

データ駆動型科学に向けた 研究支援体制の整備

—豪州事例を中心に

研究データ管理(RDM)説明会2022 in 大阪

～研究データポリシーと研究データ基盤の活用について～

2022年12月7日

国立情報学研究所 船守美穂

Outline

1. データ駆動型科学の動向とデータプロフェッショナルの必要性
2. データ駆動型科学に向けた組織的支援

データ駆動型科学の動向と データプロフェッショナルの必要性



The Fourth Paradigm: Data-Intensive Scientific Discovery

Tony Hey

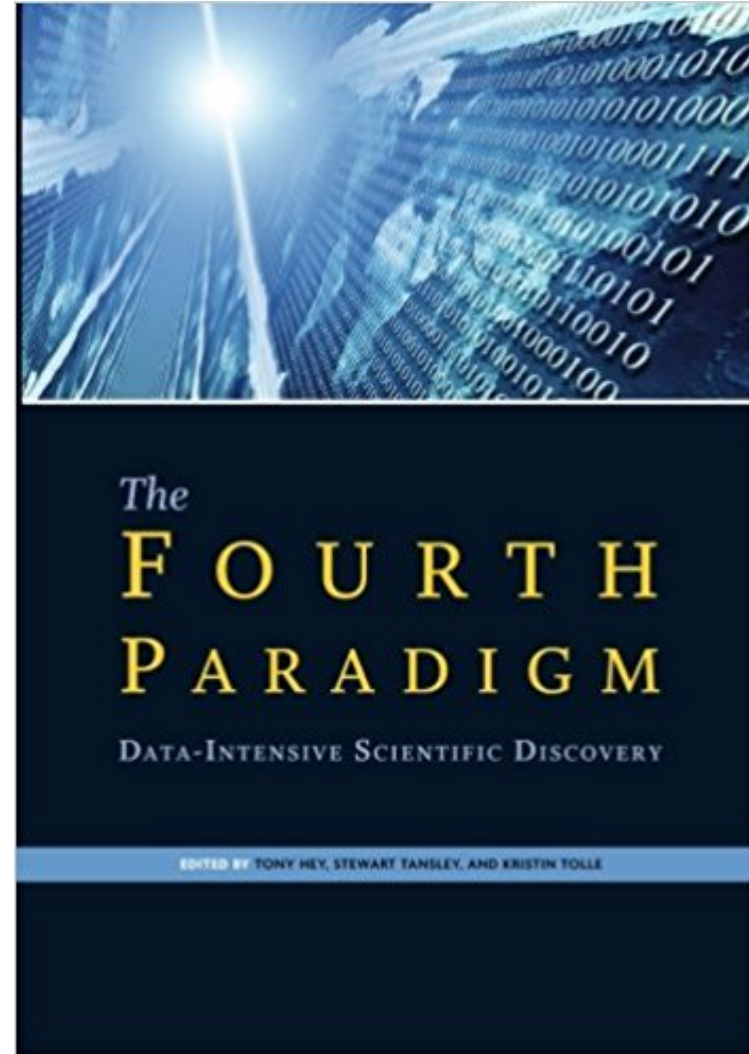
Corporate Vice President

Microsoft External Research



Emergence of a Fourth Research Paradigm

1. Thousand years ago – **Experimental Science**
 - Description of natural phenomena 実験科学
2. Last few hundred years – **Theoretical Science**
 - Newton's Laws, Maxwell's Equations... 理論科学
3. Last few decades – **Computational Science**
 - Simulation of complex phenomena 計算機科学
データ集中科学
4. Today – **Data-Intensive Science**
 - Scientists overwhelmed with data sets from many different sources
 - Data captured by instruments
 - Data generated by simulations
 - Data generated by sensor networks
 - eScience is the set of tools and technologies to support data federation and collaboration
 - For analysis and data mining
 - For data visualization and exploration
 - For scholarly communication and dissemination



With thanks to Jim Gray



This work is licensed under a [Creative Commons Attribution 3.0 United States License](https://creativecommons.org/licenses/by/3.0/).



