

戦略策定に向けた主な視点について マテリアルズ・インフォマティクス

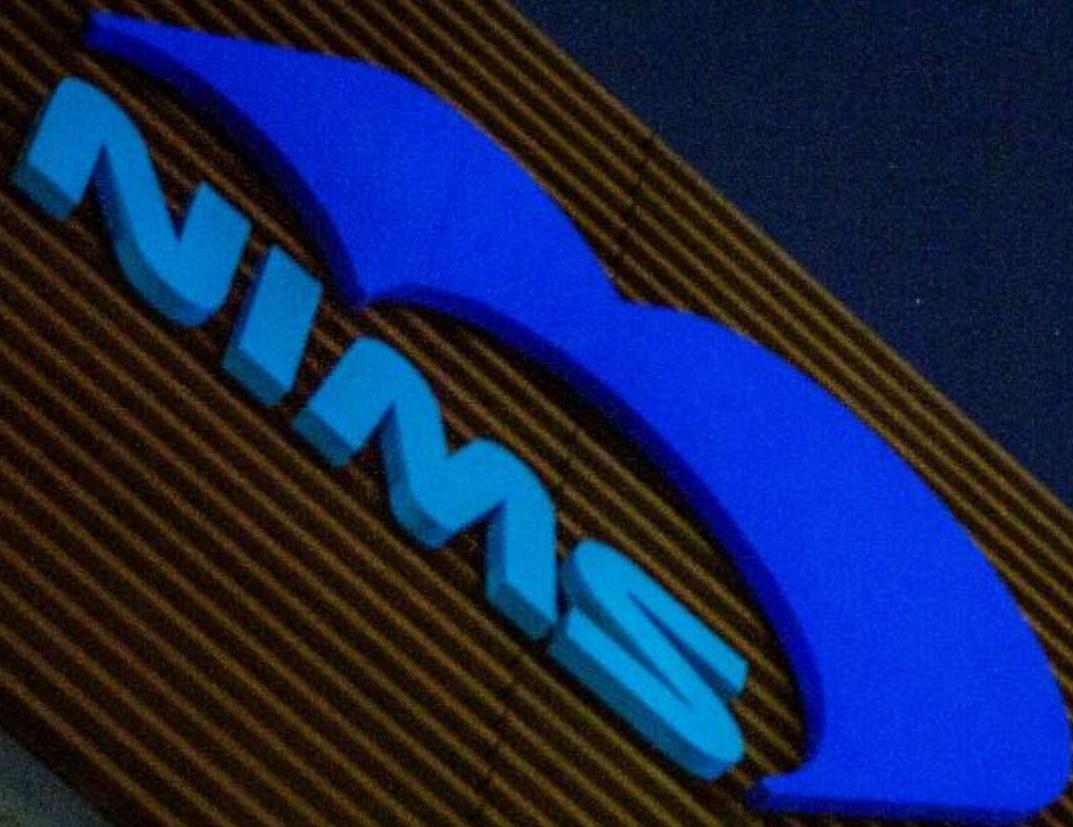


令和 2 年 1 0 月

内閣府

令和2年10月21日

NIMSにおけるマテリアル データ戦略



物質・材料研究機構(NIMS)

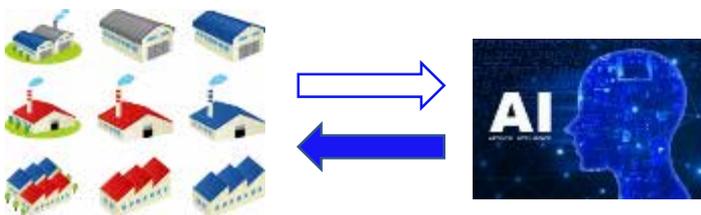
理事長 橋本和仁

マテリアルデータをめぐる世界の覇権争い

1. AIに強みを持つ企業 (IBM, Google, etc.,)

日本産業新聞
2018年3月27日

材料企業からデータ提供を受け
付加価値を付けて返すビジネス



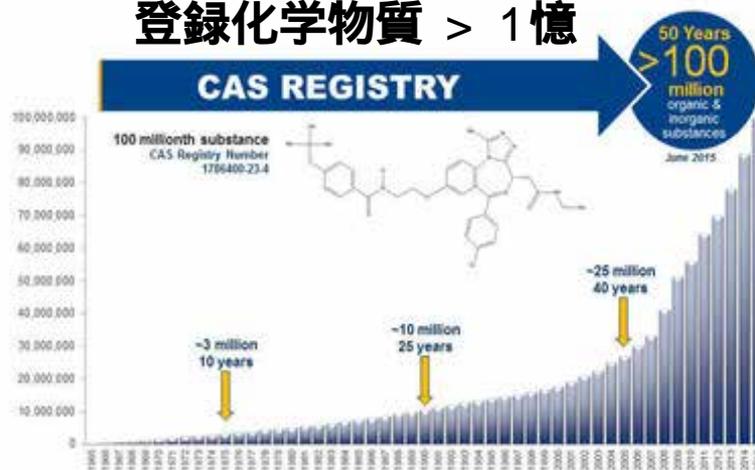
材料企業からデータを根こそぎ収集

2. 学術誌出版社 (エルゼビア、学会傘下の出版社)

米化学会



登録化学物質 > 1億



- 完全電子化システムで情報分析
- 人海戦術 (3000人体制?)
- 論文・特許情報 (合成法等)

論文、特許からデータを収集

NIMS : 世界最大級の材料データベース



MatNavi 専門家がキュレーションした12のデータベース

無機材料



世界最大の無機材料データベース

	Atom Work	Atom Work Adv.
結晶構造	82,000	303,885
状態図	15,000	42,406
特性	55,000	365,517
更新	なし	年1回



高分子



33万物性
(2019.4現在)

国産高分子データベース

- ・ 学術論文からの精選データ
- ・ 人手による高品質データ
- ・ 過去20年の蓄積 (10人体制)

