



AI等の活用を推進する研究データエコシステム構築事業



背景・課題

- ・ポストコロナの原動力として「デジタル」「AI」が最重要視されているが、AI・データ駆動型研究開発に必要な大規模かつ高品質なデータの利活用を推進していくことが鍵である。このため、全国に散逸する研究データをつなぎ、必要なデータを利活用できる環境を整備することが必要。

未解決の主な課題

- ✓ 分野・機関を越えた全国的な研究データ共有・利活用の基盤が未実装
- ✓ 各大学のデータマネジメント体制やルール構築が進んでいない（研究データマネジメントポリシーを策定している大学は国立大学でも全体の2割程度かつ必ずしも全学的な内容になっていない）
- ✓ DXによる研究手法の変革、AI・データ駆動型研究の進展が不十分

【経済財政運営と改革の基本方針2022】（令和4年6月7日閣議決定）

- ・国際性向上や人材の円滑な移動の促進、大型研究施設の官民共同の仕組み等による戦略的な整備・活用の推進、情報インフラの活用を含む研究DXの推進、各種研究開発事業における国際共同研究の推進等により、研究の質及び生産性の向上を目指す。

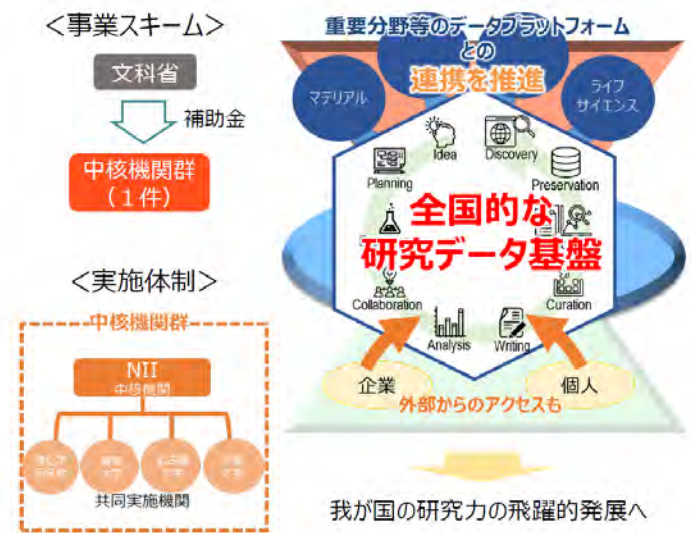
【新しい資本主義のグランドデザイン及び実行計画】（令和4年6月7日閣議決定）

- ・大学等や国の機関が保有するデータは、それぞれの機関に分散し、データ形式もバラバラとなっているが、他のデータ基盤との接続を可能とし、民間企業等の利活用を進める。

