

理研での情報系研究と 研究DX

2021年6月30日

理化学研究所
美濃 導彦

理化学研究所概要

- 大正6年(1917年)創立、日本初の自然科学の総合研究機関
- 日本最高、世界トップクラスの研究の質を誇る
- 研究センター方式で、国の政策に沿った戦略的研究開発を集中的に実施
- 最高水準の研究基盤の構築と、研究コミュニティへの利用機会の提供
- 体制・人員・予算

拠点：国内10拠点・海外5拠点(英国、米国、中国、シンガポール、ベルギー)

職員数：3,542名

研究系3,017名、事務系525名 *2019年4月1日時点

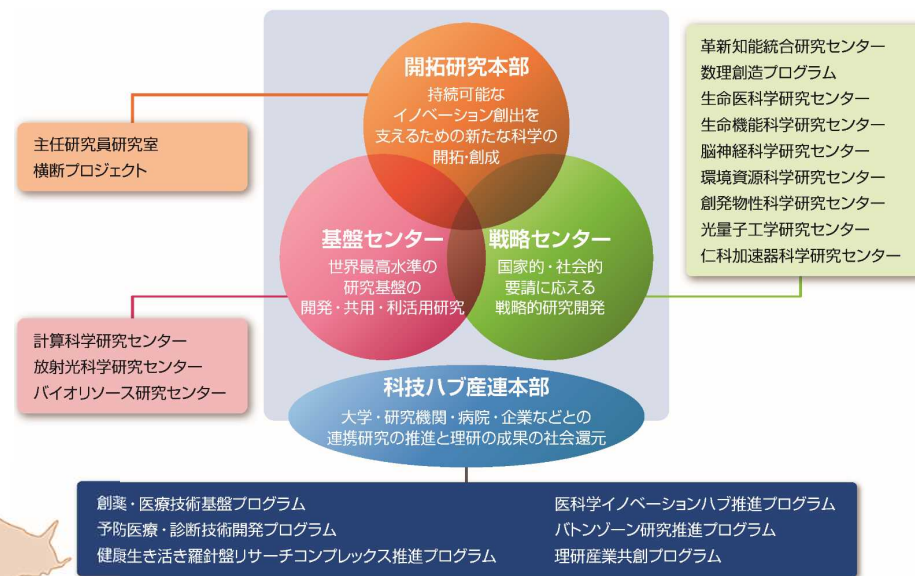
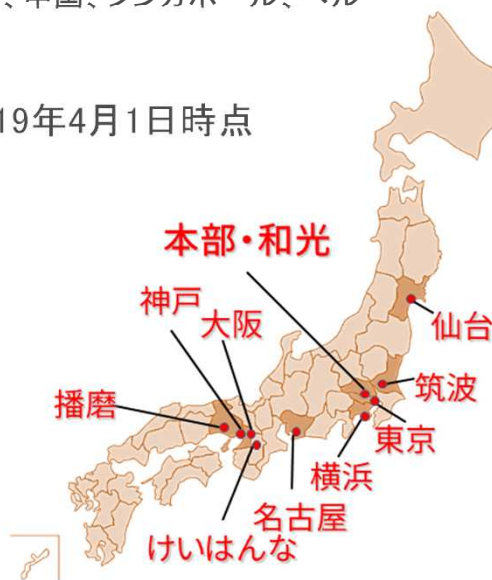
事業費：988億円(2019年度予算額)

運営費交付金:531億

AIP補助金:31億

SP-8/SACLA補助金:83/55億

富岳等補助金145億 他



理研 4つの強み

1. 大学には無い、大型施設の開発と運用・共用
2. 一研究室規模が大きい
長期・大型プロジェクトを計画、実行できる
3. 研究支援（技術者等、研究推進事務）が充実
4. 新領域、学際領域の開拓の容易性。
研究室間の垣根が低い、日常的な研究議論

所内情報研究のハブ機能・情報システム基盤/利用支援の提供

研究データ基盤の構築等による情報環境の強化

- ・オープンサイエンスを見据えた**研究データ共有・公開基盤の構築**のほか、データ利活用支援等に関する**研究開発や人的交流のための体制**を構築。
- ・集約されたデータやオープンデータを基にした**データ駆動型研究**の推進。
- ・**脳 x AI x ころ**の要素を取り入れた人の心に寄り添ったロボットを実現。

AI研究

革新知能統合研究センター (AIP)

HPC研究

計算科学研究センター (R-CCS)

数理科学

数理創造プログラム (iTHEMS)

量子力学

量子コンピュータ研究センター (RQC)

エンジニアリング

科技ハブ産連本部 (RCSTI)
光量子工学研究センター (RAP)

バイオリソース事業のメタデータ整備

バイオリソース研究センター (BRC)

生命科学系公開データベース

生命機能科学研究センター (BDR)
生命科学研究センター (IMS)
脳神経科学研究センター (CBS)

メタボロームデータ

環境資源科学研究センター (CSRS)