金属・合金



500種以上の鉄鋼材料の
機械的性質、クリープ、
疲労 : 82,700以上



CCT曲線図 : 214; 硬度:
2213; 金属組織図 : 627

マテリアル研究所の特性を活かして世界
の材料データプラットフォームへ

NIMSにおける材料データプラットフォームの環境整備

📍 平成28年度補正予算



M-cube (M³)棟
4974.6/m²

📍 平成29年度補正予算



データサーバー
10PB/0.5PFLOPS

📍 平成30年度、令和元年、2年度 運営費交付金、理事長裁量経費

- ・ NIMSで作りに出されるデータをすべてデータベースに
- ・ 化学4社と高分子データベース作成 (MOP)

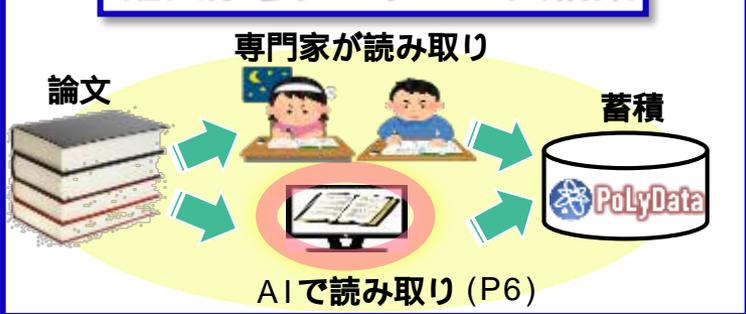
📍 令和3年度 概算要求中

- ・ 全国の大学等で作りに出されるデータをデータベースに
- ・ 高分子特許をデータベース化 (経産省)

NIMSの取り組み (マテリアルデータプラットフォーム)

データをつくる

論文からデータベース構築



実験データの自動収集 -スマートラボトリー化



データをためる

集めたデータを 統合・機械可読化



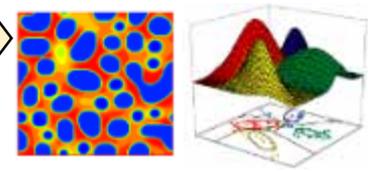
データをつかう

材料開発

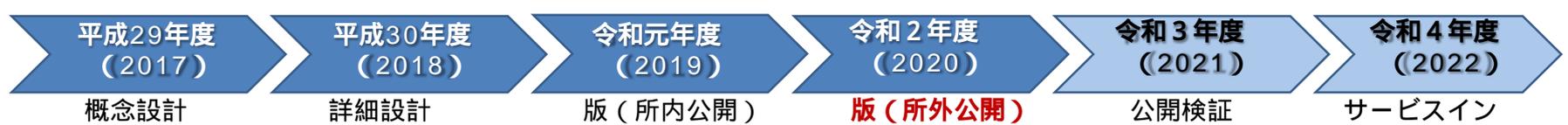
- 解析
- 評価
- 予測
- 物性選択
- 材料選択

AI

解析アプリ
変換ツール



産学連携

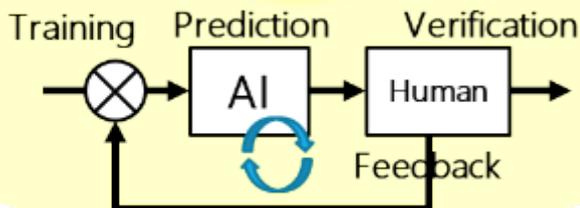


テキストデータマイニングによる 論文からの自動データ抽出

論文

AIでまとめ取り(アカデミックディスカウント)

高分子データ基盤



自動抽出成功例：

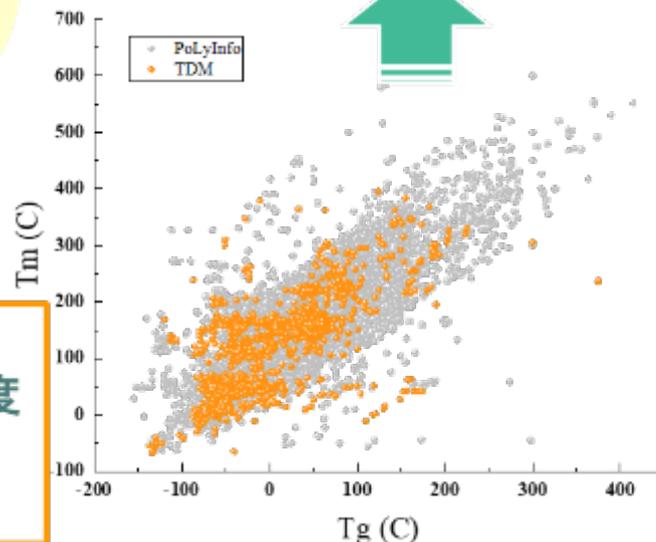
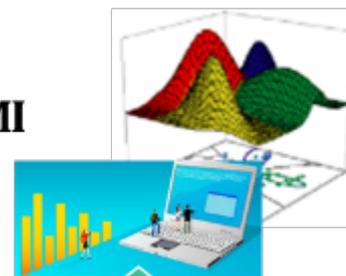
- ・ 平均分子量 21,169 点
- ・ 熱分解温度 10,085 点
- ・ ガラス転移温度 7,581 点
- ・ 溶融温度 5,583 点
- ・ 破壊伸び 4,193 点
- ・ 引張弾性率 4,187 点
- ・ 結晶化度 3,349 点
- ・ 引張破壊応力(強さ) 3,318 点
- ・ 結晶化温度 3,144 点
- ・ 溶融熱 3,035 点
- ・ 電気伝導度 2,209 点
- ・ 密度 2,083 点
- ・ 結晶化熱 1,371 点

61,046論文から
 ・ 7,581点のガラス転移温度
 ・ 5,583点の融点
 を自動抽出

今この瞬間も増量中

85物性 (対PoLyInfo
72.0%)
86,415点 (同 25.8%)