必要な取組 事業期間：R4年度～R8年度

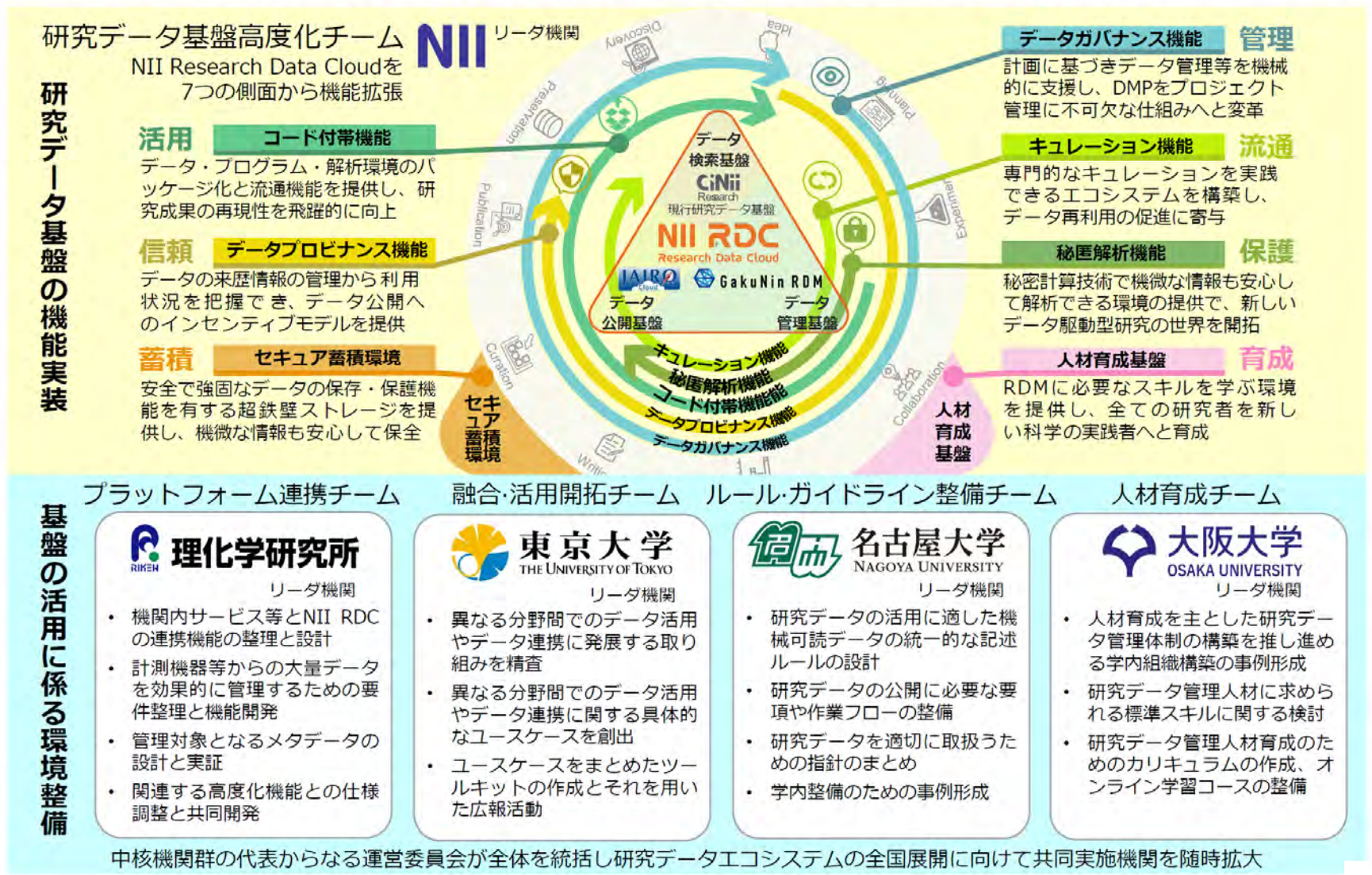
①各分野・機関の研究データをつなぐ**全国的な研究データ基盤の構築・高度化・実装**と、②**AI解析等の研究データ基盤の構築・活用に資する環境の整備**を行う、研究DXの中核機関群を支援する。

- ①**全国的研究データ基盤の構築・高度化・実装**
- 以下の機能を有した全国的な研究データ基盤の整備を推進
- ・管理データの取捨選択やメタデータ付与、データの出所・修正履歴の管理など、研究データ管理にかかる作業を効率的に実施するための機能
 - ・秘匿すべき情報が含まれる研究データを適切・安全に解析可能とする環境構築を支援する機能
 - ・各機関リポジトリや各分野のデータプラットフォームとの連携・接続による分野・機関を越えた研究データの検索機能（**拡充**）等
- ②**研究データ基盤の構築・活用に係る環境の整備**
- ・効率的なAI活用のための、機械可読データの統一化や標準化等を含めたルール・ガイドライン整備、データマネジメント人材育成支援
 - ・大学における研究データマネジメントにかかる体制・ルール整備の支援（**新規**）
 - ・異分野等の多次元データ解析など最先端の数理・情報科学と分野の融合によるAI・データ駆動型の新たな研究開発手法の開拓（**拡充**）等