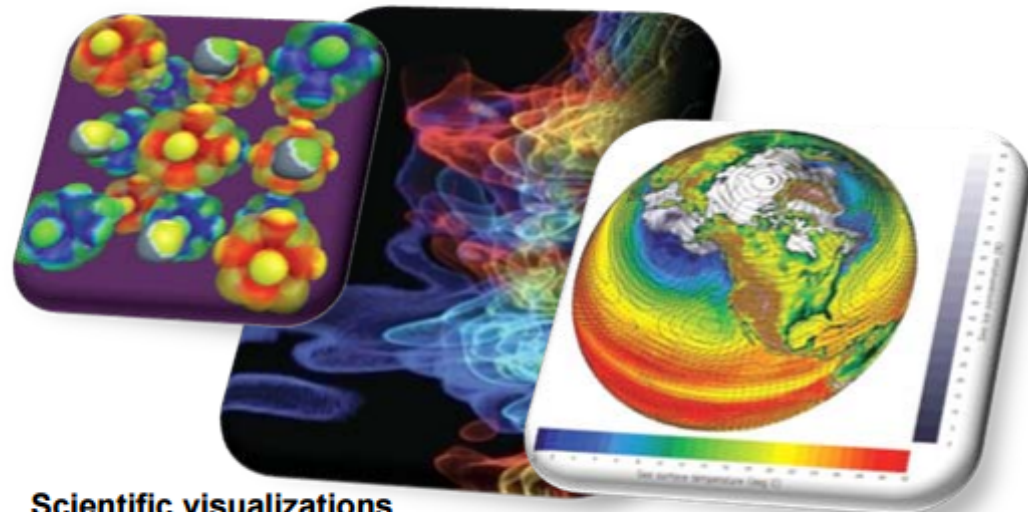
A Digital Data Deluge in Research

- Data collection データ収集時
 - Sensor networks, satellite surveys, high throughput laboratory instruments, observation devices, supercomputers, LHC ...
- Data processing, データ加工・解析 analysis, visualization 可視化
 - Legacy codes, workflows, data mining, indexing, searching, graphics ...
- Archiving データ保存時
 - Digital repositories, libraries, preservation, ...



SensorMap

Functionality: Map navigation
Data: sensor-generated temperature, video camera feed, traffic feeds, etc.



Scientific visualizations

NSF Cyberinfrastructure report, March 2007



This work is licensed under a [Creative Commons Attribution 3.0 United States License](https://creativecommons.org/licenses/by/3.0/).



Helmholtz Data Federation

...参加機関

Helmholtz Data Federation (HDF)



- Initial set of scientific cases identified by the six Helmholtz partners

- Reviewer statements:

- “Sharing of infrastructure, services and datasets opens the way for **improved cross-disciplinary working and re-use of data.**”
- “...will be a **first example of a national shared research data infrastructure in Europe.**”