先端データサイエンス

情報科学と生命科学を融合した新たな枠組みを構築し、社会問題の解決と理研内のハブ機能を担う。

ガーディアンロボット

情報学を中心に脳科学、心理学、機械工学等を統合した研究をすすめ理研内のハブ機能を担う。

データ駆動型
研究の推進

基盤研究開発

データマネジメントシステムによるオープン/クローズ戦略や分野特性に合わせたメタデータ付与によるオープンサイエンスを推進する。

情報システム基盤・研究基盤・情報サービス・情報環境の提供

情報セキュリティの強化

情報システム部

令和3年度からの理化学研究所研究推進体制図

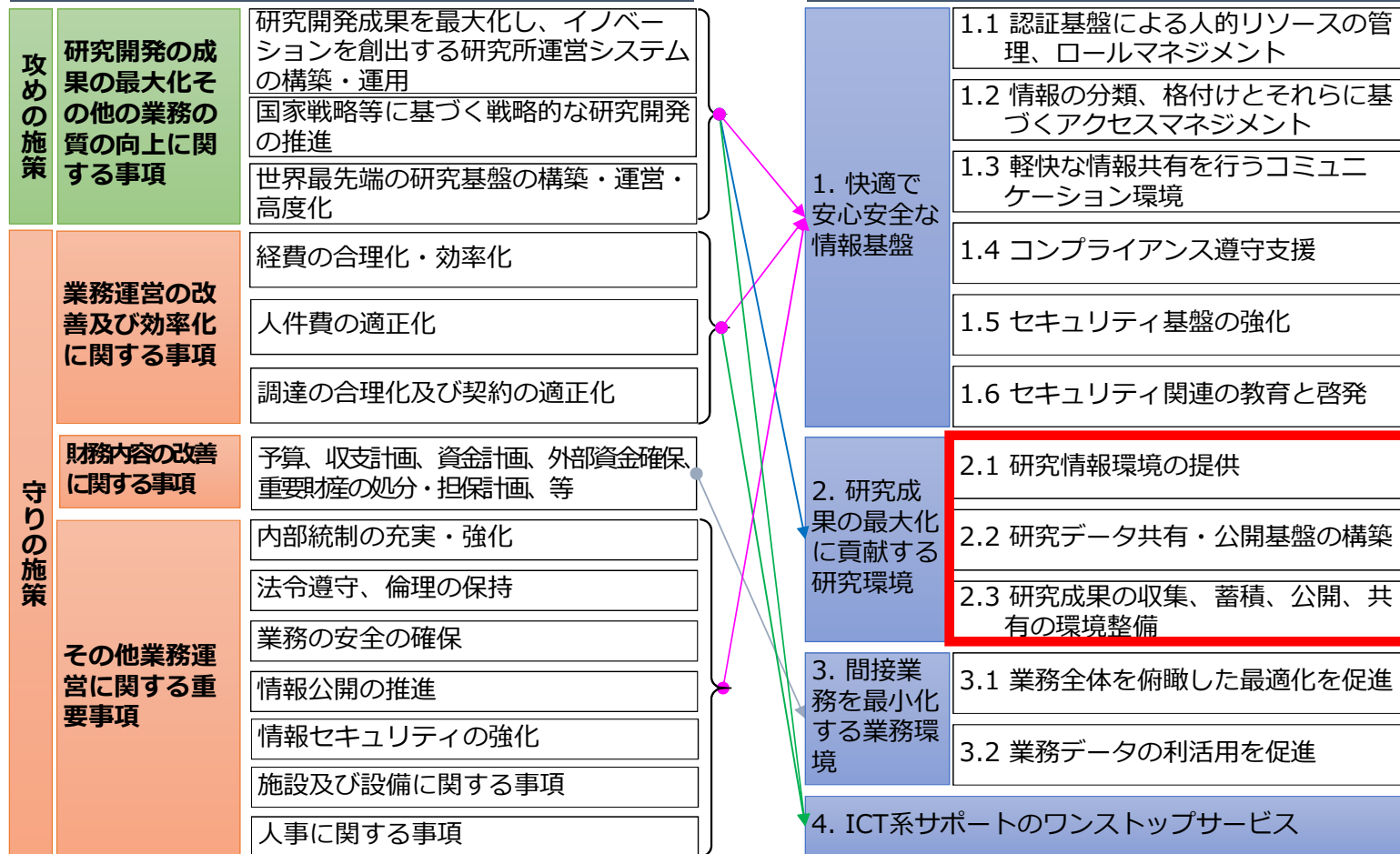


研究データの公開
ーオープンサイエンス事業ー

ICT戦略 概要

あるべき姿（中長期目標）

打ち手（第4期中長期期間のICT戦略）



中長期目標の達成

オープンサイエンス事業

オープンする研究データの種類

1. 研究者が公開する研究データ
 - プロジェクトとして推進するもの
 - FANTOM、革新脳プロジェクトによるデータなど
 - 研究者が個別に公開するもの
 - データ共有ユニット・協議会で分野ごとのコミュニティを作り推進
2. 研究者が義務として公開するデータ
 - 研究証拠データ（研究成果・業績として登録）

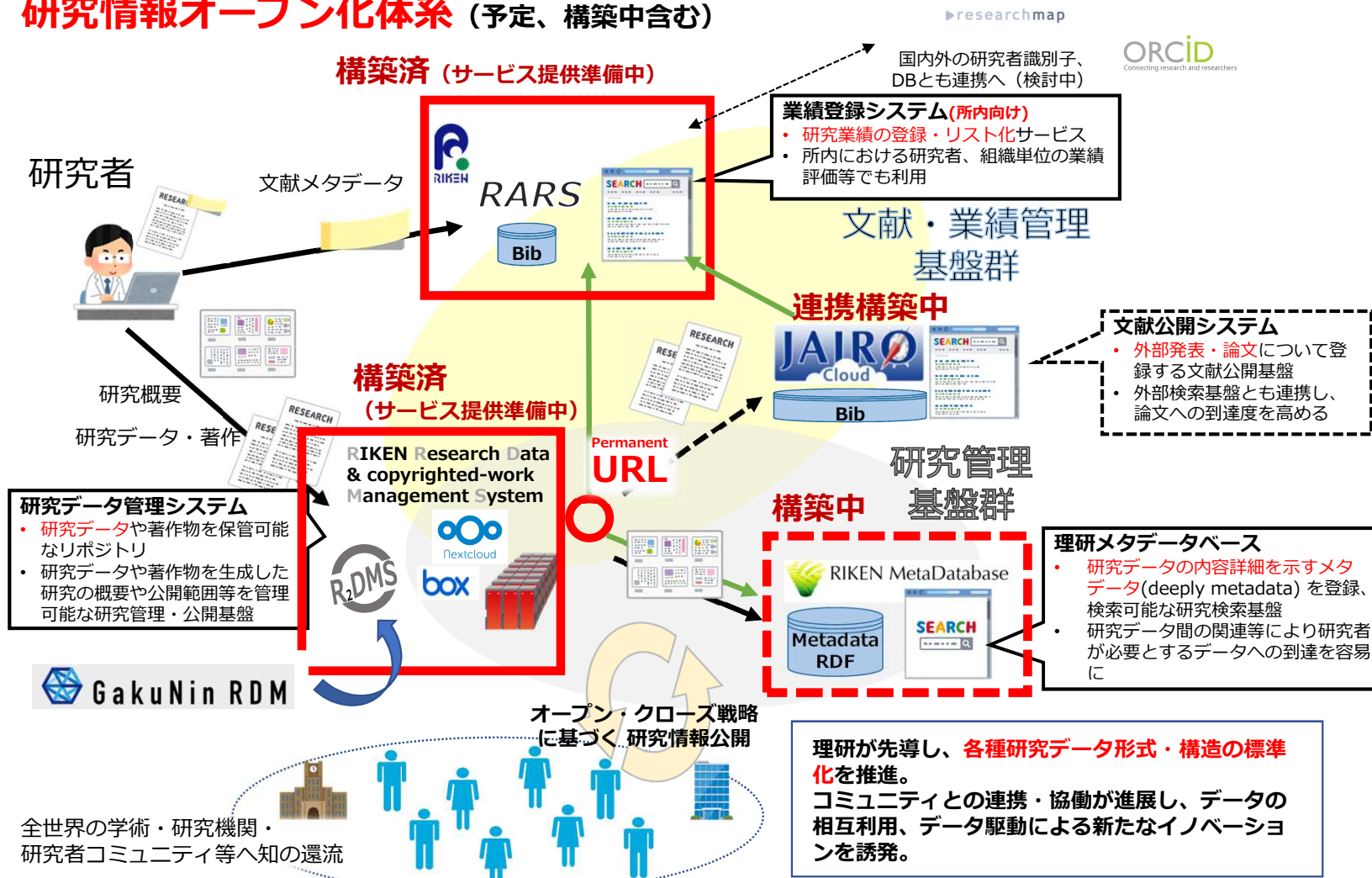
オープンサイエンス事業に関する実績

- 研究データの公開及び組織・分野横断的な利活用を目指し、収集・管理・共有・公開を一体的に行うための研究情報管理システムを含む管理基盤を構築
- 研究分野ごとの特性に合わせた組織（生命科学ユニット、医科学ユニット）を新規設置し、研究データを集積、公開を推進
 - 先行して生命科学分野の研究データの集積。データの二次利用のためのメタデータ付与や標準化について研究開発を推進
- 理研データポリシーを踏まえた研究センター等毎の詳細な研究データガイドラインを策定・公開（2021年3月）