MI



共同研究機関



国立大学法人
奈良先端科学技術大学院大学
NARA INSTITUTE OF SCIENCE and TECHNOLOGY

実験データの自動収集

測ったそばから**使える形**でためる！

実験装置いろいろ

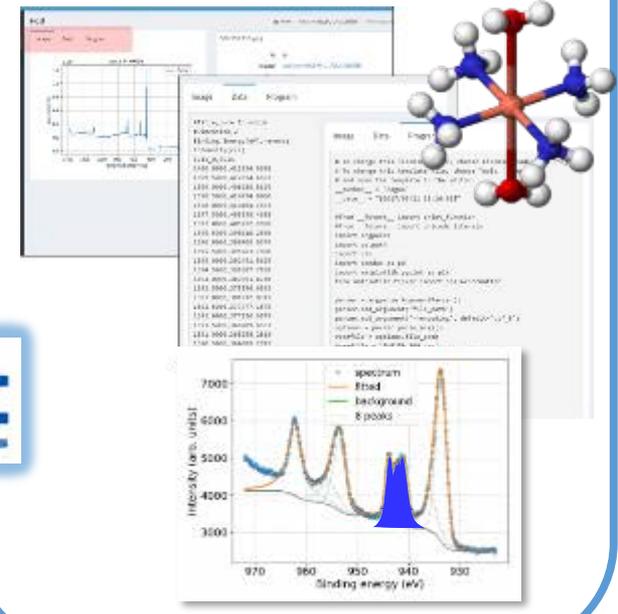


装置出力データ

70	8E	74	00	00	00	00	00
00	18	7F	DA	BB	E4	4	
00	00	00	00	00	00	00	0
00	00	00	00	00	00	00	0
58	C2	48	F8	43	C0	4	
48	C2	23	BD	48	E1	7	
7F	B7	48	40	A3	B5	4	
BC	AC	48	CE	60	B0	4	
46	AA	48	5C	29	A9	4	

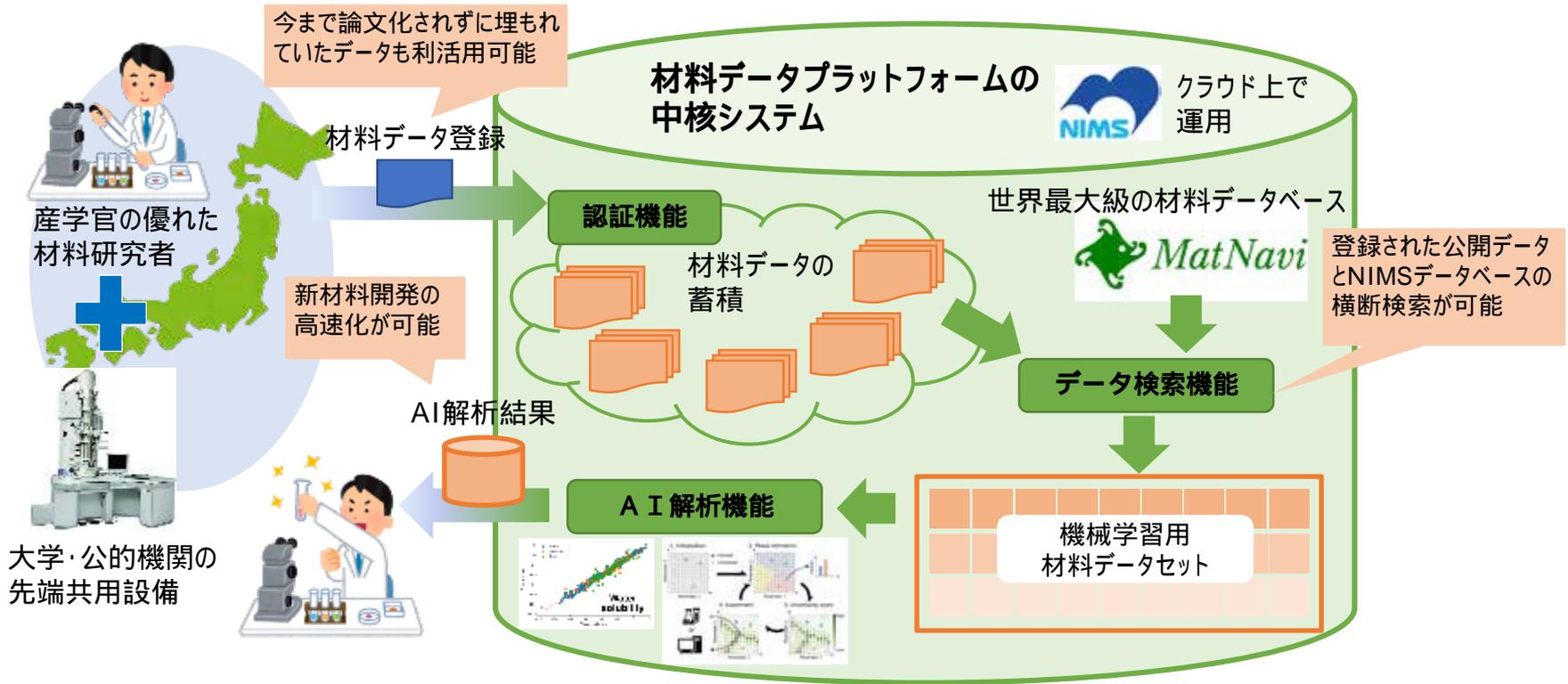


データをためる



実験データを解析に使える形に！
計測機器メーカーの協力で、XPS及びXRD
について**自動翻訳ツール**を製作し、一般公開

材料データの中核拠点へ（概算要求中） 日本全国の材料データの集約



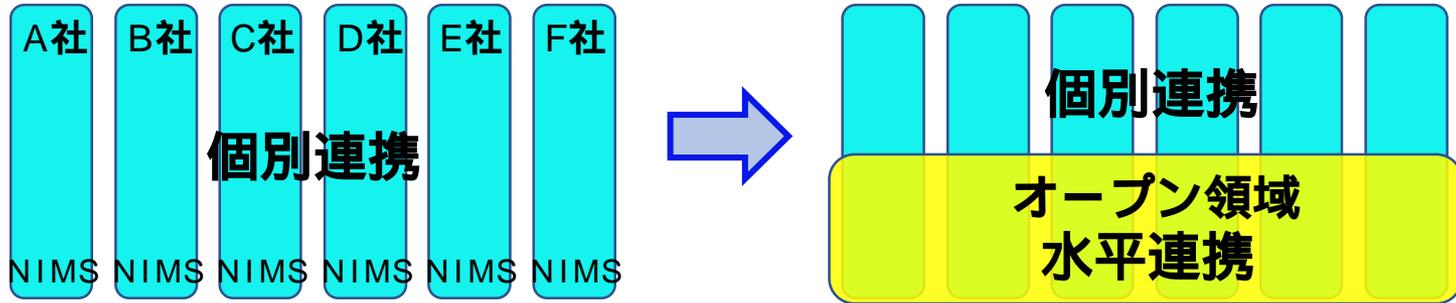
全国展開データ流通基盤開発



17	18	19	2020	2021	2022	2023	2024	2025
----	----	----	------	------	------	------	------	------

