

# 研究のデジタルトランスフォーメーション (DX)

ポストコロナ社会におけるニューノーマルを研究活動においても実現するため、研究のデジタルトランスフォーメーション(DX)をソフト・ハードの両面から取り組む必要がある。文部科学省においては、ソフト面として**研究データを戦略的に収集・共有・活用**するための取組を強化すると同時に、ハード面では、実験の自動化や遠隔地からの研究インフラへのアクセスを可能にする**研究施設・設備のリモート化・スマート化**、更に**次世代情報インフラ**である高速通信ネットワークと高性能計算資源の**整備**を加速する。

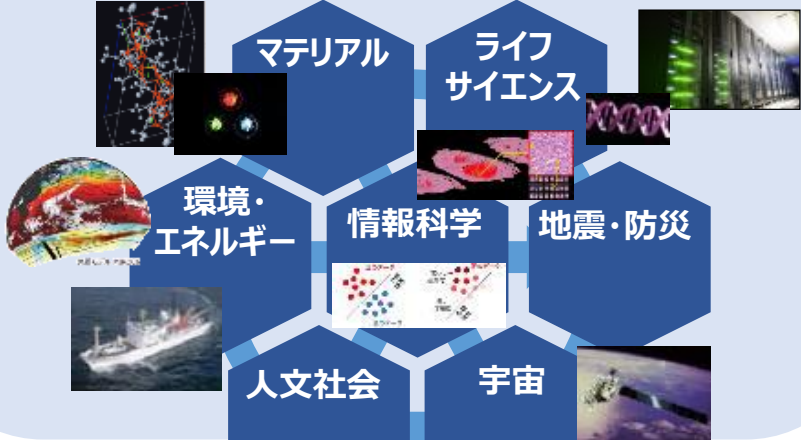
## 1. データ基盤の構築と AI・データ駆動型研究の推進

研究システムをデジタル転換するにあたって重要となるのは研究データである。

そのため、それぞれの分野の特性を生かしながら、**高品質な研究データの収集と、戦略性を持ったデータの共有のためのデータプラットフォームの構築、人材の育成・確保**に取り組み、更に、データを効果的に活用した、先導的な**AI・データ駆動型研究**を推進する。

- ▼ 関連施策
  - ・マテリアルDXプラットフォーム構想実現のための取組
  - ・創薬等ライフサイエンス研究支援基盤事業
  - ・気候変動対策や省エネ社会実現に向けた研究基盤技術の強化
  - ・AIP：人工知能/ビッグデータ/IoT/サイバーセキュリティ統合プロジェクト
  - ・統計エキスパート人材育成プロジェクト
  - ・研究大学強化促進事業

等



## 2. 研究施設・設備のリモート化・スマート化

大型共用施設から研究室まであらゆる研究現場において、リモート研究を可能にする環境構築や、実験の自動化を実現するスマートラボ等の取組を推進し、**時間や距離に縛られず研究を遂行できる革新的な研究環境を整備**する。



- ▼ 関連施策
  - ・大学・研究開発法人等の施設・設備の遠隔化、自動化
  - ・世界最高水準の大型研究施設におけるDXの推進
  - ・研究のDX推進のための共用体制整備 等



遠隔観察

## 3. 次世代情報インフラの整備

全国的な研究のDXを支える学術情報基盤としての役割を果たすため、**全国の大学・研究機関を超高速かつ大容量につなぐ学術情報ネットワーク「SINET」の機能を強化・拡充**する。



また、AI・データ駆動型研究を支えるため、スパコン「富岳」をはじめとした**高性能・大規模な計算資源の整備**と、それらを徹底活用した更なる成果創出を加速する。

# 第6期科学技術・イノベーション基本計画 (目指すべき社会像と情報分野の主な取組)

## 我が国が目指すべき社会(Society 5.0)

サイバー空間とフィジカル空間の高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する  
人間中心の社会

これにより、**持続可能性と強靱性を備え、国民の安全・安心を確保**

【持続可能性の確保】

- **地球環境の持続**
- 現世代と将来の世代が**豊かに生きていける**社会の実現

【強靱性の確保】

- 災害や感染症をはじめ、様々な脅威に対する**総合的な安全保障**の実現

**一人ひとりの多様な幸せ(well-being)を実現**

【経済的な豊かさと質的な豊かさの実現】

- 誰もが**能力を伸ばし、多様な働き方を可能**に
- 生涯にわたり**健康で社会参加**
- 夢を持ち続け、**自らの存在を肯定し活躍**

### 国民の安全と安心を確保する 持続可能で強靱な社会への変革

- (1) **サイバー空間とフィジカル空間の融合による新たな価値の創出**
  - ・ 政府のデジタル化、データ戦略(ベースレジストリ整備等)
  - ・ Beyond 5G、スパコン、量子技術、半導体等の次世代インフラ・技術の整備・開発
- (2) **地球規模課題の克服に向けた社会変革と非連続なイノベーションの推進**
  - ・ **カーボンニュートラル**に向けた研究開発
- (3) **レジリエントで安全・安心な社会の構築**
- (4) **価値共創型**の新たな産業を創出する基盤となるイノベーション・エコシステムの形成
- (5) **次世代に引き継ぐ基盤**となる都市と地域づくり(スマートシティの展開)
- (6) **様々な社会課題**を解決するための研究開発・社会実装の推進と総合知の活用
  - ・ **国家戦略(AI、バイオ、量子、マテリアル等)**による研究開発等