研究DX
研究データの
管理・利活用

AI等の活用を推進する研究データエコシステム構築事業





研究データエコシステムの構築に向けた具体的な取組イメージ（現行）



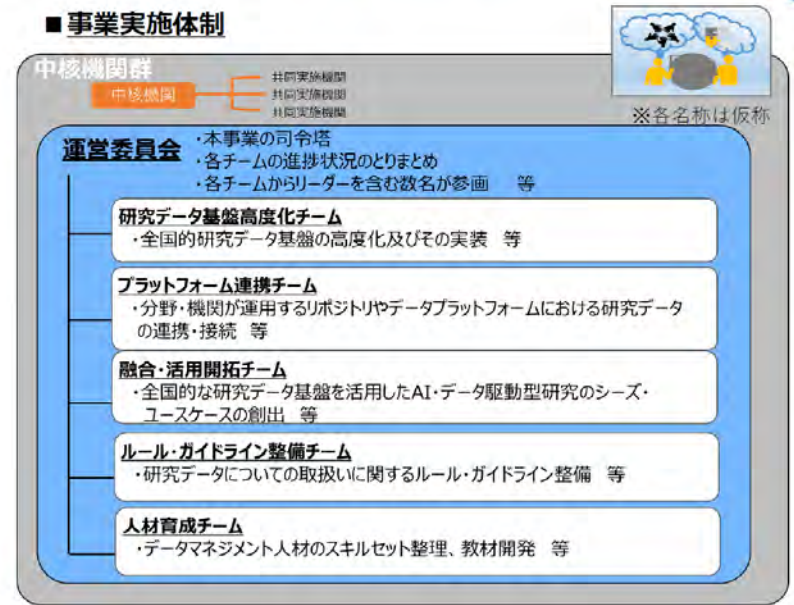
中間まとめ

■全国的な研究データ基盤

全国的な研究データ基盤を中心に、オープン・アンド・クローズ戦略に基づいた研究データの管理・利活用が持続的に行われる仕組み（研究データエコシステム）の構築を目指す。そのために、全国的な研究データ基盤の構築・高度化・実装等や研究データ基盤の活用に係る環境整備を実施



■事業実施体制



中核機関群

・研究データ基盤の構築・高度化・実装の中心的役割を担う中核機関が、複数の共同実施機関と連携した体制
・中核機関群には、本事業の司令塔となる運営委員会及びそれぞれ推進すべき事業の内容に基づき中核機関群による機関から構成したチームを構成して事業推進

<各チームのミッション>

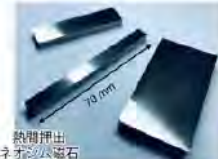
- ①研究データ基盤高度化チーム
・全国的な研究データ基盤の高度化及びその実装
【最低限求められる機能】
- データマネジメントプラン（DMP）の内容に基づく研究データの取扱い（管理対象データの取捨選択やメタデータの付与等）を支援する機能
- 研究データの出所や修正履歴等を管理し、対外的にその真正性を説明可能とすることを支援する機能
- 研究者が管理する研究データを共有・公開した際、その研究データの利用状況や公開後の再利用の状況をモニタリングし、研究者にフィードバックする機能
- 研究者が管理する研究データ・プログラム・解析環境等の情報を、再現性の検証や再利用可能な形でパッケージ化し公開することを支援する機能
- 秘匿すべき情報が含まれる研究データについて、その管理をシステム上で適切に行うとともに、これらの研究データを秘匿したまま安全に解析可能とするための環境構築を支援する機能
- ②プラットフォーム連携チーム
・分野・機関が運用するリポジトリやデータプラットフォームにおける研究データの連携・接続
・全国的な研究データ基盤上で統合的なメタデータ検索の実現
・全国的な研究データ基盤との接続を希望する産学官の機関等に提示する要件等の整備
- ③融合・活用開拓チーム
・全国的な研究データ基盤を活用し、異なる分野間でのデータ連携を前提としたAI・データ駆動型研究のシーズ・ユースケースの創出
・全国的な研究データ基盤のユーザー拡大に向けた周知・広報活動
- ④ルール・ガイドライン整備チーム
・研究データの効果的・効率的な活用のための機械可読データの統一化や標準化等や個人情報等の研究データを取り扱う上で留意すべきデータを含む研究データについての取扱いに関するルール・ガイドラインの整備
- ⑤人材育成チーム
・研究データ管理に必要なデータマネジメント人材に必要なスキルセット等の要件の整理
・データマネジメント人材の育成に向けた教材開発やコンテンツ整備等



研究DXの先行的な成果事例

**AIが約6600万通りから最適な作製条件を探索
→ AI解析前後で材料性能が1.5倍に向上**

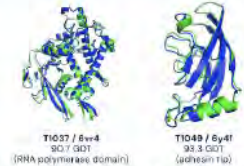
- 世界の電力消費量の約50%を占めるモーターの効率化・省エネ化のために、永久磁石の高性能化が期待。
- ネオジム磁石の作製プロセス条件は約6600万通りであり、網羅的な探索が不可能。AI解析により提案された最適な作製条件に基づいて実験計画を立てることで、わずか40回程度の実験で**ネオジム磁石の強さをAI解析前の約1.5倍向上**させることに成功。
(2021.11.15 プレスリリース)



実証実験で表示される「3D 雨雲ウォッチ」アプリイメージ

**AIによるタンパク質の立体構造解析ツール
→ 数年かかっていた作業を圧倒的に短縮**

- 2014年にGoogleが約670億円で買収したAI開発企業DeepMindが無償公開した遺伝子配列情報からタンパク質の立体構造を解析するAI「**AlphaFold2**」。
- 従来、タンパク質のアミノ酸配列の構造を特定するには数か月～数年かかり、多大な時間と費用が掛かっていたが、**AlphaFold2では、限られた情報から構造と機能を推定することが可能に。**