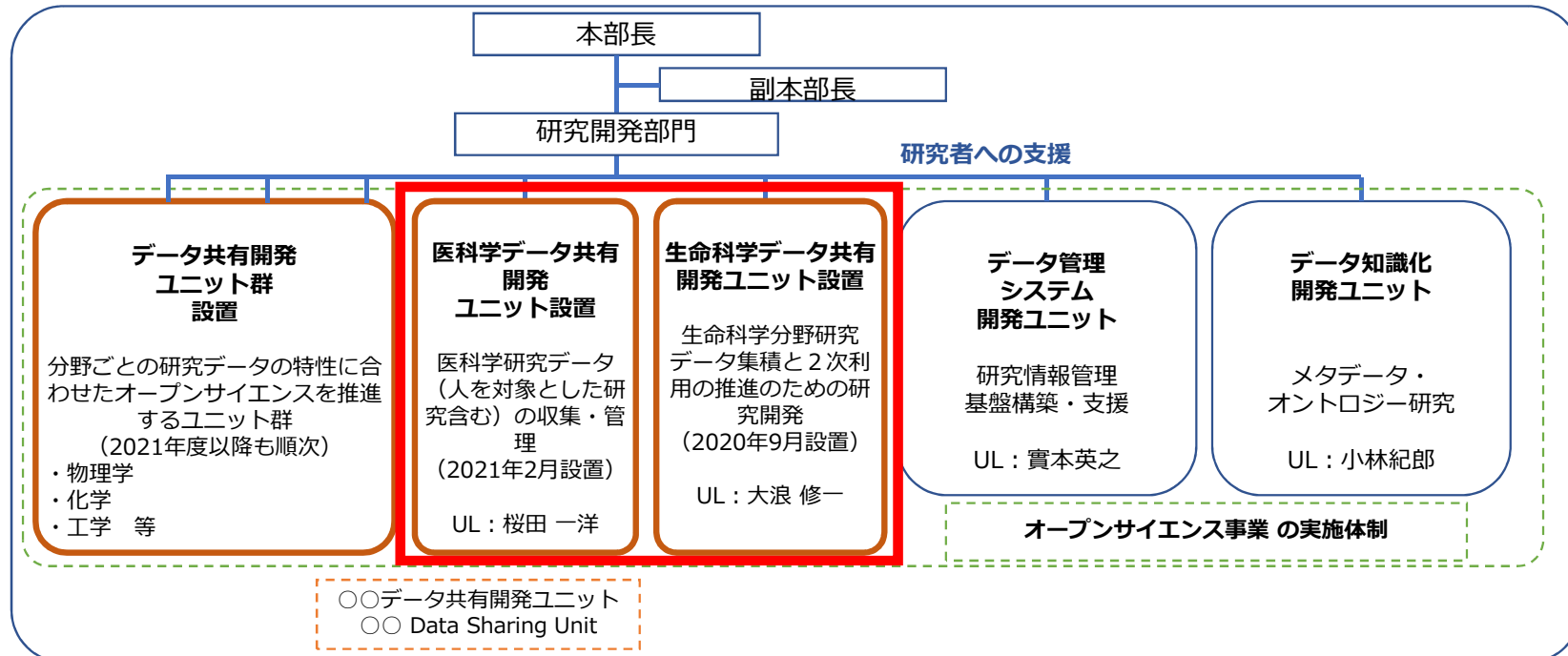
オープン化に向けた 研究情報管理基盤の整備・連携準備



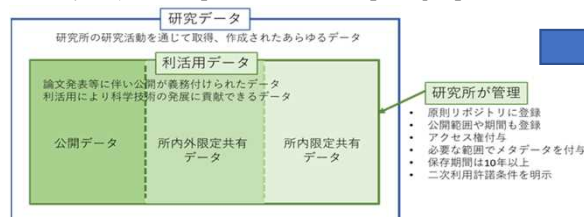
研究情報オープン化体系 (予定、構築中含む)



分野ごとの研究データオープン化に向けた 体制・制度の整備



利活用データ ガイドライン (詳細)



理研データポリシー (2020年3月策定) で定める利活用データ

分野に即したセンターごとのガイドライン策定によりその詳細を策定

研究データ・利活用データに関する詳細 (データ生成者・データ管理者の役割、公開・共有フラグ設定やデータ削除の手順、データ公開猶予期間、利用許諾等)

オープンサイエンス事業 R2年度活動実績

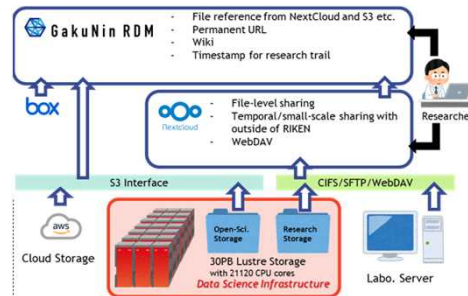
研究情報管理システムの構築

研究データと著作物の保管・共有・公開をプロジェクト単位で一気通貫で管理するシステムを、オープンサイエンス推進の基盤として構築。

研究情報管理システム

特徴

- ・ 認証システムによるセキュアなアクセス権限の管理
- ・ 様々なストレージが利用可能な汎用的な保存基盤
- ・ 大規模ストレージ連携による種々のビッグサイエンスにも対応



国立情報学研究所との連携・協力協定締結

唯一の情報学の学術総合研究所であり、研究データ基盤を構築するNIIと様々な分野を有する理研との連携により、オープンサイエンスに資する学術研究分野のインフラを強化・発展させる体制を構築。



2021年4月13日 協定締結式

- ・ オープンサイエンスの推進に向けた戦略的組織間連携のベストプラクティスを推進
- ・ データプラットフォームの開発および運用、オープンサイエンスの推進、データ駆動型研究の推進
- ・ 多くの研究分野や研究機関のモデルとなり、日本全体の研究環境の改善と研究力の強化につなげる

バイオイメージデータのデータ形式標準化とリポジトリ整備のための国際提言

- ・ バイオイメージデータの世界標準について、欧州、米国、日本の11カ国からバイオイメージデータの専門家が集い、議論、提言。
 - ・ 生物学・医学研究で用いられるイメージングデータのデータ形式の標準化およびデータ共有リポジトリの整備により、研究者に留まらず、ツール開発者が膨大なバイオイメージデータを共通に取扱うことが可能。

(2021年5月 *Nature Methods*,)
BDR発生動態研究チーム 大浪修一TLら

理研オープンサイエンスポータルを設置



理研内のオープンサイエンスに関わる様々な情報を一元的に発信

- 理研における公開データカタログ データセット
- データポリシーおよびセンターごとのガイドライン
- 活動組織やプロジェクト、イベントなど

統合を目指す公開DB

BDR

- SSBD:database
- SSBD:repository
- 1細胞遺伝子発現データ
- Reptiliomix
- CVG
- eLeaves
- Squalomix
- gVolante
- Softshell turtle genome assembly
- 理研微細構造画像メタデータベース (RIKEN MIM)