### 知のフロンティアを開拓し価値 創造の源泉となる**研究力の強化**

- (1) **多様で卓越した研究を生み出す環境の再構築**
  - ・ 人文・社会科学の振興と総合知の創出(ファンディング強化、**人文・社会科学のDX**)
- (2) **新たな研究システムの構築(オープンサイエンスとデータ駆動型研究等の推進)**
  - ・ 研究データの管理・利活用、スマートラボ・AI等を活用した研究の加速
  - ・ 研究施設・設備・機器の整備・共用、**研究DX**が開拓する新しい研究コミュニティ・環境の醸成

⇒ **研究DX**

- (3) **大学改革の促進と戦略的経営に向けた機能拡張**

### 一人ひとりの多様な幸せと 課題への挑戦を実現する **教育・人材育成**

探究力と学び続ける姿勢を強化する**教育・人材育成システム**への転換

- ・ **初等中等教育**段階からの**STEAM教育**や**GIGAスクール**構想の推進、**教師の負担軽減**
- ・ 大学等における多様なカリキュラムやプログラムの提供、**リカレント教育**を促進する環境・文化の醸成(MOOC等多様なデジタルコンテンツの活用、対面とオンラインのハイブリッド化等)

# 第6期 科学技術・イノベーション基本計画における 研究DXに係る文部科学省の主な取組

## あるべき姿

社会全体のデジタル化や世界的なオープンサイエンスの潮流を捉えた研究そのもののDXを通じて、より付加価値の高い研究成果を創出し、我が国が存在感を発揮することを目指す

## 主な取組

### <信頼性のある研究データの適切な管理・利活用促進のための環境整備>

- ・研究データ基盤システム（NII Research Data Cloud）の普及・広報と必要な改良・持続的運営
- ・研究データ基盤システムと分野毎のデータ連携基盤との間で、相互にデータの利活用を図るための仕組みの構築
- ・研究者の研究データ管理・利活用を促進するための人材確保
- ・研究データの管理・利活用に関する相互運用性の向上等国際連携の推進
- ・大学、研究開発法人等によるデータポリシーの策定、機関リポジトリへの研究データの収載
- ・データマネジメントプラン（DMP）、メタデータの付与を行う仕組みの公募型研究資金への導入
- ・研究データの管理・利活用の促進に向けた評価体系への導入

### <研究DXを支えるインフラ整備>

- ・SINETの増強と学術情報基盤を支える技術の研究開発、大学等の知を生かせる我が国の社会基盤インフラとして、民間と連携しつつ利活用できる環境整備の方策の検討
- ・「富岳」の本格共用の推進、大学、国立研究開発法人等の計算資源の増強、次世代計算資源の方向性の検討
- ・大型研究施設や共用施設・設備のリモート化・スマート化
- ・プレプリントを含む文献など、研究成果に係る情報を広く利用できる環境整備
- ・基盤分野（OS、プログラミング、セキュリティ、データベース等）を含めた数理・情報科学技術に係る研究の加速