実験結果とシミュレーションの結果がほぼ類似

**仮説・実験・検証・修正のサイクルを自動化
→ これまで着手が困難だった希少疾患等の創薬開発が推進**

- ケンブリッジ大学とマンチェスター大学により開発されたロボットサイエンティスト「イブ」は、一日当たり1万化合物のスクリーニングを行うとともに、**実験結果を統計学・機械学習により解析し、より高い活性をもつ新たな化合物の構造を予測し、実験・検証・修正のサイクルを繰り返す、仮説主導のハイスループット研究を自動化した。**
- このような手法を導入することにより、製薬企業が手を出しにくい**希少疾患等に対する創薬開発が推進**されることが期待。

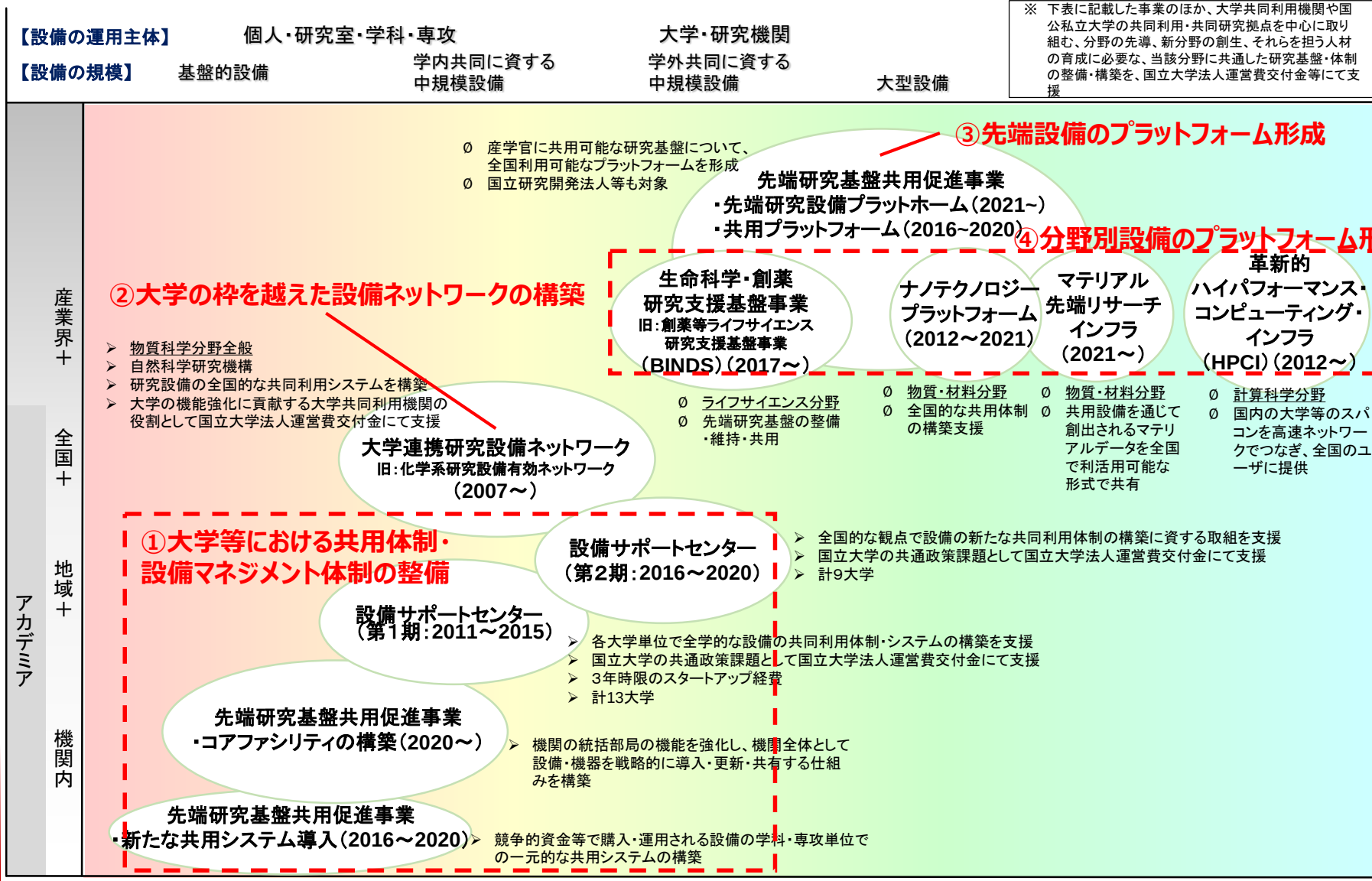


Robot scientist Eve

「AI」、「データ」、「リモート化・スマート化」による圧倒的な生産性の向上とハイインパクトな成果が生まれ始めている。
しかし、まだ一部の研究者・研究領域のみ。この変革の動きを、日本全体に発展させることが必要。