BRC

- バイオリソースカタログデータ
- MoG+
- コンソミックマウス表現型データベース
- マウス表現型間（機能間）関係性データベース
- PhenoPub
- 持続可能な循環型社会を実現する「農業環境エンジニアリングシステム」の開発のメタデータ
- IMPCデータ

IMS

- FANTOM web resource
- ZENBU
- ZENBU-reports
- scPortalen
- refTSS

CSRS

- PRIMEデータベース

RAP

- ヒト人体力学特性データベース
- 人体形状モデルデータ
- 細胞画像データ (3D, 4D)

CBS

- 革新脳データポータル
- Marmoset Gene Atlas
- Neuro Tycho
- INCF日本ノードポータル
- Neuroimaging PF
- Visioime PF
- Dynamic Brain PF
- Cerebellar PF
- CBSN PF
- Invertebrate Brain PF
- BSI-NI
- Brain Atlas
- CelLoc-3D

研究データ公開の今後の展開

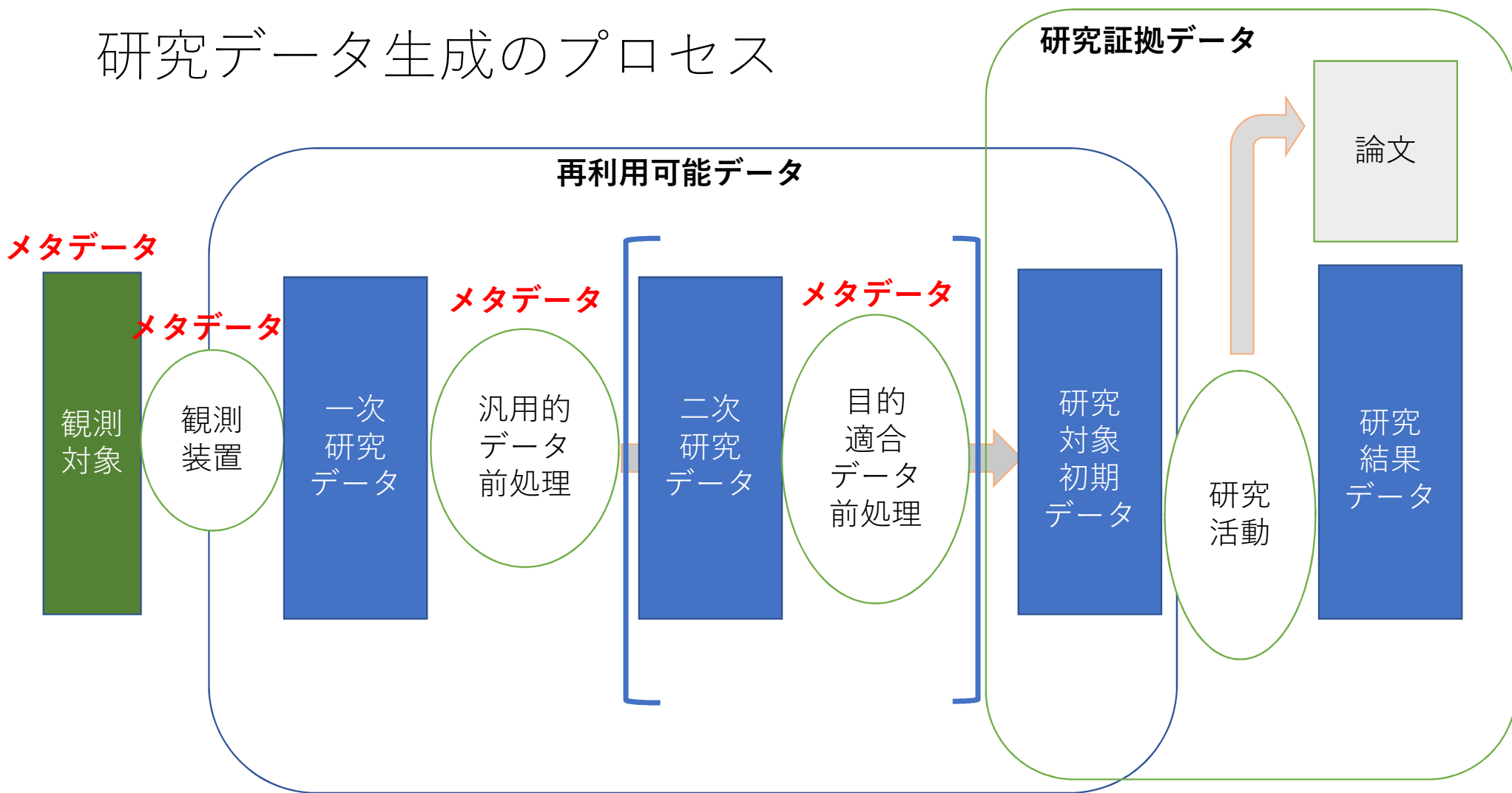
研究データの作成・蓄積・利用のプロセス

- 研究データの作成・収集（機関の研究者が行う）
 - 研究者が研究のためにデータを作成
 - 他の人が利用できるようにデータを整理、メタデータを付与（インセンティブが必須）
 - 整理したデータをリポジトリに入れる（公開期限やデータ管理の方法を指示）
- 研究データの蓄積・公開（機関のIT部門）
 - リポジトリ（長い期間、安定的）の提供
 - データの維持管理（オープンクローズ戦略、データ利活用期間の設定）
 - 利活用の促進、利用上の質問対応
- 研究データの利用（一般の研究者）
 - メタデータによる検索
 - データ駆動型研究の推進
- 研究利用者のために作成、蓄積、公開する？
 - データ作成者のインセンティブをどう付与するかが重要
 - 理研ではデータ生成活動、オープン化のKPIの議論を今年度から開始

研究データの作成・収集

- 研究データはできるだけ発生した現場で管理
 - 大量データを動かすのはコストも時間もかかる
- 研究者は**仮説検証方式**で研究を行う場合がほとんど
 - 仮説を立ててデータを集める
 - 研究証拠データは仮説検証のためのものが多い
 - 論文を書くので、そこにデータの説明がある
 - 他の研究に役立つ可能性は高くない
- データ取得そのものは多くの場合**研究支援者、技術者、学生**が行う
 - 研究データはできるだけ一次データであることが望ましい
 - 観測装置がメタデータを付与できる仕組み
 - この段階での研究データをリポジトリへ入れてはどうか
 - 一次データは加工しないと使い難いが特殊な加工をすると汎用性がなくなる
 - データを加工する人、プログラムがメタデータをつけてリポジトリへ入れる
 - その後、研究者が研究データを利用

