情報学研究所オープンサイエンス基盤研究センター大波純一氏

2 . 各研究機関における研究データマネジメントの事例

内閣府「研究データ基盤整備と国際展開WG」
報告書図表集（令和2年10月公開）より再掲

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

理化学研究所の事例についてはアップデートされた内容を含む

各研究機関におけるデータマネジメントの実例

		JAMSTEC	NIMS	AIST		理研	
事業内容		津軽海峡東部海洋レーダーデータサイト	船舶観測データ・サンプルの計測・取得	材料データプラットフォーム	人体形状の計測	歩行データの計測	FANTOMプロジェクト*
データ公開状況	生データ	一部公開：レーダーデータは利用申請をすることでダウンロード可能。水温データはWebサイトでグラフを公開、ダウンロード不可。	公開：航海・潜航に関する品質管理を施したデータ	データごとに適切な公開条件を付して公開（利活用を図るために登録されるデータ）	公開：人間計測データ	公開：歩行データ 非公開：転倒歴など	公開：独自技術（Cap-Trap法、CAGE法）より得られた完全長cDNA配列やRNAの5'配列データ
	加工データ	公開：レーダーデータから算出した海表面の流向・流速データ	-	データごとに適切な公開条件を付して公開（研究成果物として公表されたもの）	非公開：解析・二次データ 有償公開：統計量	非公開：転倒スコアなど	公開：哺乳類電子の機能情報（遺伝子の位置や発現機構など）データファイルならびに専用ビューワーを介して提供
	メタデータ	公開	公開	公開	公開	公開	公開
メタデータの内容		<ul style="list-style-type: none"> データ取得日 データ取得方法 データ取得者 	<ul style="list-style-type: none"> データ取得日 データ位置情報 データ取得方法 データ処理方法 データ処理者 データ形式 キーワード 	<ul style="list-style-type: none"> データ作成日、公開日等 公開と利用の条件 データ作者情報 データ来歴、装置、試料等の情報 	<ul style="list-style-type: none"> データ取得日 データ取得方法・項目 実験対象者の性別・年齢 データ形式 	<ul style="list-style-type: none"> データ取得期間 データ取得方法・項目 実験対象者の性別・年齢・身長・体重 データ形式 	<ul style="list-style-type: none"> 使用サンプルの詳細情報 サンプルの分類情報 配列データの取得方法 データ形式
データ公開猶予期間		30分ごとの準リアルタイム	航海メタデータは原則2ヶ月 データは原則2年。ただし、定常観測データは処理が済み次第公開	データポリシおよびデータリポジトリ運用規則に沿ってデータ登録者が設定する期間	既に公開済み	既に公開済み	論文発表と同時公開（既に公開済み）
データ公開猶予理由		-	学会発表、論文作成、特許作成	学会発表、論文作成、特許作成など	-	-	論文作成のため

*) 理研主導による国際共同研究プロジェクト (FANTOM = Functional ANnotation Of Mammalian genome)。哺乳類を対象として、網羅的に遺伝子の機能 (ゲノム中の遺伝子の位置や発現機構) を解析。

各研究機関におけるデータマネジメントの実例

	JAMSTEC		NIMS	AIST		理研
第三者機関によるデータ活用 の有償・無償基準	<ul style="list-style-type: none"> 科学的・教育的目的利用は無償 産業利用は有償 	<ul style="list-style-type: none"> 科学的・教育的目的利用は無償 産業利用は有償 	原則無料。データポリンおよびデータリポジトリ運用規則に沿って有償等の可能性あり	無償	無償	無償
公開媒体	Webサイト	<ul style="list-style-type: none"> Webサイト（メタデータ・データファイル） オフライン提供（生データ、高解像度映像、サンプル） 	データリポジトリ	ホームページ	ホームページ	<ul style="list-style-type: none"> 専用Webサイト（理研内） 国際塩基配列DBコレクション 公開リポジトリサービス NBDC（JST内）の生命科学系データベースアーカイブ
データ帰属先	<ul style="list-style-type: none"> レーダーデータはJAMSTECに帰属。 水温データの一部はJAMSTECに帰属、一部は観測を実施している他機関に帰属。 	原則JAMSTECに帰属	NIMSに帰属。ただし共同研究契約等がある場合はその契約に依る。	原則AISTに帰属	原則AISTに帰属	理研に帰属するが、データはCC BY 4.0で公開