### <高付加価値な研究の加速>

- ・マテリアル分野、ライフサイエンス分野、環境・エネルギー分野、海洋・防災分野等におけるデータ駆動型研究の振興



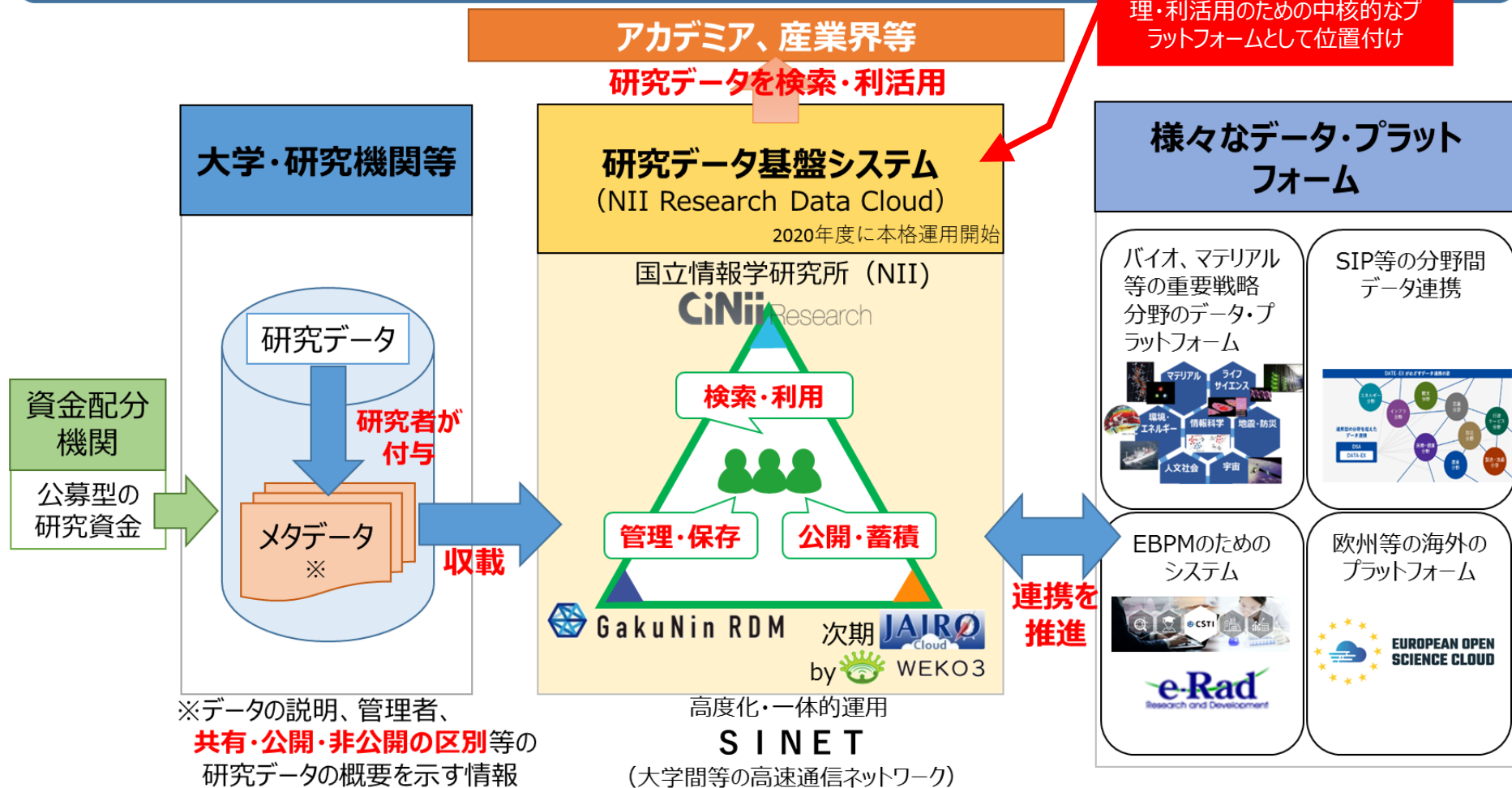
データ基盤の構築とその利活用促進によるデータ駆動型研究の振興が大きな課題

# 公的資金による研究データの管理・利活用に関する基本的な考え方について

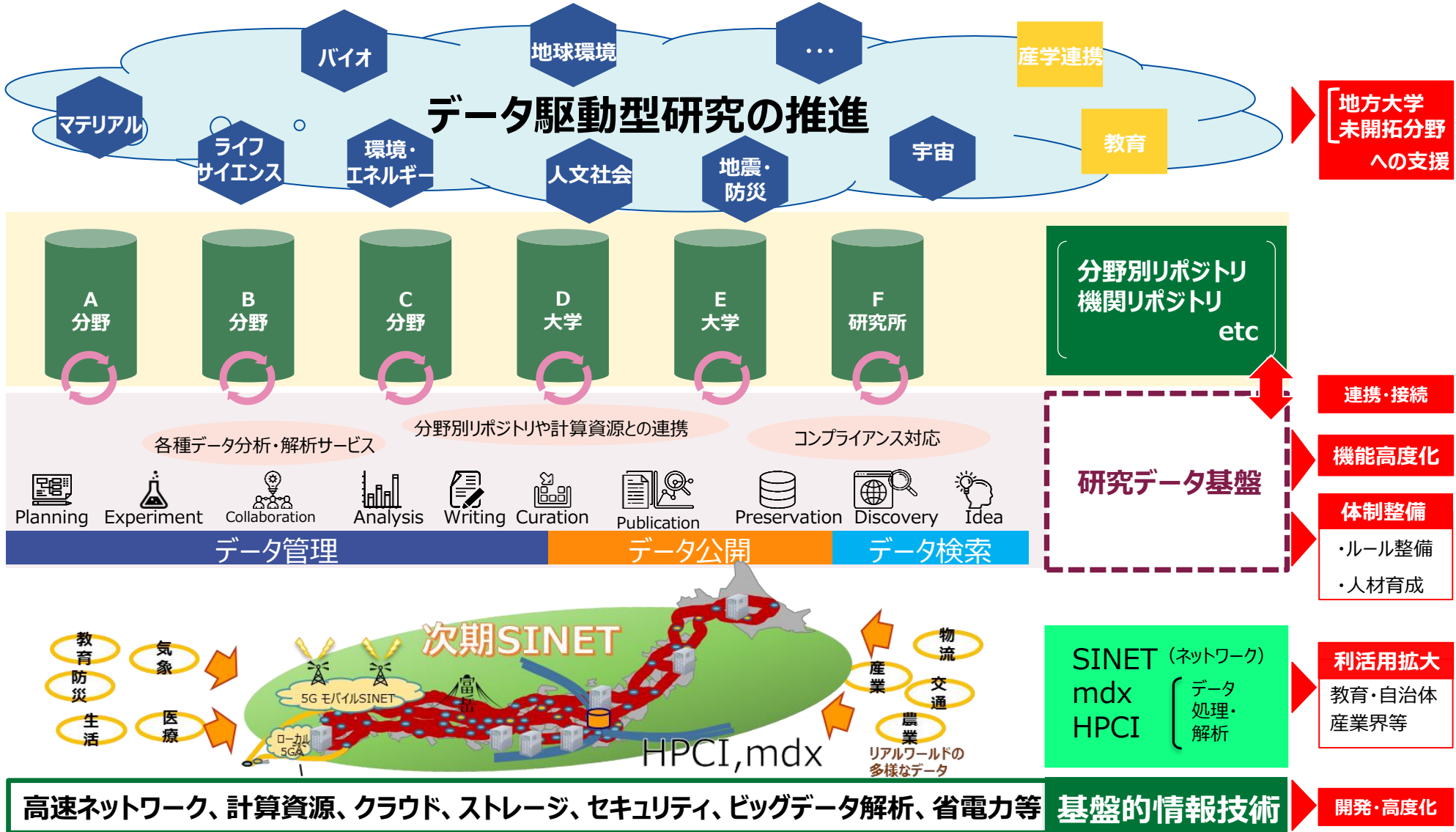
## 研究データ基盤システムを中核としたデータ・プラットフォームの構築

- 研究データの公開・共有を推進、産学官のユーザがデータを検索可能
  - **ムーンショット型研究開発制度**における試行(2020年度開始)、その後、次期SIPに導入
- ➔ **全ての公募型の研究資金**の新規公募分に導入(2023年度まで)

我が国における研究データの管理・利活用のための中核的なプラットフォームとして位置付け



# 研究DXの推進により実現すべき未来像



# 研究DX推進に向けた研究データの共有・利活用を進める上での 主な論点例（案）

---