研究データ生成のプロセス

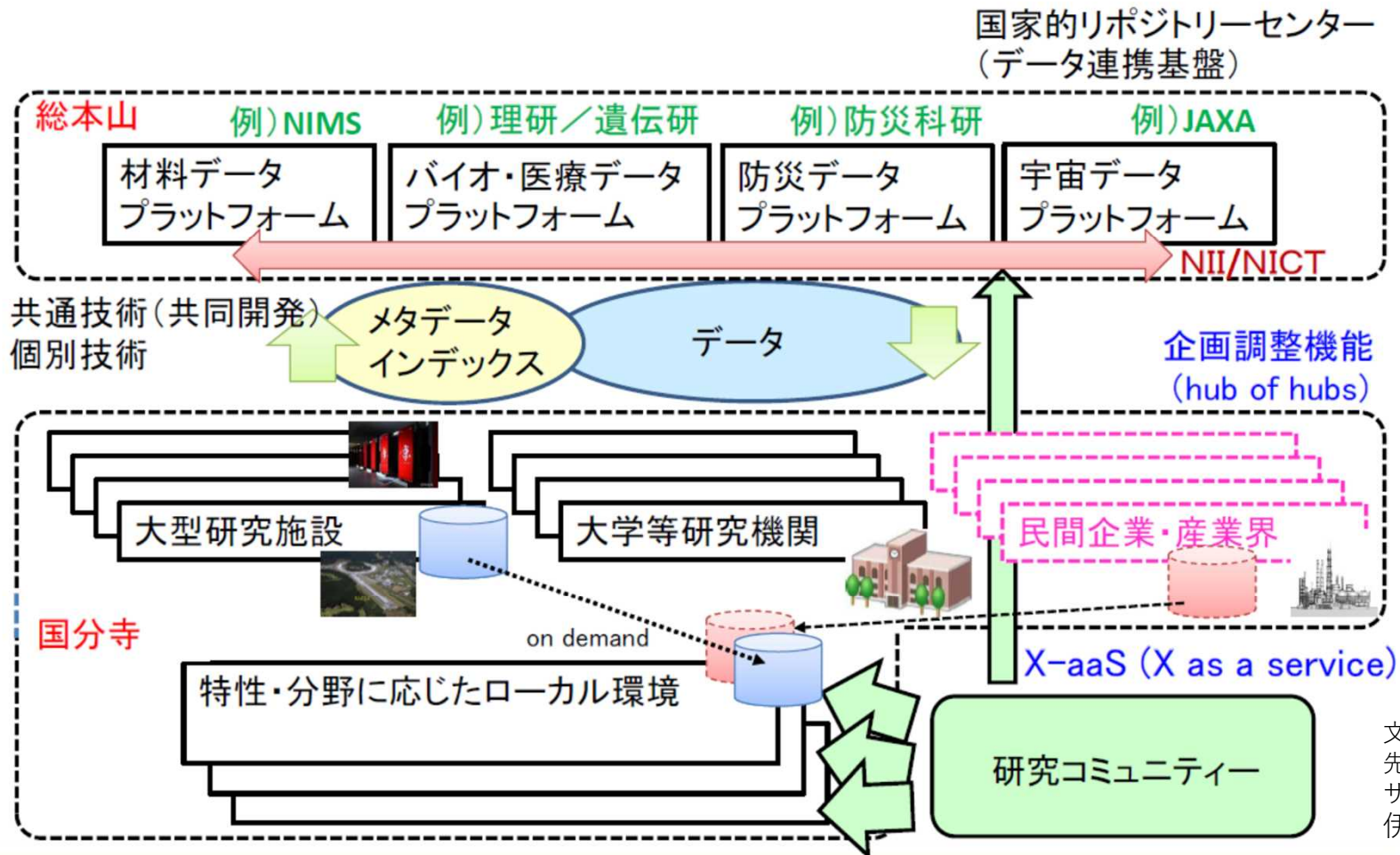


研究データの蓄積・公開

- 研究データは発生した場所の近くで蓄積
 - 大規模な大学、国研等の研究機関が責任をもってデータを維持管理
 - 研究室任せではデータが散逸する
 - 中小規模の大学から出るデータはNIIなどが引き受ける必要がある
- 研究分野の中核機関（データを多く持っている機関）がデータのフォーマットや取り扱いの標準化を推進
 - 他の機関は中核機関の下、分散型でデータ管理を実現
 - 中核機関が日本を代表して国際的に連携
 - 理研は生命科学の分野で中核を目指す
- 中核機関はデータの利活用を促進する
 - 対象分野の研究者等がデータを利用した研究を進められるよう支援する

持続可能な科学技術データ連携・流通エコシステム

Mi²i



文科省マテリアル
先端リサーチインフラ事業
サブプログラムディレクター
伊藤聡氏より提供

産官学連携共同活動の推進

マテリアルズインフォマティクス学理の構築

研究データの利活用

- メタデータを利用したデータの検索
 - 世界規模で連携してゆく必要がある
 - 研究データに利用方法等が記述されていることが望ましい
- オープンサイエンス
 - データを作成している研究者によるオープンデータの活用は容易
 - データを作成しない研究者がデータ駆動型研究を進められる必要がある
- シチズンサイエンス
 - 高校生や一般市民が分かりやす形でデータを利用する
 - 研究データの意味をわかりやすく説明する作業が必要
 - 研究成果や研究プロセスを社会に広く説明するために有用

研究データ共有の大きな壁

- 個人情報の問題
 - 多くの分野で人に関する情報を扱う
 - 社会学的な研究は人に関する情報が必須
 - 医学的な研究は人に関する情報が必須
- 統計データから特定個人同定を全く不可能にする方法は存在しない
 - 特異データの存在するので、完全オープン化は無理
 - 匿名化でノイズを入れると科学的手法と相反する
 - 仮名化したうえでコミュニティオープン等のアイデアが必要
- 理研ではまず、自然科学を対象に進める
 - 物理学、工学、生命科学（人は除く）
 - 今年度から人を対象にした医科学データの議論を始める予定

研究DX



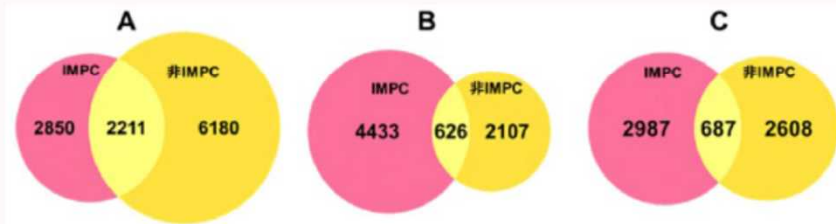
吉木淳¹ 綾部信也¹ 田村勝² (1 実験動物開発室, 2 マウス表現型解析開発チーム)

5,061遺伝子のノックアウトマウスを整備 - 国際連携 (IMPC) によるマウスリソースの拡充 -

成果発表 Nature Genet (2021)

概要・世界的位置づけ

ヒト遺伝子の機能や疾患における役割を解明するため、5,000を超える遺伝子をノックアウト (KO) した高品質なマウス系統を世界各国のマウスリソース機関と連携して作出し、マウス系統を公開・提供。合わせて、KOマウス系統の国際標準となる表現型情報を公開し、生命科学の研究基盤を拡充。



- A: 2,850遺伝子は世界初のKO系統。
- B: 遺伝子発現を可視化するレポーター系統が3倍に拡充。
- C: 組織特異的な機能解析が可能なコンディショナルKO系統が2倍に拡充。

成果の意義・社会貢献への期待

遺伝性希少疾患や未診断疾患の発症に関わる原因遺伝子の特定や病態メカニズムの解明、さらに創薬や治療法の開発に貢献すると期待。さらに、研究リソースの完全公開により、生命科学分野でのオープンサイエンスを先導。