産業界との連携(1)

同業多社によるオープンイノベーション

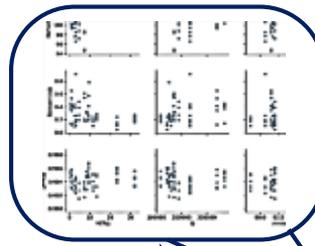


オープン領域：データを共有し、MI解析を共同で

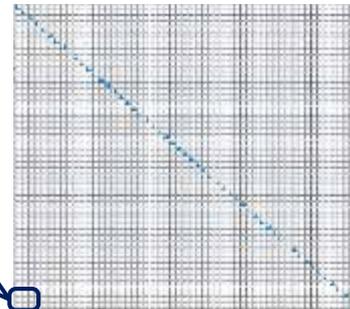
- ・ **データは新たにNIMSが測定** ・ **ポリオレフィン** ・ **サンプルを共有**

化学MOPのデータセット：**品質のそろった**ポリオレフィン実験データ約4万件

237項目

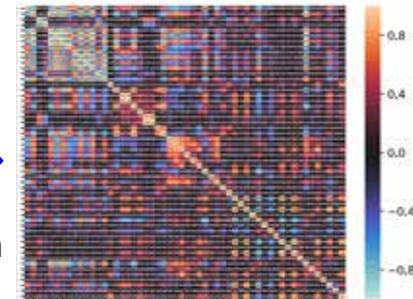


一部を抽出



相関関係

Python
で解析



相関係数

10

業界団体（オールジャパン）で結集

令和2年2月26日
日本経済新聞 朝刊 1面

特許情報のデータベース化を 産業政策としての呼びかけ

（経産省製造局素材課）

- 1 特許情報のAI用のフォーマット・材料特性データベースの整備：共通フォーマット
- 1 化学系企業のMI責任者とNIMSなどでAI用フォーマットを設計。
- 1 化学系企業の知財部員・研究員等が特許情報からデータを手動で抽出：信頼性担保
- 1 データはNIMSのデータサーバーへ蓄積・保管：PolyInfoなど他のデータベースとの連結

企業の素材開発の迅速化のため、各社のAI活用をサポートする共通情報基盤を構築し、日本企業に提供



課題1：データフォーマット

MIに使うためにはフォーマットをそろえることが必須

- **マテリアルでも分野により異なる**
高分子材料：複雑な命名規則（名寄せの工夫）・マルチスケール構造
無機材料：単純な命名規則（小規模辞書による統制）・微細構造
- **アカデミアと産業界で異なる（高分子を例に）**
アカデミア：基礎物性・新規ポリマー重視
産業界：応用物性・既存ポリマー・コンパウンド重視
- **産業界でも各社の戦略により異なる（高分子を例に）**
A社主要製品：イオン交換膜
B社主要製品：接着剤
- **データガバナンスを優先するか、拡張性を優先するか**
学術論文志向型：データ型の統制で利便性を重視 他DBとの連携に有利
特許志向型：自在な型で拡張性を重視 DBの独立展開に有利

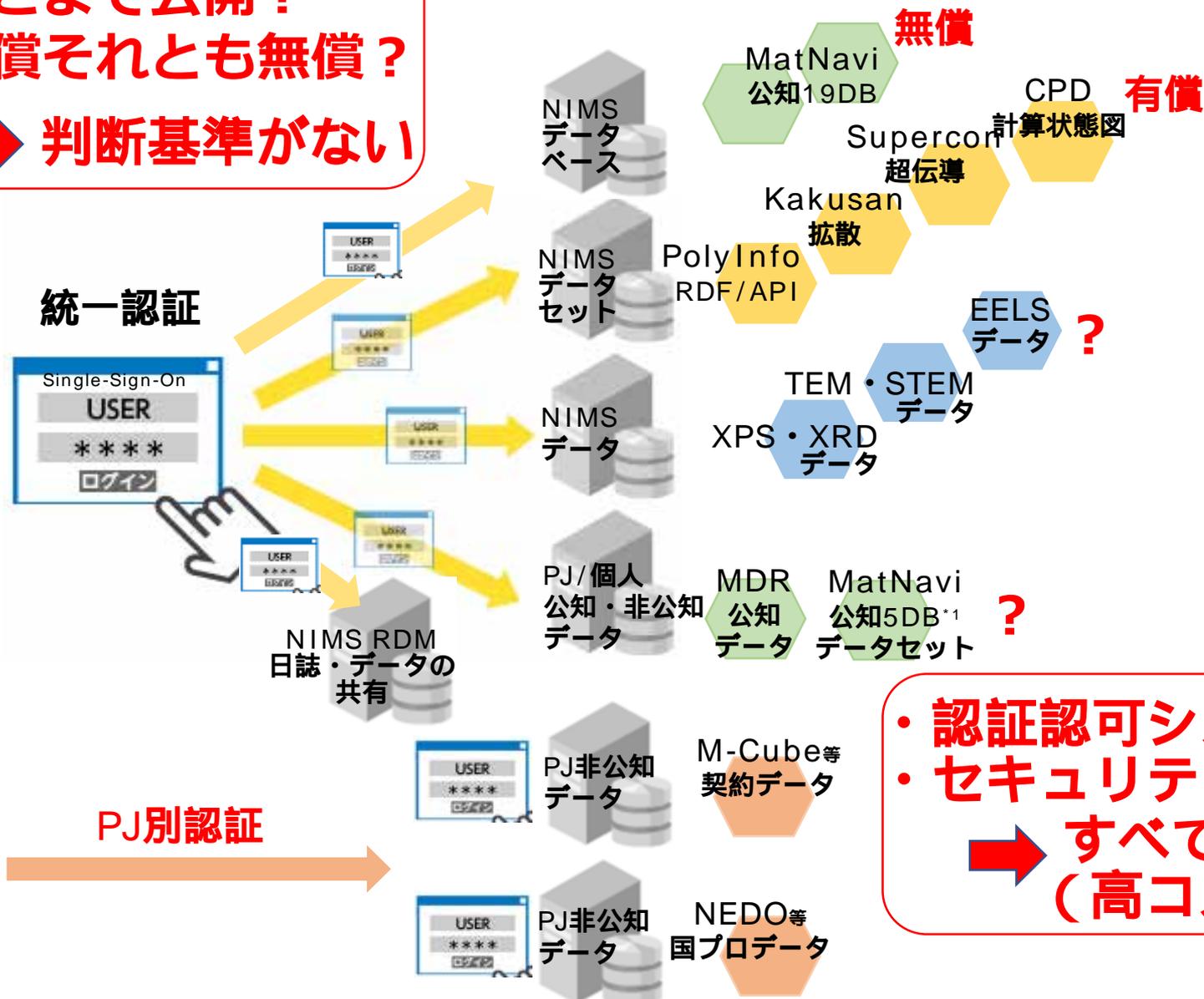
...