- データ駆動型研究やデータの共有を促進するために有効な仕掛けはどのようなものが考えられるか
- 各分野における研究データや研究の基盤となるデータを戦略的に蓄積していくために必要な方策は何か
- 各分野の研究データをつなぎ、全国的な研究データ基盤を構築していく上で必要な機能は何か
- データの特性に応じた取扱いルールをどのように決めていくべきか
- 研究データの共有・利活用を進める上で、必要な人材をどのように確保していくべきか
- データ駆動型研究を進めていく上で、情報分野の研究者と各分野の研究者との連携をどのように進めていくべきか

等

## 参考資料

---

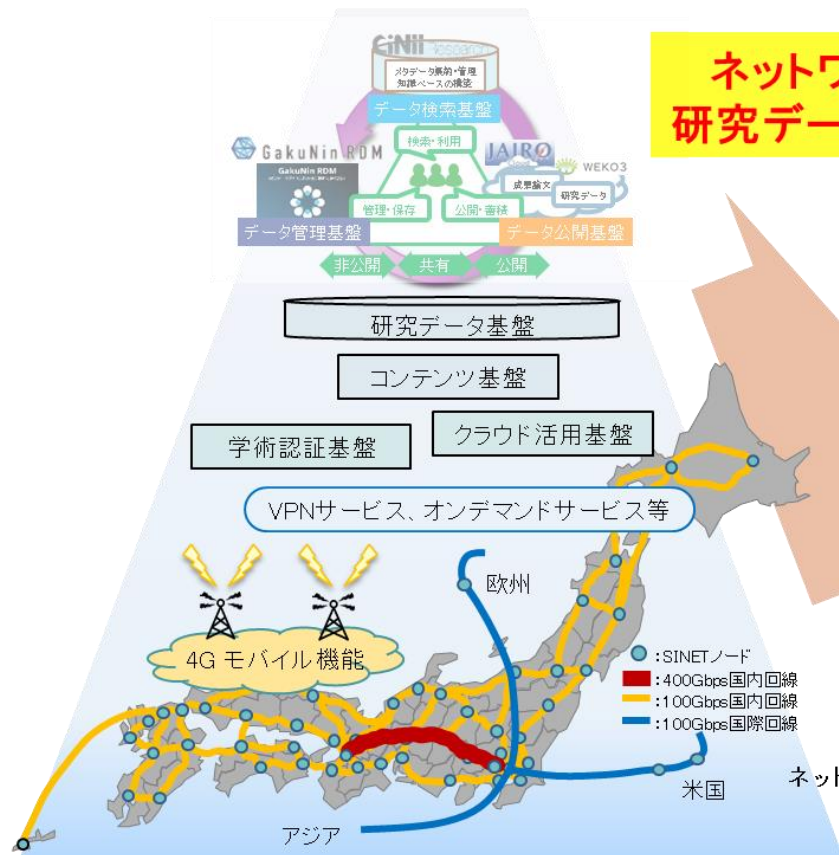
- 今後の次世代学術情報ネットワーク・データ基盤整備の方向性
- HPCI（革新的ハイパフォーマンス・コンピューティング・インフラ）
- データ活用社会創成プラットフォーム（mdx）の整備・運営
- 各分野における研究データ基盤の構築状況
- マテリアル分野におけるAI・データ駆動型研究による研究手法の革新（研究DX）
- データサイエンスに係る各国の取組