機械可読の共通言語によりヒト疾患とマウスの
遺伝子型・表現型の情報を対応付ける

オープンサイエンスの起源

- 生命科学分野が先導
 - 世界中が協力して一つの問題（ヒトゲノム解析）の解決を目指す
 - このために解析したデータ共有
 - 国際会議を開催し、どのチームが次にどの部分を解析するかを議論
- サイエンスの研究は世界が協力して効率化を進める
 - 同じ研究を異なる国でできるだけ行わないようにしたい
 - 研究結果のデータを世界的に共有
 - 自律分散的に同じ研究を行わない枠組みを構築
- オープンサイエンス
 - データの公開とそのデータを利用した研究（研究の分業化）
 - データを取得し公開する研究者とオープンデータを利用して研究する研究者に分業が進む？
 - プラエが観測データを残し、ケプラーがそのデータをもとにケプラーの法則を発見



米倉 功治

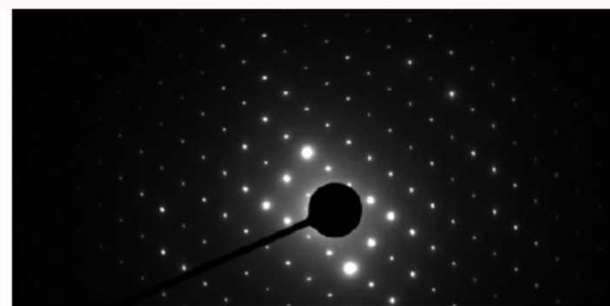
電子線回折の自動測定システムを開発

成果発表 Journal of Structural Biology (2020)

概要・世界的位置づけ

クライオ電子顕微鏡は、2017年のノーベル化学賞を受賞対象となるなど、生体分子、薬剤、高分子材料などの立体構造解析に非常に有効な手法である一方、電子顕微鏡を常時操作しながら測定を行う必要があり、**時間と手間がかかるという欠点**があった。

本研究では、世界最高精度を実現したクライオ電子顕微鏡の測定プログラムを、自動測定用にさらに高度化し、**測定開始後は完全に自動で測定するシステムを開発**し、このシステムを用いて多くの分子の構造を高効率かつ高精度に解析した。

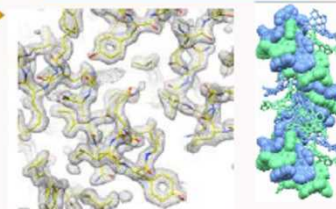


微小結晶から得られた電子線回折パターン
1 Åの分解能を超える回折点が観察できる。

成果の意義・社会貢献への期待

測定の自動化により、高効率で精度の良いデータを多数取得できるようになり、これまで決定が難しかった重要な分子構造の理解が可能となったことで、**創薬や新しい素材開発などへの本格利用**が期待。

さらに本システムは汎用的な電子顕微鏡でも利用できることから、**開発したプログラムと利用手順を公開**した。今後、多くの大学、共用施設、企業などにおいて、**測定標準としての利活用**が期待される。



クライオ電子顕微鏡が
創薬やタンパク質等の構造解析でさらに重要なツールに



古澤 カチームリーダー

ロボットによる微生物の大規模進化実験に成功

成果発表 Nature Communications(2020)

概要・世界的位置づけ

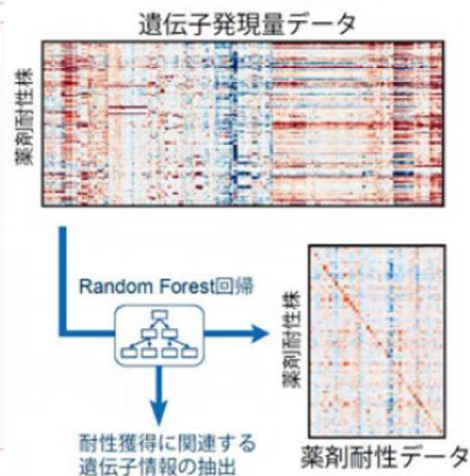
微生物を長期に培養し薬剤耐性進化の過程をハイスループットに解析できる「進化実験ロボット」を構築し、大量解析が可能となった。大腸菌をさまざまな薬剤を添加した環境で進化させ遺伝子発現量やゲノム配列変化などのデータを取得し、機械学習により薬剤耐性進化を特徴づける状態量の抽出に成功した。



進化実験ロボット

成果の意義・社会貢献への期待

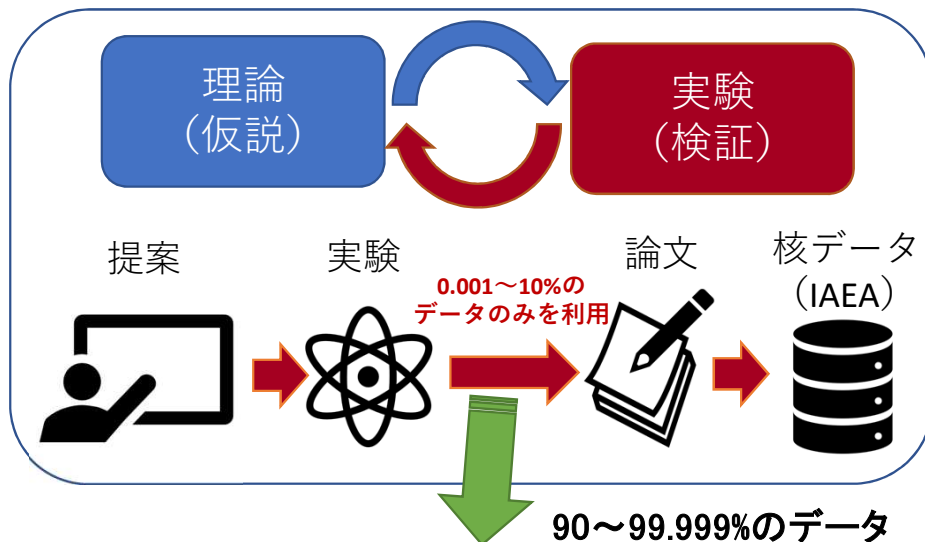
抗生物質への耐性獲得を抑制する手法や新しい抗生物質の開発への貢献、微生物進化の予測と制御による工学・農学分野への応用が期待できる。本解析手法は増え続ける生物データの解釈に貢献する基盤技術となることも期待される。



物理学(加速器科学)におけるオープンサイエンスの取り組み(計画)

従来の加速器科学実験は、1つの目的のために1つの実験を行っていた

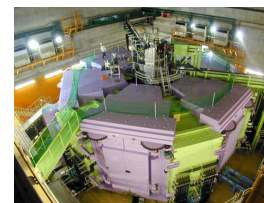
理研RIビームファクトリーでは、10~100の現象を含む、
様々な物理事象を1度の実験で捉える事ができるようになった



データ利活用へ

様々な人がデータ解析できるよう、
・実験条件などのメタデータを整備
・データ解析ツールやノウハウ等の共有化を行い、オープンデータとして公開

- ・ 加速器科学への参入が難しい大学やアジア諸国へのデータ展開
- ・ 原子力分野へのデータ展開 (社会貢献)