極域氷床、海洋研究

素粒子物理

ゲノム研究、放射治療

バイオメド、植生

健康科学、原子核物理

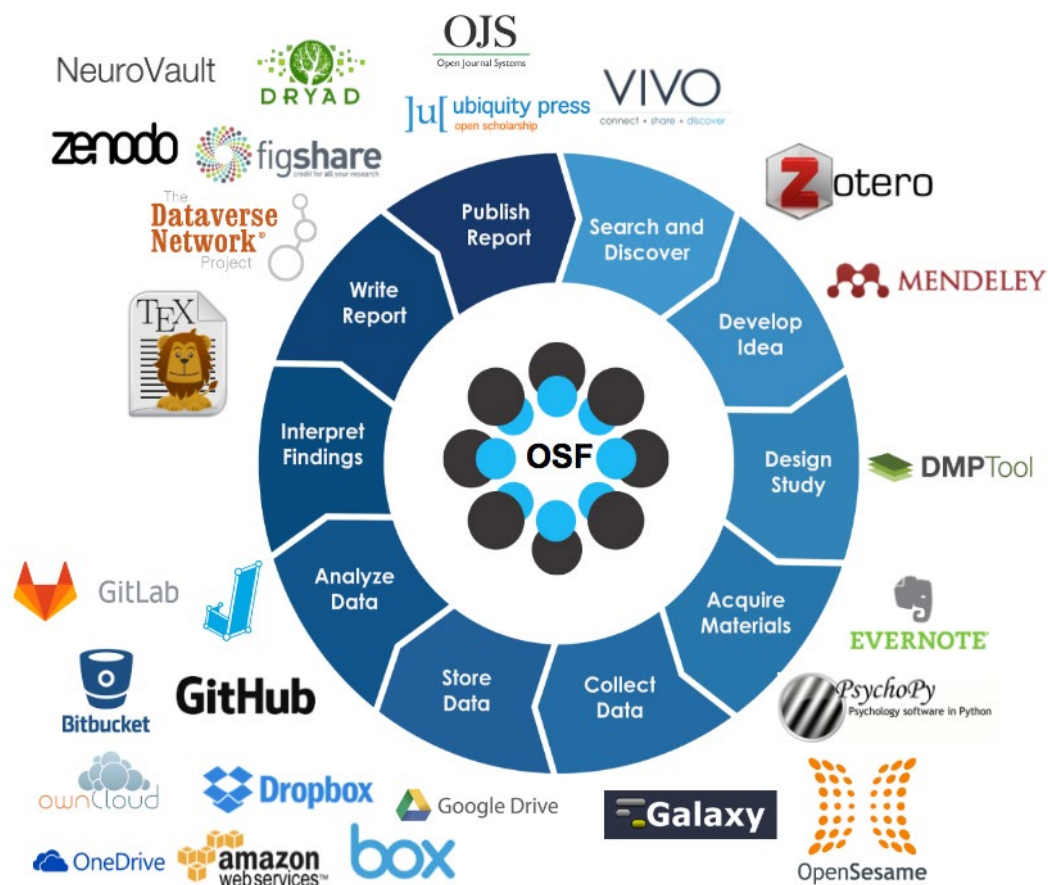
気候、エネルギー、宇宙物理

グローバル化と協働作業

□ インターネットの出現により、国際共同研究の機会が拡大

■ 情報の共有・保存のためのプラットフォーム

■ オンライン・コラボレーションのプラットフォーム





欧州・地中海 植物保護機関（国際機関）



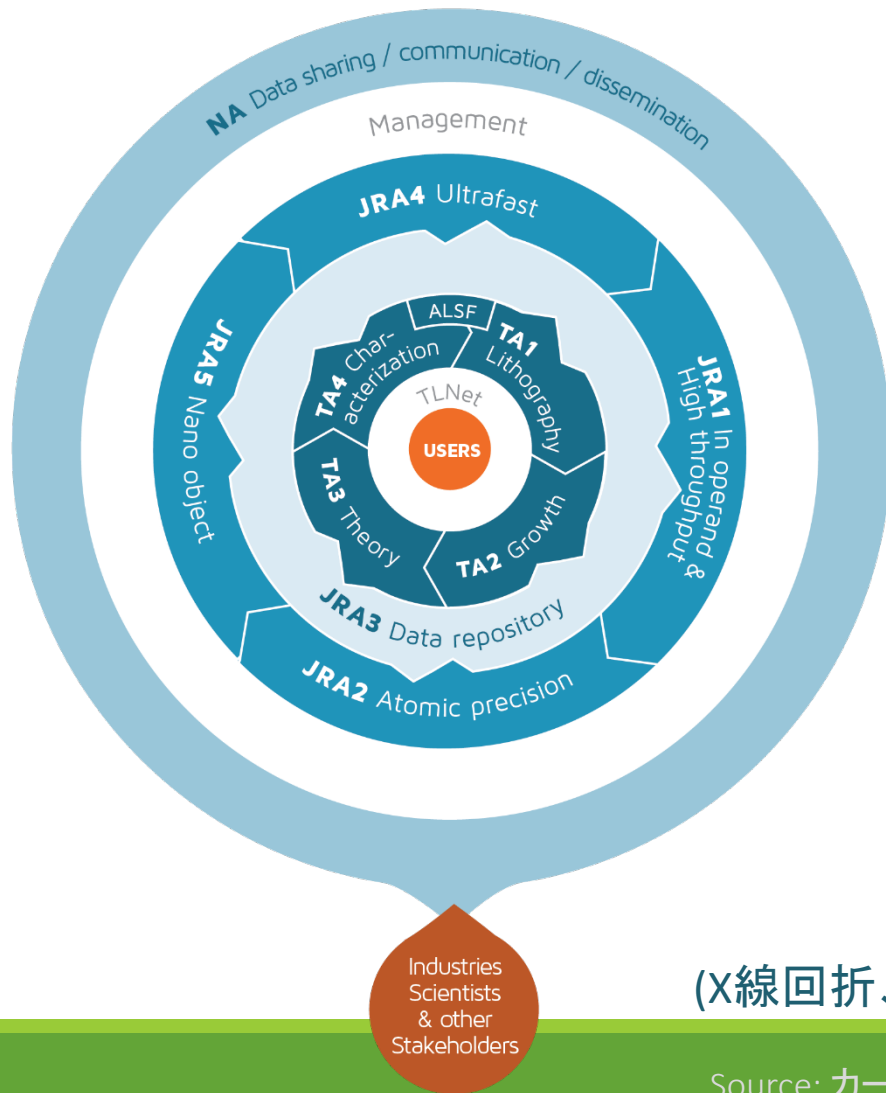
- 国際植物防疫条約（IPPC）に基づき、欧州・地中海地域51カ国について設置された、植物保護の協力・調和機関。
- 事務局は20名程度だが、研究者は計4500名規模。
- 各観測点における植生や害虫、天候等の情報をデータベース化。
- 各種基準や標準のガイドラインを作成。