# 今後の次世代学術情報ネットワーク・データ基盤整備の方向性

- 次世代学術研究プラットフォーム：ネットワーク基盤と研究データ基盤の一体的運用
- ◆ 日本全国を400Gbpsで接続、国際回線も200Gbpsに増強・整備
- ◆ 研究データライフサイクルに沿った研究データ基盤の開発、運用（順次開発・機能拡充予定）
- ◆ SINET接続点増設でアクセス環境改善、5G対応モバイル基盤の本格運用

SINET5（2016～2021年度）

次世代学術研究プラットフォーム（2021～2027年度）



## ネットワーク基盤と 研究データ基盤を融合

研究データ基盤

- 3つの基盤を有機的に繋ぐ管理・活用・信頼・流通機能の実現
- オープンサイエンスの実践に不可欠な人材育成の仕組みを提供

ネットワーク基盤

- 全国を400Gbps以上、国際を200Gbps以上で整備
- 無線（全国セキュア5G+ローカル5G）と超高速有線との融合





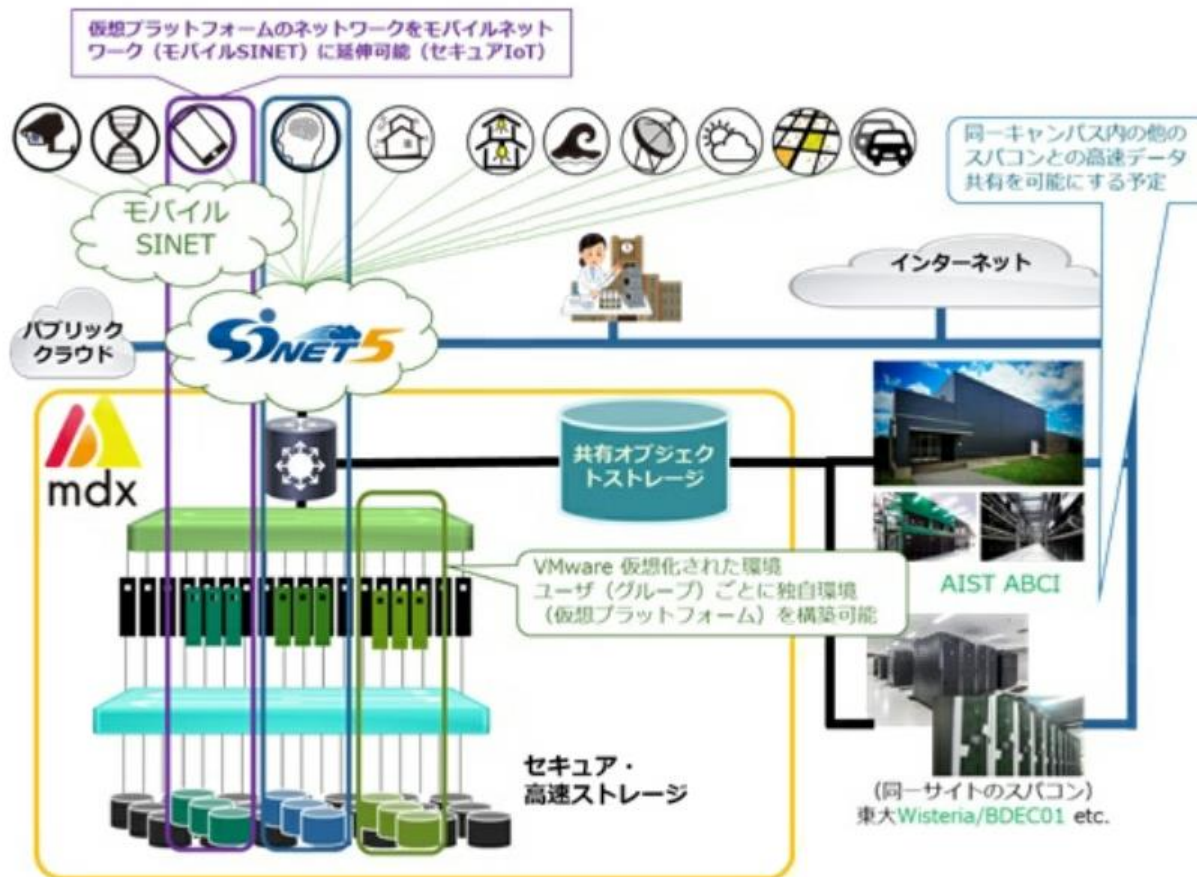
# HPCI (革新的ハイパフォーマンス・コンピューティング・インフラ)

- 国内の大学等のスパコンを高速ネットワークでつなぎ、利用者が一つのアカウントにより様々なスパコンやストレージを利用できるようにするなど、多様なユーザーニーズに応える環境を構築し、全国のユーザーの利用に供する。