誰かが先頭に立ってユーザーの意向をまとめる必要がある

極めて強いリーダーシップと汗をかく人が必要

課題2：オープン・クローズ戦略

- どこまで公開？
 - 有償それとも無償？
- ➡ 判断基準がない



課題 3 : 費用負担

最高のデータプラットフォームを保ち続けるには お金がかかる

- ・ サーバー管理費用、更新費費用、 etc.
- ・ 公表データ（学術論文、特許）の文献料、
取り込み費用、キュレーション費用、 etc.

国費？

- ・ 国研や大学の運営費交付金では限界
- ・ 不安定な補正予算を財源にするのは危険

産業界も負担？

- ・ 役に立つことがわかれば支出する
- ・ 役に立つことがわかるまでは出せない!?

産学官が一致してオールジャパンで取り組むべきでは

文部科学省における データを基軸としたマテリアル研究開発の推進について

令和2年10月
文部科学省

データを基軸とした研究開発の推進について

< 基本的な考え方・目標(データを介したマテリアル研究開発のイノベーション創造) >

データを基軸とした研究開発手法を全国の産学官の研究者が広く活用することを可能とし、「人×資金×データ再利用率」により、マテリアル分野において我が国が圧倒的な研究力を発揮していく。そのため、信頼性の高いマテリアルデータを全国で広く共有・活用し、研究成果最大化+データを介した産学官の連携を加速する仕組みを構築。

< 取り組むべき解決課題 >

どのような形でデータを蓄積・共有するか(主体、データ形式)?

どのようにデータ蓄積・共有を押し進めるか(ルール、指針)?

どのように全国でのデータ利活用を実現するか(システム)?

データを基軸とした研究開発の推進について

どのような形で蓄積・共有するか？

→ 産学、産産、学学連携を加速するための、データハブ(特定大学、国研)を形成
機関を越えたデータ共有を可能とするためのデータフォーマット(データ構造)

データハブの形成

<データの蓄積>

・先端設備共用支援体制により全国の材料研究者が創出する良質なデータを蓄積・共有していく。そのため、データ創出のための設備、人材の配備等を通じ、特定大学・国研をデータ創出ハブとして整備(来年度全国6ハブ整備を想定)。

<データの共有>

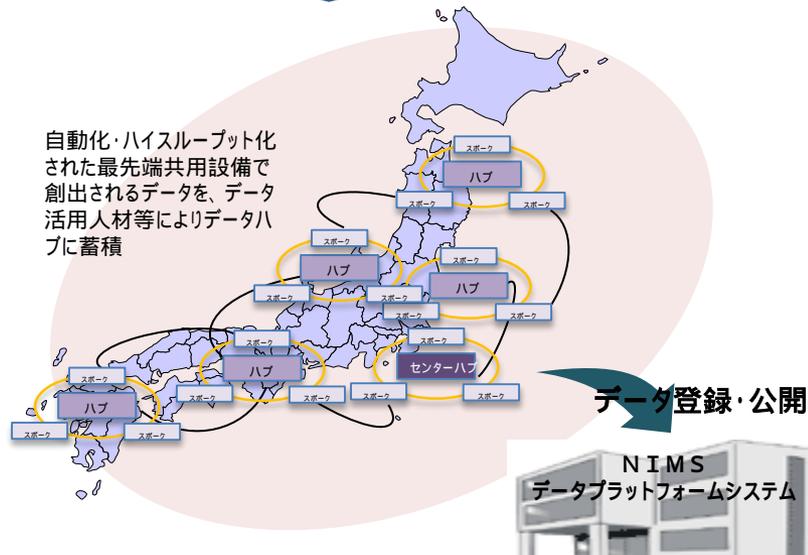
・データ創出ハブで創出したデータは、公開可能な状態になり次第、NIMSの公開システムで順次公開。また、NIMSが保有する世界最大級の材料データベースと一体的に利活用(マテリアル・インフォマティクスに利用)。

大学・国研の先端設備共用支援体制(ナテクプラットフォーム)

・電子顕微鏡や半導体加工装置など、先端共用設備の利用をトータルサポート
(年間3000件を越える利用実績、
ユーザー 産:3割、学:7割)



データ創出ハブ形成



データを基軸とした研究開発の推進について

データ構造

< 課題 >

- 研究設備や研究内容により、データの形式・必要なデータ項目(種類)が異なる一方で、データの蓄積・共有・利活用のためにデータ間で互換性を確保する必要。
- 異なるデータを共有・活用するためには、データの補足情報(メタデータ)を系統的に記載すること及びデータ標記のゆれの解決が課題。
- 上記課題に対応の上で、各研究者が研究課題解決用いたデータセットの作り方(データ構造)は研究者間で共有し、新規課題解決に生かすことが重要。また、経産省が主導する、日系計測・分析機器メーカーのデータ標準化検討の成果活用も期待。