〈データベース〉

- EPPO Global DB & EPPO codes
- EPPO Standards on Efficacy Evaluation of Plant Protection Products
- PQR - EPPO database on quarantine pests
- EPPO database on Diagnostic expertise
- CAPRA (Computer Assisted Pest Risk Analysis)
- EPPO Pest Reporting System (restricted access)
- EPPO Plant Protection Thesaurus

〈EPPO標準〉

- EPPO Standards on plant protection products (3標準)
- EPPO Standards on phytosanitary measures (10標準)



□ EU内の複数の測定機器により得られた測定値を一つのリポジトリに集め、**データ統合、解析するツールを提供する。**

(利用場面)

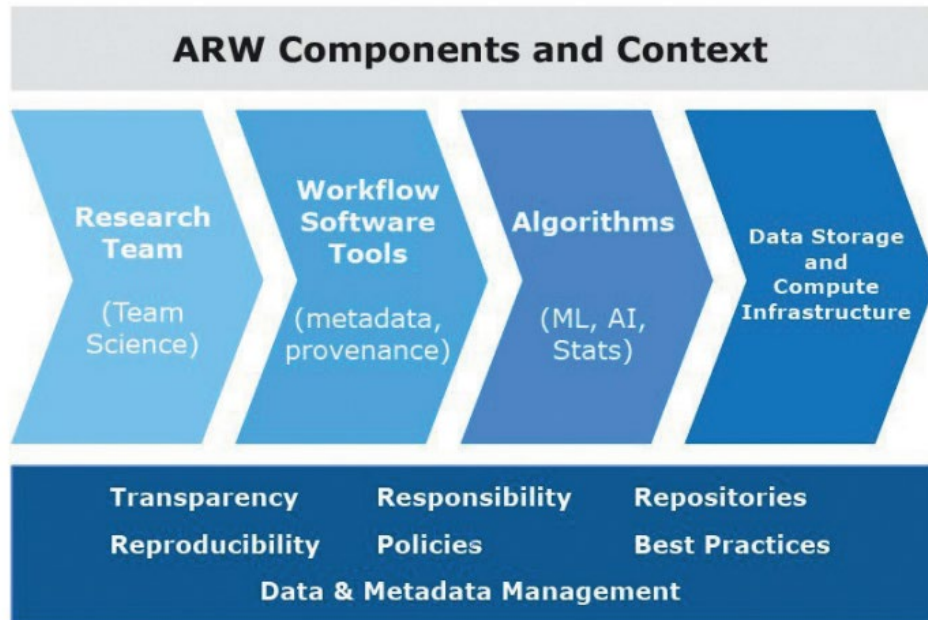
- 一研究者が特定サンプルの計測のために、複数の施設を利用。データをNFFAのデータリポジトリに保存。
- 研究グループがデータを共有、共同解析
- 他の研究者が、研究データを利用、解析