# データ活用社会創成プラットフォーム (mdx)の整備・運営

- ・ 9大学2研究所が連合して共同運営する、データ活用にフォーカスした高性能仮想化環境であり、データの入出力等の様々な目的に使用できる汎用ノード群と、機械学習等の演算を高速に処理できる演算加速ノード群を有し、Society 5.0実現に貢献。
- ・ 利用者は広域網とストレージ、計算機等からなるIT環境を利用でき、mdxとSINETを用いて広域でデータを収集・集積・解析する情報基盤を容易に構築可能。2021年3月9日に稼働を開始し、2022年度より正式運用を行う予定。



※ 9大学2研究所：  
北海道大学、東北大学、  
筑波大学、東京大学、  
NII、東京工業大学、  
名古屋大学、京都大学、  
大阪大学、九州大学、AIST

# 各分野における研究データ基盤の構築状況

## マテリアル

### ●物質・材料研究機構 (NIMS)

▷無機材料や高分子材料、金属構造材料の物質構造等、NIMSの研究活動で得られたデータ、論文等から収集したデータに加え、論文等で発表した公知データをメタデータ（試料・装置・手法などの情報）を付与した形で収録したもの。

- ・物質・材料データベース (MatNavi)
- ・材料データリポジトリ (Materials Data Repository : MDR)

### ●情報統合型物質・材料開発イニシアティブ (MI<sup>2</sup>I)

▷ NIMSの研究活動で得られた物質・材料の基礎データや、データ科学的手法等の解析ツールを格納。

## 環境・防災

### ●データ統合・解析システム (DIAS)

▷国内外の各機関が保有する衛星観測、大気・地上気象観測、海洋観測、気候変動予測モデルの出力DB

### ●防災科学技術研究所 (NIED)

▷調査観測実験データ、シミュレーションデータ、画像、図面等の調査研究で得られた各種情報等

- ・高感度地震観測網 (Hi-net)
- ・基盤的火山観測網 (V-net)
- ・日本海溝海底地震津波観測網 (S-net) 等

### ●東京大学地震研究所

▷関東甲信越地域、紀伊半島とその周辺地域、瀬戸内海西部とその周辺地域の地震活動地図、地震波形データ、及び国立大学微小地震観測網データを用いた震源カタログ。

- ・地震活動に関するデータベース

## ライフサイエンス

### ●理化学研究所

▷理研における研究活動において取得されたデータやデータベースに関するメタデータのDB

- ・理研メタデータベース
- ▷天然化合物の名称・構造、生物活性情報のほか、物性データや機器分析データなど。
- ・NPEDIA
- ▷生命科学分野の画像データと生命システム動態の定量データ
- ・SSBD:database
- ・SSBD:repository
- ▷革新脳プロジェクトによる、マーマセットの脳のMRI画像、マクロレベルの結合、遺伝子発現マップ、ECoG電極アレイによる機能データ
- ▷タンパク質の量子化学計算値データベース
- ・FMOデータベース (タンパク質の量子化学計算値データベース)
- ▷理化学研究所が運用する全天X線監視装置MAXIの観測データ。
- ・全天X線監視装置MAXIデータベース

### ●日本医療研究開発機構 (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」適用のAMED研究開発課題における、研究成果に紐づくゲノムデータ及び臨床情報や解析・解釈結果等を含めたゲノム情報等。

- ・AGD (AMEDゲノム制限共有データベース)
- ▷東北メディカル・メガバンク計画のゲノム・コホート調査由来の、健康調査及び全ゲノム配列情報を含む生体試料の解析情報（基本属性情報、調査票情報、生理学検査情報、検体検査情報、ゲノム・オミックス情報、診療情報、MRI画像情報等）を統合したDB等
- ・dbTMM
- ・jMorp

### ●国立遺伝学研究所

- ▷DNA塩基配列と機能注釈データのDB
- ・DNA Data Bank of Japan (DDBJ)
- ・Japanese Genotype-phenotype Archive (JGA)
- ・AMED Genome Group Sharing Database (AGD)

### ●大阪大学蛋白質研究所

- ▷生体高分子の立体構造に関するDB
- ・Protein Data Bank Japan (PDBj)

## 海洋・原子力・宇宙

### ●海洋研究開発機構 (JAMSTEC)

▷機構の研究設備等による調査研究の調査観測データ、シミュレーションデータ、画像、図面等の各種データ等。

- ・航海・潜航データ・サンプル探索システム (DARWIN)
- ・深海映像・画像アーカイブ (J-EDI)
- ・地震研究情報データ提供システム (J-SEIS) 等

### ●日本原子力研究開発機構 (JAEA)

▷各々の試験条件、試験結果等や緩衝材の基本特性試験によって得られたデータなど

- ・ガラスの溶解に関するデータベース
- ・緩衝材基本特性データベース 等

### ●自然科学研究機構国立天文台

▷国立天文台及び大学の観測装置で共同利用等により取得した天文観測データ。

- ・SMOKA
- ・すばる望遠鏡アーカイブシステム (STARS) 等

## 人文・社会科学

### ●人間文化研究機構国文学研究資料館

▷日本古典籍ナショナルカタログと、国文研及び国内外の大学等が所蔵する日本文学を含む古典籍の全冊画像のデータベース。

- ・日本古典籍総合目録データベース (目録)
- ・新日本古典籍総合データベース (画像閲覧)