理研RI(放射線同位元素)ビームファクトリー
「世界TOPのRIビーム施設」

- ・ 多い時には約1万のセンサーが稼働し、1回の実験で約500TBのデータを産出

現状課題

- ・ 実験データから物理量の算出(較正)には、機器条件や資料の位置測量結果等を加味し膨大な手間を要する(通常2,3年/学生)
- ・ 著名な学術論文(NatureやPhysical Review Lettersなど)でも、執筆に関わる高々10%程度のデータのみが発表
- ・ 残りのデータの中にも理研でのみ取得可能な貴重なデータが含まれている

解決手段

- ・ 大きな実験データを解析に適した大きさに分割して公開、分散解析を行う
- ・ さらに解析ツール類を共有・公開し、オープン・シチズンサイエンスを実施

加速器科学分野ではデータの共有・利活用は進んでいない

ハード面(加速器能力)だけでなく、ソフト面(データ利活用)でも世界をリードする

大型の共用機器のデータ

- 仁科の加速器
- SPRING8のデータ

- 利用できる人が限られている
- オープン化して研究を進める意義は大きい

- データ取得研究などを研究提案として受け入れることが必要

研究のDXのこれから

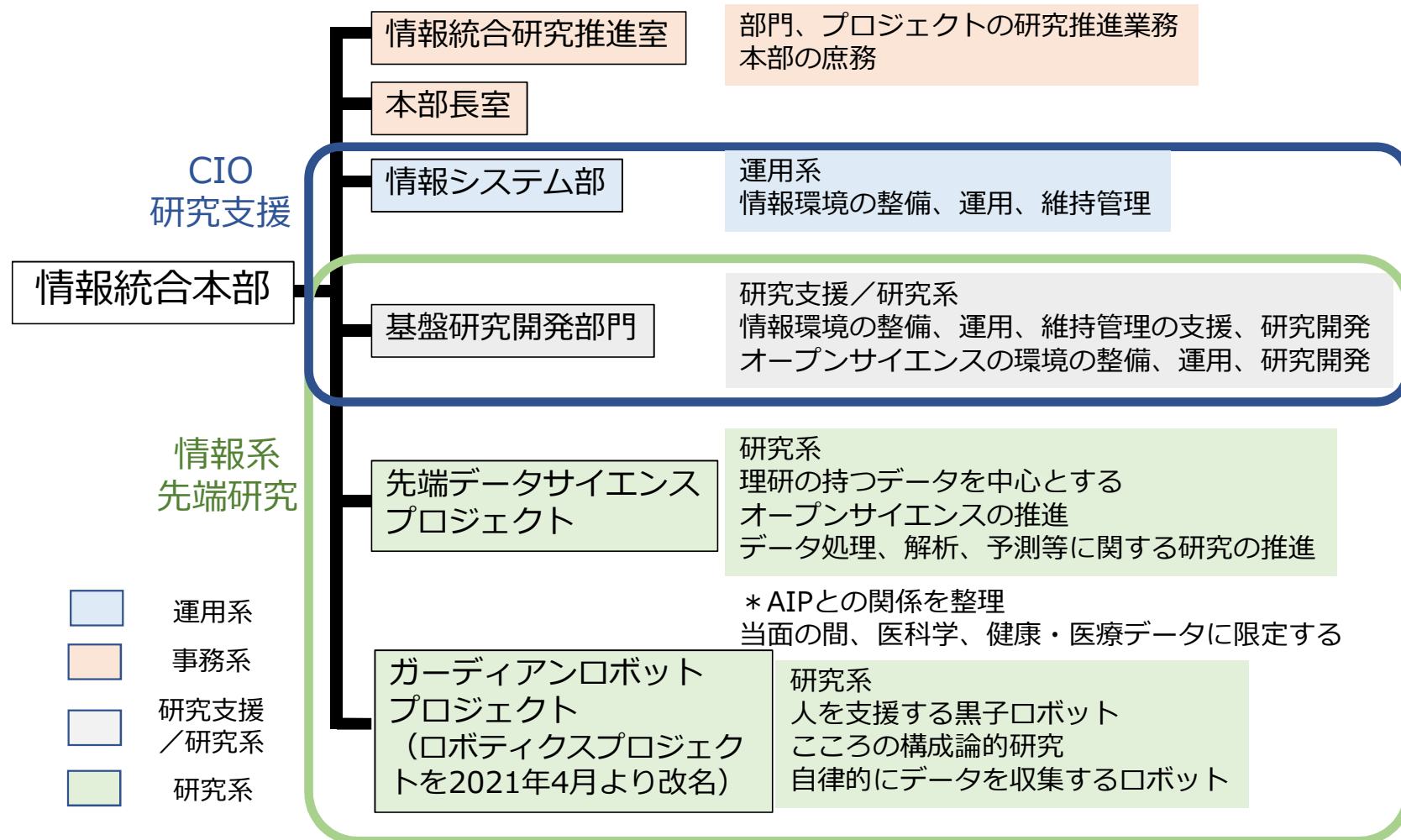
- データ共有による研究のDX
 - 文系の研究は資料の価値が大きい
 - 研究資料の共有により学問の方法が変わる
- 研究の効率化
 - 実験の自動化、遠隔化
 - 実験の再現性向上のためにロボットの導入
- 研究の自動化
 - データ計測からデータ解析までをAIで実現
 - 解析結果の評価もAIで行い、自動で次の実験計画を立てる
 - AIの能力に対する誤解（本質的に無理な話）
- 新たな価値を生み出すことが重要
 - データ駆動科学：仮説検証とは異なる研究の新たな方法論
 - ビッグデータの失敗を教訓として進める必要

ご清聴ありがとうございました。

情報統合本部設立の目的

- データ駆動型研究の推進
 - 理研が創出する研究データのオープン化
 - 利用可能なオープンデータを活用したデータ駆動型研究の推進
 - 脳型ロボットのためのデータの収集、オープン化と利用可能なデータを利用したロボットの知能化
- Society5.0の推進のための基礎技術の研究
- 理研内の情報系研究者の人事交流、ハブ機能
- 理研内の情報系研究の外部からの見える化
 - 情報系の学生、研究者が理研を就職先と考える環境の構築