(X線回折、NMR、質量分析、電子顕微鏡、微細加工・・・)

研究自動化 (ARW)

Automated Research Workflow

- ARWは研究プロセスを大幅に加速する。また、研究の再現性なども確保する。
- このため、ARW推進のための情報資源や情報技術の開発が必要である。
- ARWは、異分野の研究者やデータプロフェッショナルの協働(チームサイエンス)を必要とするため、多様な研究貢献を評価するインセンティブが必要である。
- さらに、個人情報の取扱い等、法制度や技術的課題の解決が必要である。



ESFRIにおけるe-インフラの重視



Professor John Womersley

前ESFRI議長 John Womersley

A lot of the research infrastructures are really data infrastructures or networks that bring together datasets from existing facilities or large single-sited infrastructures that generate very, very large amounts of data. It is no longer possible or sensible to simply look at the facility without also considering the data handling.

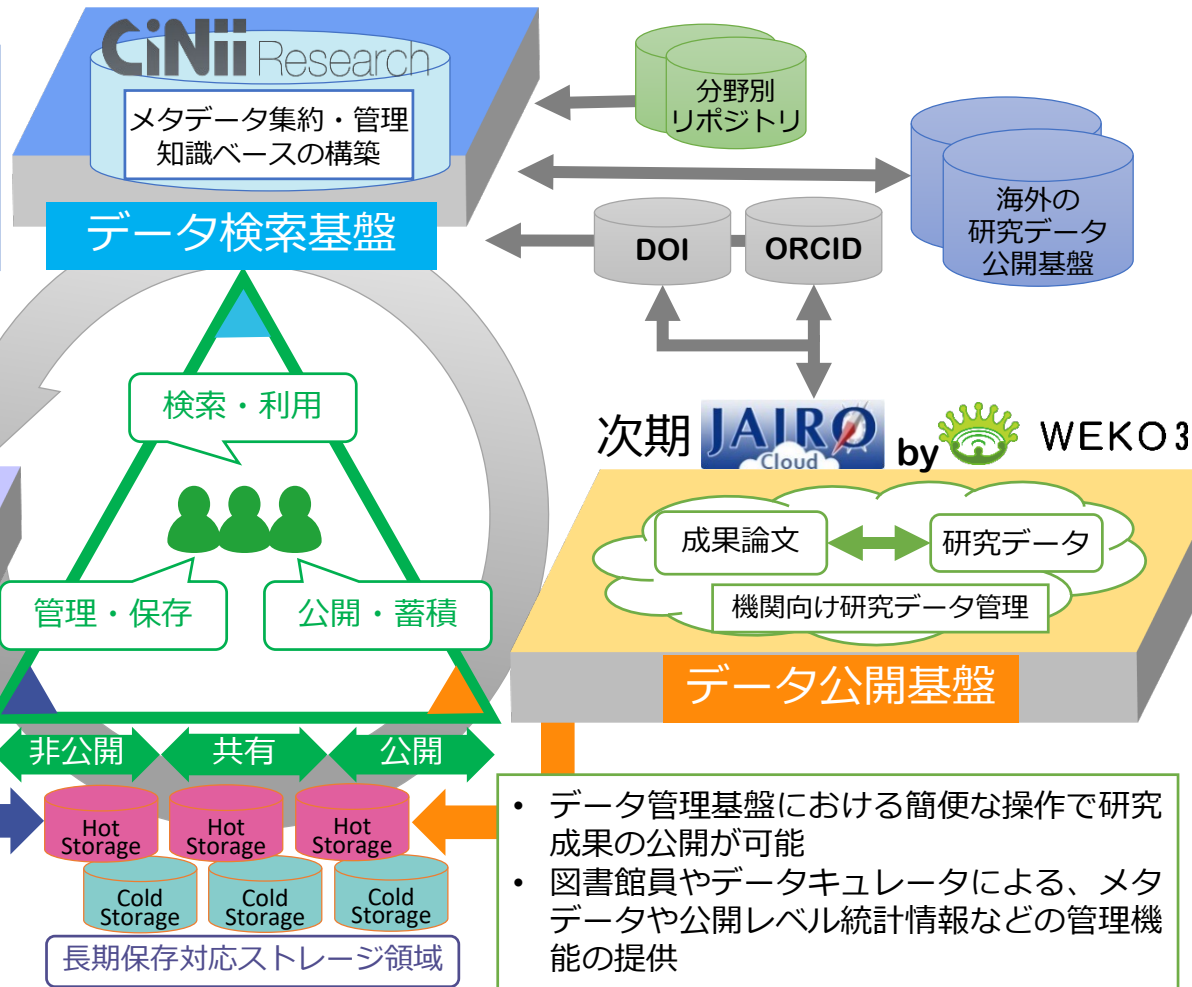
The emerging “e” component of research infrastructure