### ●一橋大学経済研究所

▷近代日本経済の歴史統計を経済活動の諸分野にわたって推計、加工して体系的に集めた統計書である『長期経済統計』全14巻 (1965-1988年) のデータベース。

- ・長期経済統計データベース (LTES)

# マテリアル分野におけるAI・データ駆動型研究による研究手法の革新（研究DX）

## 政策的意義

- AI・量子・バイオ・環境に続く政府戦略である「**マテリアル革新力強化戦略**」では、革新的・高速な研究開発のための**AI・ビッグデータによるデータ駆動型研究の推進が主要課題**。データ駆動型研究の開発基盤の整備と同手法による課題解決の重要性が指摘。
- 研究データのデジタル化による**データ検索**に留まらず、他国に例のない**AI・データ駆動研究による研究手法のDX**を先駆的に推進。

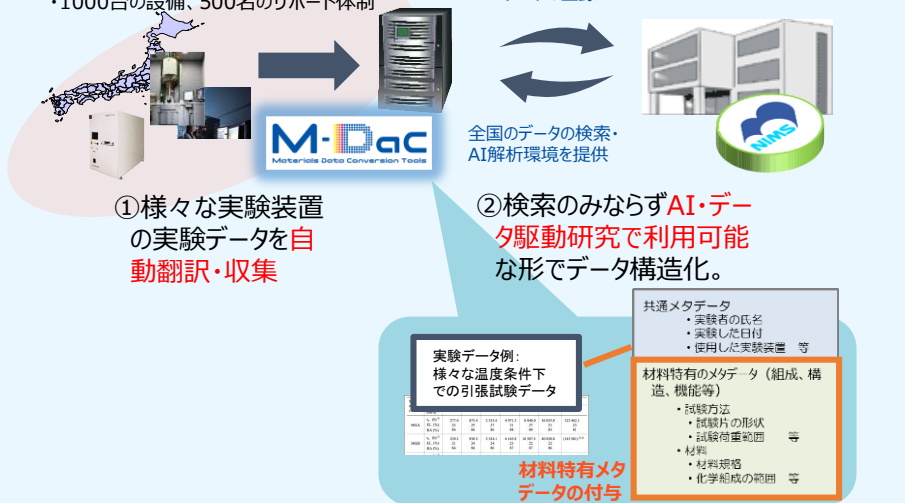
## 研究DXのロールモデルとしてのマテリアル

### ✓ 機関の壁を越えたAI・データ駆動研究環境整備を推進

- 全国の先端共用設備で創出される実験データを、検索のみならず、**AI・データ駆動研究で利用可能な形で自動収集・翻訳（本年度より本格実施）**

全国の大学等の先端共用設備  
（マテリアル先端リサーチインフラ）  
・全国25の大学等体制  
・1000台の設備、500名のサポート体制

構造化されたデータの登録  
全国データの共用・AI解析システム  
（データ中核拠点（NIMS））



- データ共有の**インセンティブ設計**やデータ利用の**ルール整備**、AI解析のための**全国データ構造化、産業界との連携**について具体的に検討・整備。
- 化学MOP（NIMS）など、データを基軸とした産業界との水平連携

### ✓ 産学でデータ駆動型研究開発による研究手法革新の高いニーズ：

- マテリアル分野では、従来、試行錯誤型で研究開発。データ駆動型研究導入で膨大なマテリアルの検討が可能となり、革新的・効率的な成果創出が実現
- 【想定例】従来30年かかったバッテリーの研究を1年に短縮
- バッテリーの電解液の組み合わせは、100万通り以上。
- データ駆動型研究で、従来30年かかる電解液探索を1年に短縮

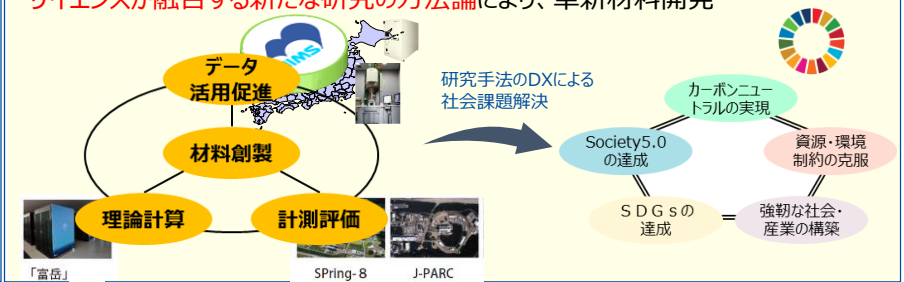


### ✓ 構築された先端大型共用施設との連携基盤：

- 大型施設（計算、放射光、中性子）との連携を既に実践し連携基盤を構築済
- 参考）元素戦略で実践した連携体制
- 参考）先端大型共用施設との連携実績



### ✓ データ創出・活用型マテリアルプロジェクトで、計算・計測・データサイエンスが融合する新たな研究の方法論により、革新材料開発

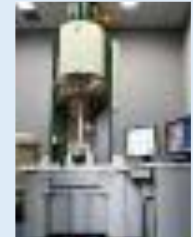


# 機関の壁を越えた全国データ共有によるAI・データ駆動型研究環境

マテリアル先端  
リサーチインフラ

## マテリアルデータをつくる

- 大学・国研の先端設備の共用支援体制を提供  
電子顕微鏡、半導体加工装置など、全国1000台以上
- 自動化・ハイスループット化により、質の高いデータを大量に創出
- データ活用人材等によりデータ蓄積をサポート