(データ互換性確保のイメージ)

データの補足情報(メタデータ)の系統的な記載

- 例えば、引っ張り強度試験結果を他の研究者が参照する際、誰が、どのような形状の材料で実験したのかといった情報を整理して記載しておくことが必要。

データ標記の統一

- 例えば、Temperature、摂氏、華氏などの項目標記の違いの解決が必要

共通メタデータ

共通メタデータ

データ

データ

共通メタデータ

- 実験者の氏名
- 実験した日付
- 使用した実験装置 等

データの種類毎の詳細メタデータ (引っ張り試験データ補足説明)

- 試験方法
- 試験片の形状
- 試験荷重範囲 等
- 材料
- 材料規格
- 化学組成の範囲 等

どのように作成されたか:
作成条件

どのような材料か:
組成データ

どんな機能が:
様々な温度条件下での
引張試験データ

NIMS referenc. code	Type of melting ⁽²⁾	Size of ingot (kg)	Description of process	Product form	Dimensions (mm) ⁽³⁾	Processing and thermal history ⁽⁴⁾	As-cast grain size number ⁽⁵⁾	Rockwell hardness (HRC)	Non-metallic inclusions ⁽⁶⁾ (%)					
MGA		3.500	Al-Si-killed		50.8 OD 8.0 WT 6.000 ID	Hot extruded and cold drawn 1 063°C / 10 min AC 780°C / 60 min AC	10.1	16	dA = 0.05 dB = 0.05 dC = 0.02					
Chemical composition (mass percent) ⁽¹⁾														
						S	Ni	Cr	Mo	V	Nb*	N	Al*	Ti*
MGA						0.010	≤0.010	8.0	0.85	0.18	0.06	0.030	≤0.04 ⁽⁷⁾	— ⁽⁸⁾
MGB						0.010	≤0.010	9.5	1.05	0.25	0.10	0.070	≤0.02	≤0.04
Tensile strength: 500 °C														
MGA						40	320	300	290	270	250			
Yield strength: 500 °C														
MGA						875.4	2 333.4	4 971.3	8 840.0	36 035.8	123 442.1			
MGA						24	27	31	27	21	24			
MGA						84	86	88	89	85	81			
Elongation: 500 °C														
MGB						239.2	938.5	2 334.1	6 163.8	10 507.3	40 858.8	(163 981) ^(9,10)		
MGB						21	24	25	22	22				
MGB						84	86	86	87	87	86			

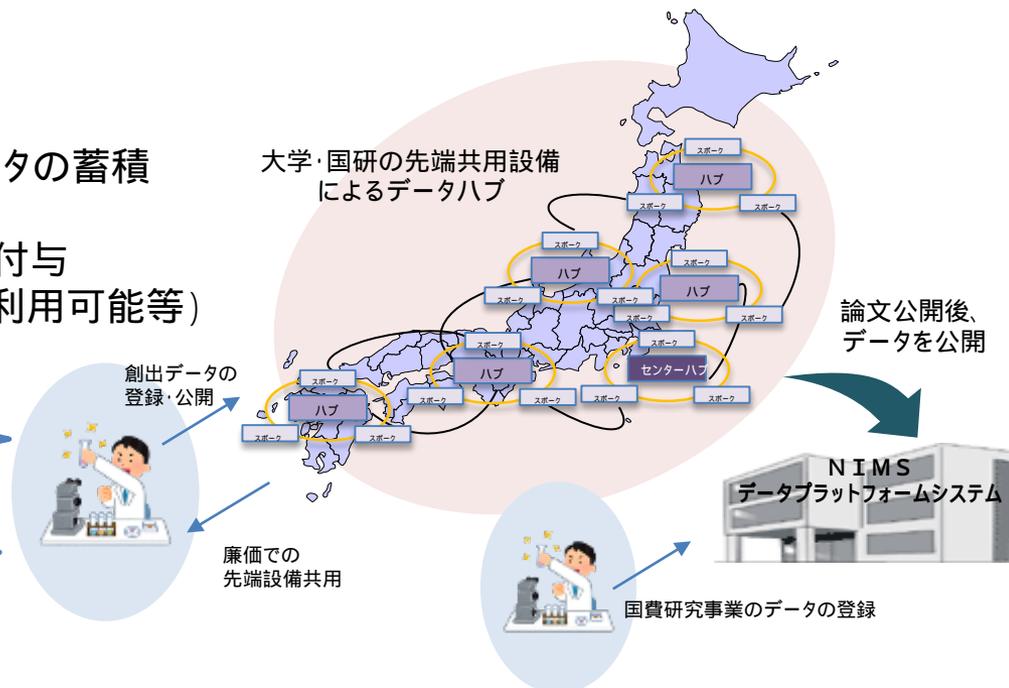
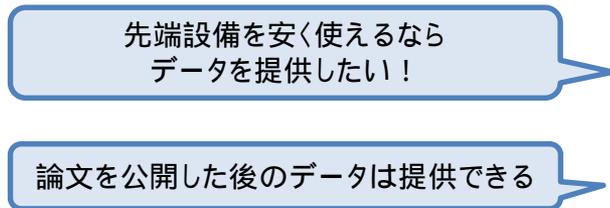
データを基軸とした研究開発の推進について

どのようにデータ蓄積・共有を押し進めるか？

- これまで蓄積・共有されなかった試行錯誤データが蓄積されるための仕組み
オープン・シェアクローズ・クローズといった種類毎に、データの取り扱いの整理、収集戦略

想定されるデータ蓄積の仕組み

- ・国費によるマテリアル研究で創出されたデータの蓄積
- ・研究者のデータ蓄積に対するインセンティブ付与
(データ提供者は低価格で先端共用設備が利用可能等)



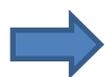
データの取扱い

- ・データ取り扱いに関する共通指針の策定(有識者、文科、経産課長級の会合を設置済)