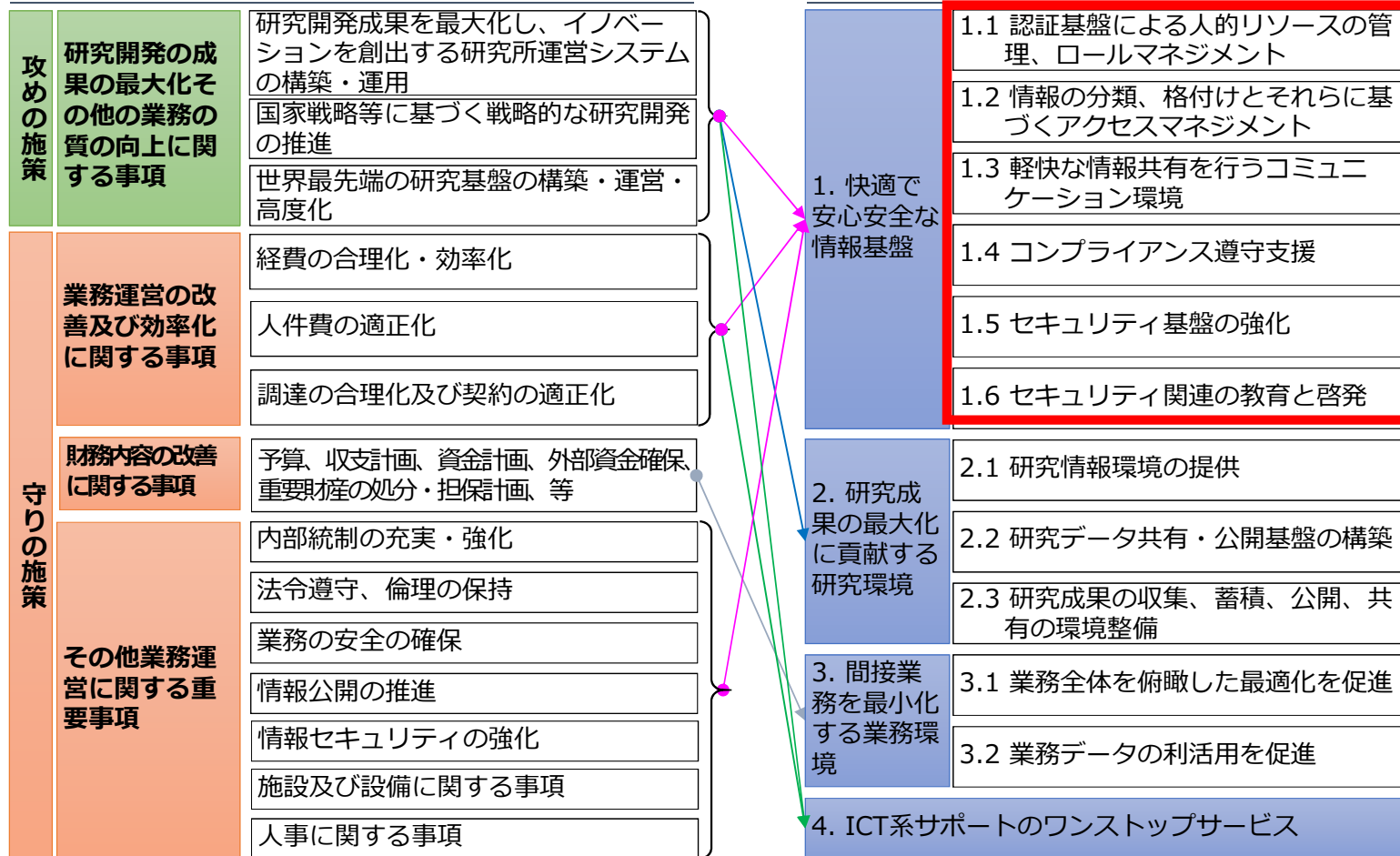
情報統合本部組織図



ICT戦略 概要

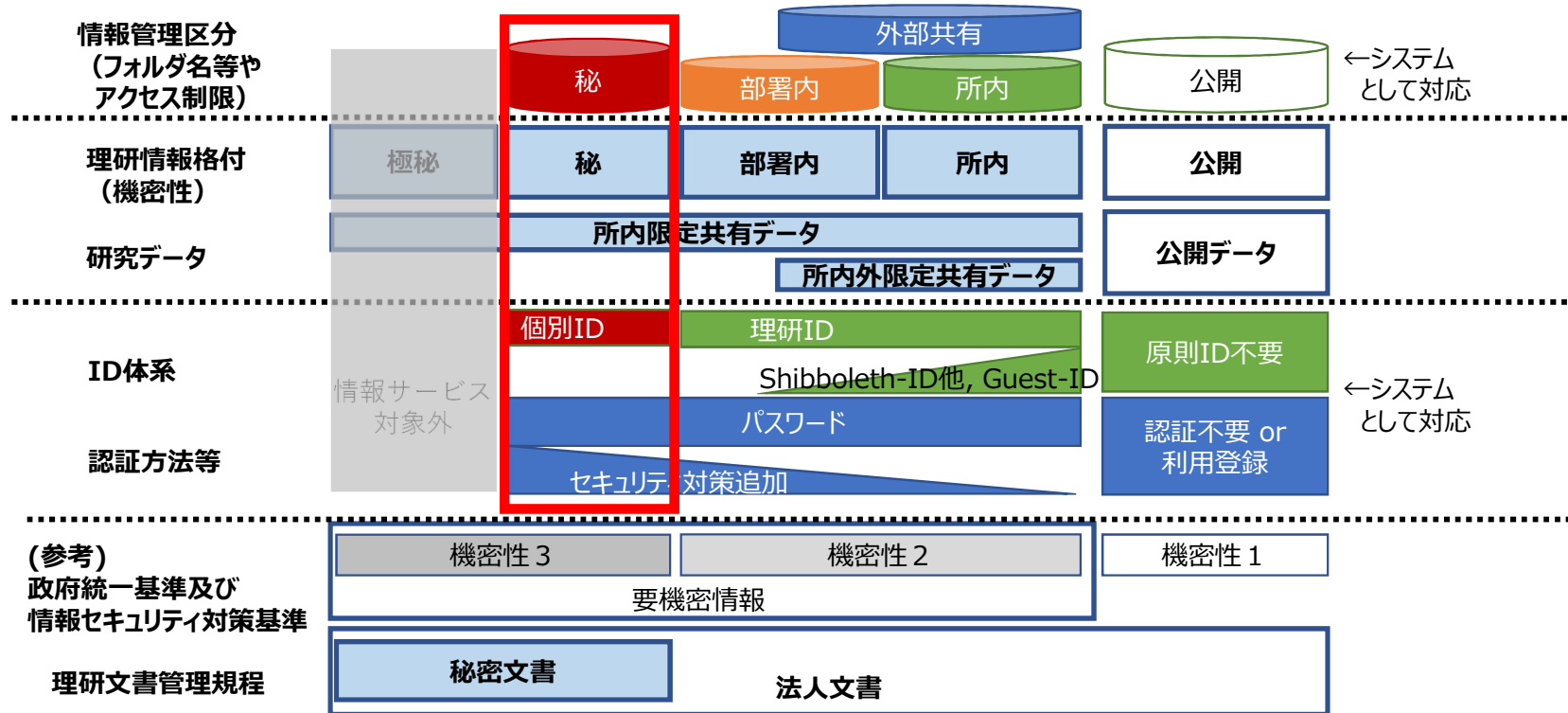
あるべき姿（中長期目標）

打ち手（第4期中長期期間のICT戦略）



中長期目標の達成

ICT戦略に基づく情報サービスでの 情報格付けと情報管理区分、認証方式



格付例

極秘： 他機関が極秘と指定した情報、外国為替及び外国貿易法の規制対象となる機微技術情報など。

秘： 個人識別符号が含まれる個人情報や要配慮個人情報で個人識別が相対的に容易なもの、核セキュリティ対策に関する情報、存在自体が秘密を要する契約情報、各部署が保有する機微な調査関連資料など。

※ 機密性に応じた安全管理措置が必要。

※ 機密性は時の経過等により変化しうるものであり、ある時点において公開以外の格付をした情報が、別の時点においても当然に不開示情報に該当するわけではなく、開示請求があった都度、開示の可否を判断しなければならない。

Security by Designによる情報システムの構築