データ中核拠点  
(NIMS)

## マテリアルデータをためる

- データ創出基盤と連携したユーザー認証機能、データ登録機能（ユーザー毎のクラウド専用領域）を提供
- ステータス（オープン、シェアクローズド、クローズド）に合わせたデータ管理機能を提供

NIMS reference code	Type of data	Year of data generation	Product name	Discussion topic	Discussion date	Processing and experimental details	Keywords	Researcher name	Research institution	Open access status
MGA	SEM	2000	Si-Al-Fe alloy	SEM/EDS	10/10/00	Electron micrograph and EDS data	SEM, EDS	Y. Imai	NIMS	Open access
MGA	SEM	2000	Si-Al-Fe alloy	SEM/EDS	10/10/00	Electron micrograph and EDS data	SEM, EDS	Y. Imai	NIMS	Open access

NIMS reference code	C	Si	Al	Fe	S	Na	Ca	Mg	Ni	Nb	N	AsP	1st
MGA	0.07	0.28	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MGA	-0.14	-0.51	-0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

NIMS reference code	Si	Al	Fe	S	Na	Ca	Mg	Ni	Nb	N	AsP	1st
MGA	0.10	0.38	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MGA	0.10	0.34	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

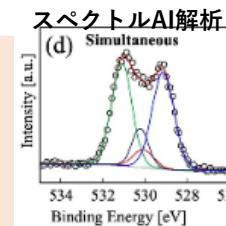
NIMS reference code	Si	Al	Fe	S	Na	Ca	Mg	Ni	Nb	N	AsP	1st
MGA	0.10	0.34	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MGA	0.10	0.34	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

データ中核拠点  
(NIMS)

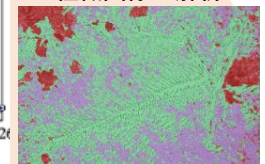
## マテリアルデータをつかう



- NIMSの世界最大級データベースと連携したデータ検索機能
- データ駆動型研究を促進するAI解析機能



組織画像AI解析



例 1) 微細構造画像データ + AI解析 → 多元系合金設計技術の開発

例 2) 圧電体特性データ + 加工プロセスデータ + AI解析 → 誘電体デバイス設計

# データサイエンスに係る各国の取組

## 【EU】

- **欧州統合データ基盤プロジェクト (GAIA-X)**  
欧州における効率的/競争的/安全/信頼できるデータ基盤サービスプロバイダーの連携を支援するプロジェクト。  
次世代の各機関連携された欧州データ基盤を設計するとともに、欧州のデータインフラにおける共通要件を定義し、参考実装を実施。
- **欧州データ戦略 (A European strategy for data)**  
「データの単一市場」である「欧州データ空間」構築を目指す(2020.2)。  
独仏が主導するプロジェクト「GAIA-X」とのシナジーも重視

## 【ドイツ】

- **ドイツ研究データインフラストラクチャ(NDFI)構築**  
分散的・時限的に保存されていた研究データを共通基盤に集積して「使えるデータ」化する構想に毎年9,000万€(2020-2029)の投資が計画。
- **データ統合シミュレーションの研究開発を実施**  
第1期は工学的アプローチからのモデル化などを実施(5年x2 合計約6,500万€)、第2期ではさらに大容量データやセンサー計測の取り扱い、データ解析手法の開発などを進め、統合システム科学に発展させる(1期7年 600万€/年 = 予定) 応用分野は、環境、健康、製造業など

## 【フランス】

- **研究とイノベーションのための包括的で全国規模のeインフラの構築に2億2400万€を資金配分**
- **大学のデジタル化に合計5600万€の投資 (フランス復興計画)**

## 【中国】

- **デジタルインフラ整備への大規模投資**  
5G、IoT、AI、科学イノベーション施設等が対象で、追加投資額は2025年までに約10兆元(約150兆円、約1兆ドル規模)。

## 【インド】

- **インドにおけるデジタルID基盤 (India Stack)**  
国民ID「アダール (Aadhaar)」と連携しているデジタルID基盤。アダールには個人の生体情報(指紋・光彩など)が登録されており、納税者番号や銀行口座などもひも付けられている。これにより、補助金や公共サービスのスムーズな提供、電子決済が実現。

### 研究データ基盤の現状

日本の研究データ基盤も2020年度に運用を開始予定であるが、研究開発経費は世界に比して極めて低い



"By 2020, we want all European researchers to be able to deposit, access and analyze European scientific data through a **European Open Science Cloud**..."  
(Speech by Commissioner Carlos Moedas in 2016)



• 世界各国は2010年代よりデータ駆動型研究基盤開発に投資し、最近ではオープンサイエンス実現に向けた基盤を整備