ESFRI: The European Strategic Forum for Research Infrastructures

- 欧州域内において戦略的に整備・運営していくべき研究インフラについて欧州連合理事会に助言する、極めて影響力の大きい機関

NII Research Data Cloud

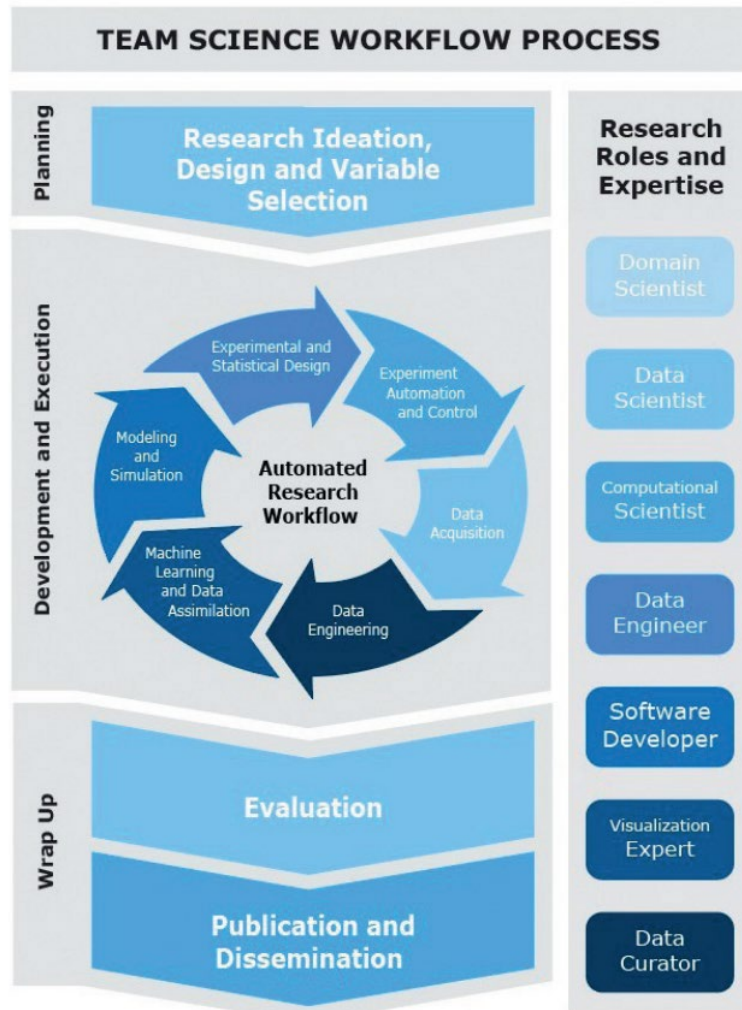
- 機関リポジトリ + 分野別リポジトリやデータリポジトリとも連携
- 研究者や所属機関、研究プロジェクトの情報とも関連付けた知識ベースを形成
- 研究者による発見のプロセスをサポート