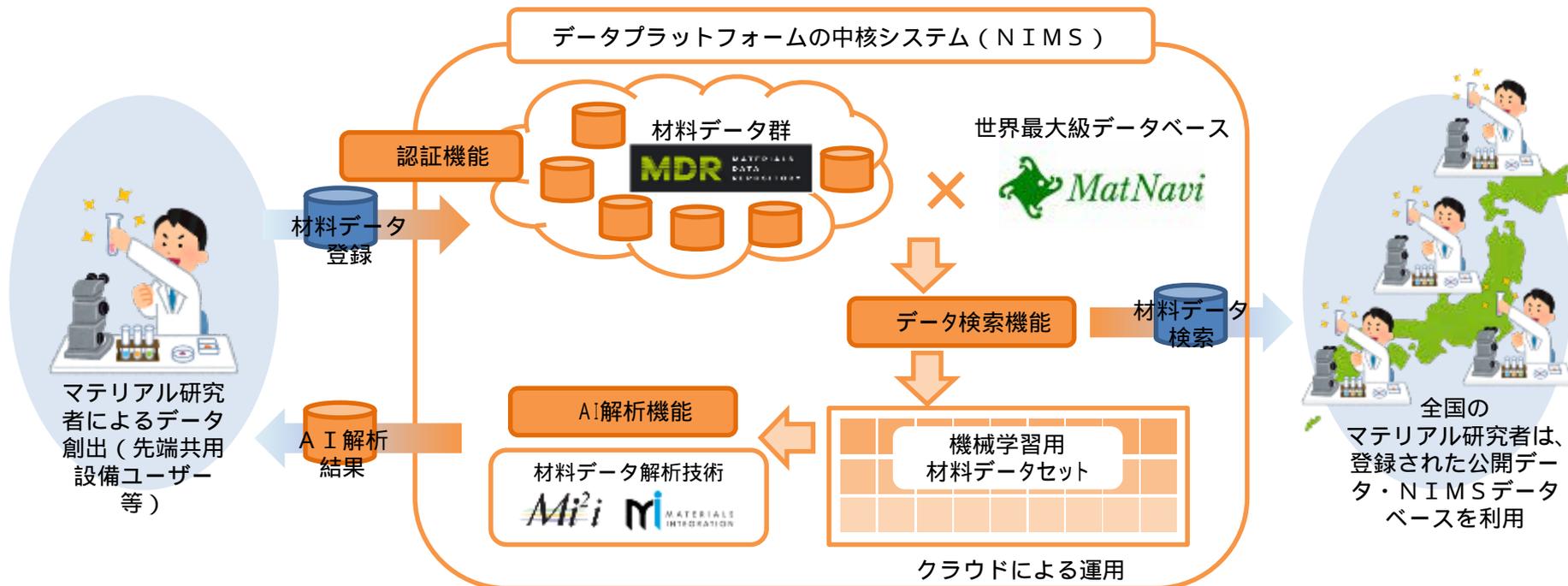
データを基軸とした研究開発の推進について

どのように全国でのデータ利活用を実現するか？

- 幅広いユーザーのデータ登録、共有を可能とする認証機能を有したシステム
- ・全国で創出されたデータにアクセスするための検索機能
- ・幅広い研究者層がデータ駆動型研究を可能とする解析ソフト、インターフェース



- ・2023年を目途に、全国の大学、国研の先端共用設備ユーザーを対象とし、認証システム、データ検索機能、解析機能を有するNIMSのデータ基盤を整備、試験運用を開始。
- ・2025年を目途に、共用設備ユーザーが本格利用開始。加えて、産学のシェア・クローズド向けのサービスも開始。



マテリアルDXプラットフォーム構想実現のための取組

令和3年度要求・要望額 11,506百万円
 (前年度予算額 2,458百万円)
 運営費交付金中の推計額含む



背景 ・ 課題

近年、マテリアル研究開発では、**データを活用した研究開発の効率化・高速化・高度化**と、これらを通じた**研究開発環境の魅力向上が重要**となっている。また、**新型コロナウイルス感染症の世界的流行に伴い**、データやAI、ロボットを活用した新たな研究開発手法や研究開発環境の本格導入の必要性が高まる中、マテリアルの研究開発現場や製造現場全体の**デジタル化・リモート化・スマート化**といった**デジタルトランスフォーメーション（DX）が急務**。我が国には、良質なマテリアルデータを生み出す**世界最高水準の共用施設・設備群、産学官の優れた人材が存在**するが、この強みを最大限に活用し、**産学官のデータを効果的に収集・蓄積・流通・利活用できる仕組み、データを持続的に創出・共用化できる仕組みは未整備**

産学官の高品質なマテリアルデータの戦略的な収集・蓄積・流通・利活用に加えて、**データが効率的・継続的に創出・共用化**されるための仕組みを持つ、**マテリアル研究開発のための我が国全体としてのプラットフォームを整備**

【統合イノベーション戦略2020(令和2年7月閣議決定)】

<データを基軸としたマテリアルDXプラットフォーム(仮称)の実現>
 ・マテリアルの研究開発力を大幅に強化する、我が国全体で高品質なマテリアルデータが持続的かつ効果的に創出、**共用化、蓄積、流通、利活用**される産学官のプラットフォームの実現に向けて、産学官の協力の下で構想・推進

【成長戦略フォローアップ(令和2年7月閣議決定)】

・「マテリアル革新力」を強化するため、以下の取組を含め検討し、政府戦略を策定する。
 - データ蓄積の中核拠点整備や、良質なデータを取得可能な**共用施設・設備の整備、データ創出・活用を牽引する研究開発プロジェクト**等について2020年度から検討を進め、速やかに実施する。

取組概要

共通的なデータ収集・蓄積・流通・利活用のための**基盤整備**を進めるとともに、**先端共用施設・設備**や**重要技術・実装領域**を対象とした**研究開発プロジェクトからデータ創出の取組**を行う

データの取り扱いについては、今後、経済産業省とともに**共通指針**を検討・策定
 本取組の総合的な進捗管理を行う**ガバナンスボード**を設置し、経済産業省等の事業と連携することを検討