【共用の範囲イメージ】



⑤ 共同利用・共同研究体制の構築

中間まとめ

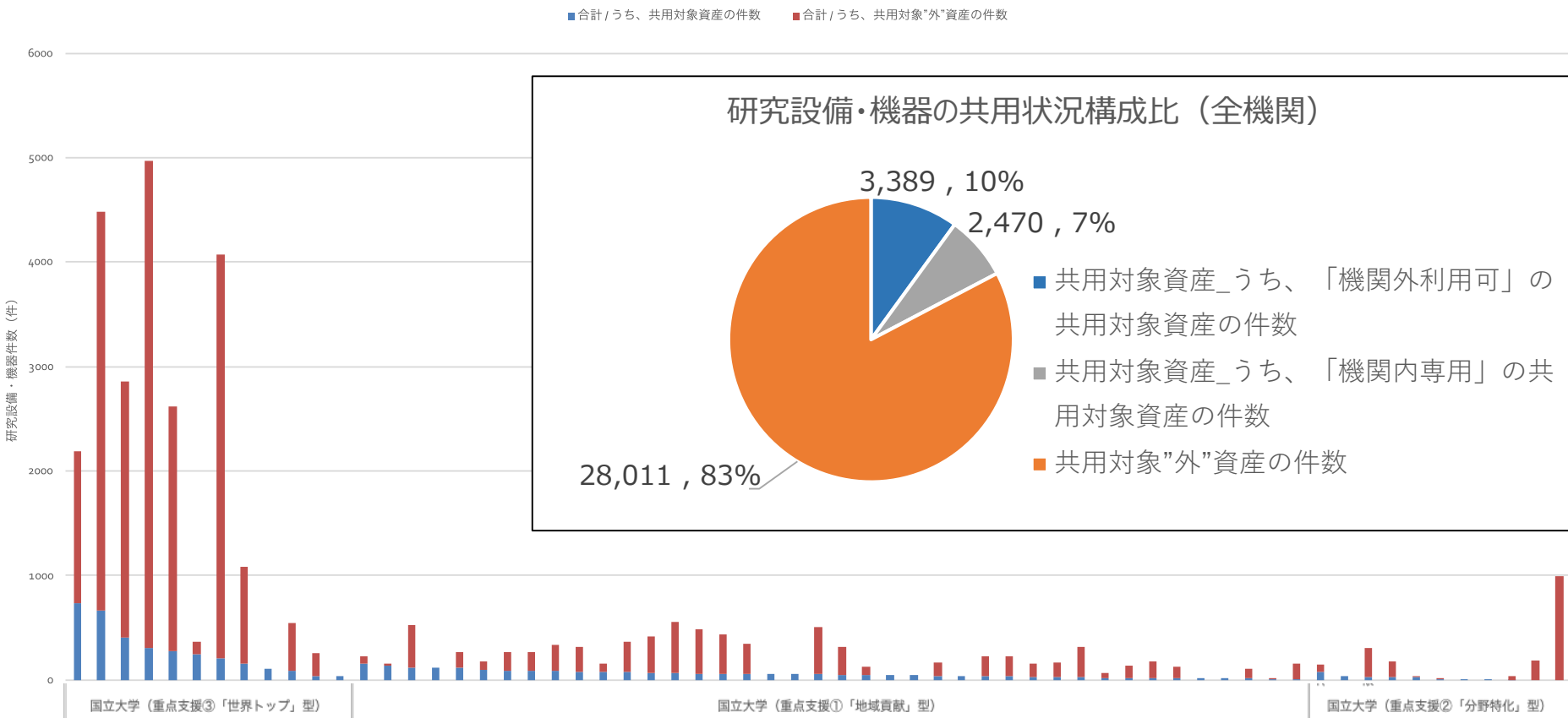
【研究設備・機器の規模イメージ】



研究設備・機器の共用件数（2020）：国大

大学類型別の共用対象資産の件数（降順）

中間まとめ



※機関によっては把握している一部の回答、共用対象資産の件数のみの回答となっている点に留意。

「共用化対象」の資産は一定のデータを収集できている（国大65/対象70機関）。共用化対象の資産は全体の約17%。ただし5機関は共用対象資産の回答なし。また12機関より、部局管理設備の共用状態は不明等、大学全体としては設備の共用状況を把握しきれていない状況あり。