データ中核拠点の形成

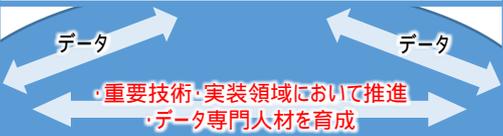
令和3年度要求・要望額 4,232百万円
 (前年度予算額 600百万円)
 運営費交付金中の推計額

- オープンデータ・シェアードデータを対象に、**セキュアな環境の下、データとデータ構造を蓄積・管理する中核拠点をNIMSに整備**



データ基盤

これまでNIMSにおいて進めてきた材料データ収集の高度化や、NIMSデータ公開基盤開発の成果を基盤とした、日本全国のマテリアルデータを集約するための**データ中核拠点を構築**



データ蓄積・利活用による論文生産や特許出願、人材育成等を通じた、産学連携の促進、研究成果の社会実装の加速

データ創出基盤の整備・高度化

令和3年度要求・要望額 4,093百万円
 (前年度予算額 1,553百万円)

- 技術支援により**先端的な施設・設備の全国共用**を行う、**ナノテクノロジープラットフォーム事業**を実施。さらに、多様な設備を持つ**ハブ**と特徴的な**技術・装置を持つスポーク**からなる**ハブ&スポーク体制**を新たに構築し、**高品質なデータとデータ構造の共用基盤を整備・高度化**

【データ共用基盤部分に係る事業内容】

- 対象機関：大学・独法等
- 事業期間：令和3年度～（10年）
- 支援規模：6ハブ、19スポーク程度
- 支援内容
 - データ対応型設備の整備
 - データ構造化等を行う
 - データ人材の確保



【データ共用基盤部分に係る事業スキーム】



データ創出・活用型プロジェクト

令和3年度要求・要望額 3,181百万円
 (前年度予算額 306百万円)
 運営費交付金中の推計額含む

- 重要技術領域において、データ創出・活用と理論・計算・実験が融合する、データ駆動型の研究開発プロジェクトを実施**

マテリアルサイエンス型

(新物質・新機能マテリアルの創出)
 令和3年度要求・要望額 120百万円 (新規)

【事業内容】

- 対象機関：大学・独法等
- 課題数：6課題程度
- 事業期間：令和3年度～（10年）
- 令和3年度：FS
- 令和4年度～：拠点形成・本格実施

プロセスサイエンス型

(新たな材料開発プロセスの創出)
 令和3年度要求・要望額 617百万円 (前年度予算額 306百万円)

【事業内容】

- 対象機関：大学・独法等
- 課題数：新規2課題程度(継続2課題)
- 事業期間：令和3年度～（7年）

【事業スキーム】



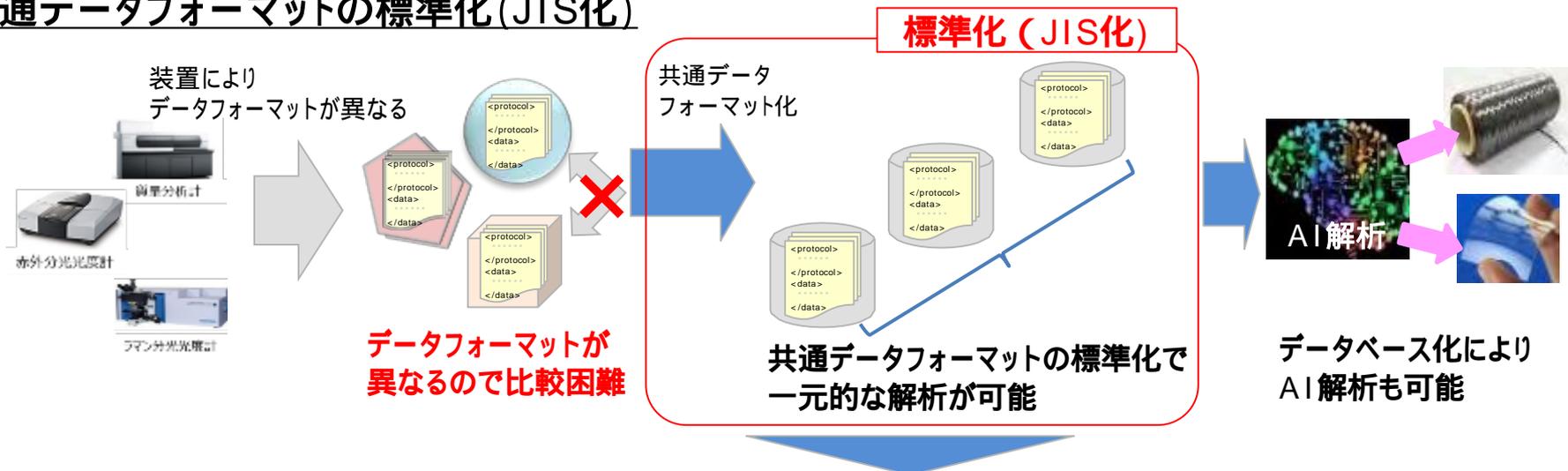
マテリアル革新力強化に向けた**基礎基盤研究(NIMS事業)** 令和3年度要求・要望額 2,444百万円(新規) 運営費交付金中の推計額

マテリアルとの融合が大きな付加価値をもたらす量子、バイオ、AI、国土強靱化分野において、データを創出・蓄積しつつ、それらを活用した研究開発を実施

(参考) 計測分析装置の共通データフォーマットの国際標準化

- 共通データフォーマットを標準化(JIS化)することで、AI解析による新材料の開発や生産性が飛躍的に向上するなど、MIの取り組みの加速化が期待。
- 先進材料の標準化に関する国際的な枠組み(VAMAS)を活用し、ISO化に取り組み日本発の共通データフォーマットの国際標準化を目指す

共通データフォーマットの標準化(JIS化)



共通データフォーマットの国際標準化(ISO化)を図り、国際的なMI基盤を整備



新材料及び標準に関するベルサイユプロジェクト(VAMAS)の枠組みも活用

日本発の共通データフォーマットの国際標準化へ