- データ収集装置や解析用計算機とも連携
- 研究遂行中の研究データなどを共同研究者間やラボ内で共有・管理
- 組織が提供するストレージに接続した利用が可能

- データ管理基盤における簡便な操作で研究成果の公開が可能
- 図書館員やデータキュレータによる、メタデータや公開レベル統計情報などの管理機能の提供

チームサイエンスを必要とするARW



分野の研究者

データサイエンティスト

コンピュータ科学者

データエンジニア

ソフトウェア開発者

ビジュアル化専門家

データキュレータ

みんなの
協力が必要。
でもそのためには
各自の貢献が
適切に評価される
必要がある！



データ駆動型科学の促進

...データプロフェッショナルの育成・活用

データの取扱・解析に係わる
データプロフェッショナル

データの共有・公開に係わる
データプロフェッショナル

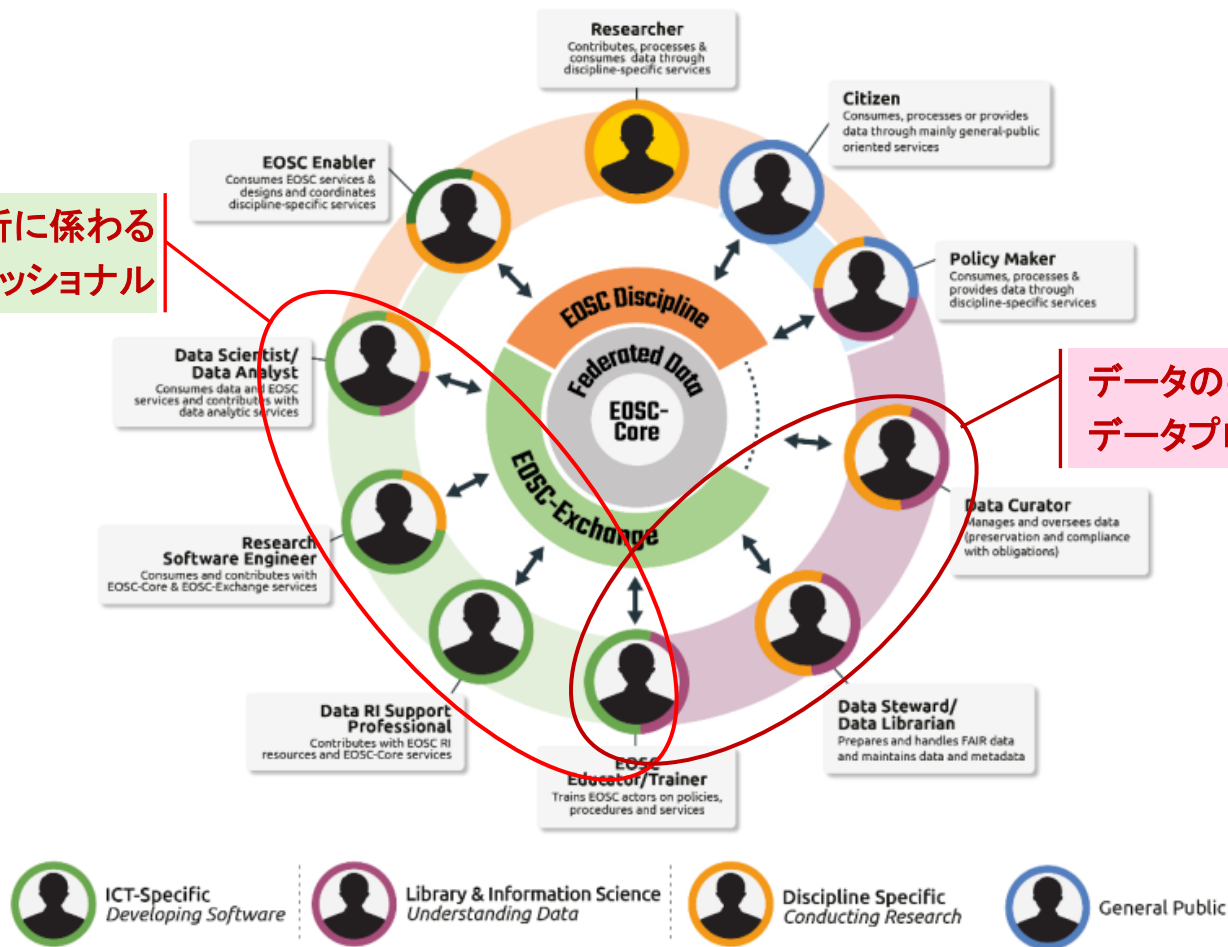
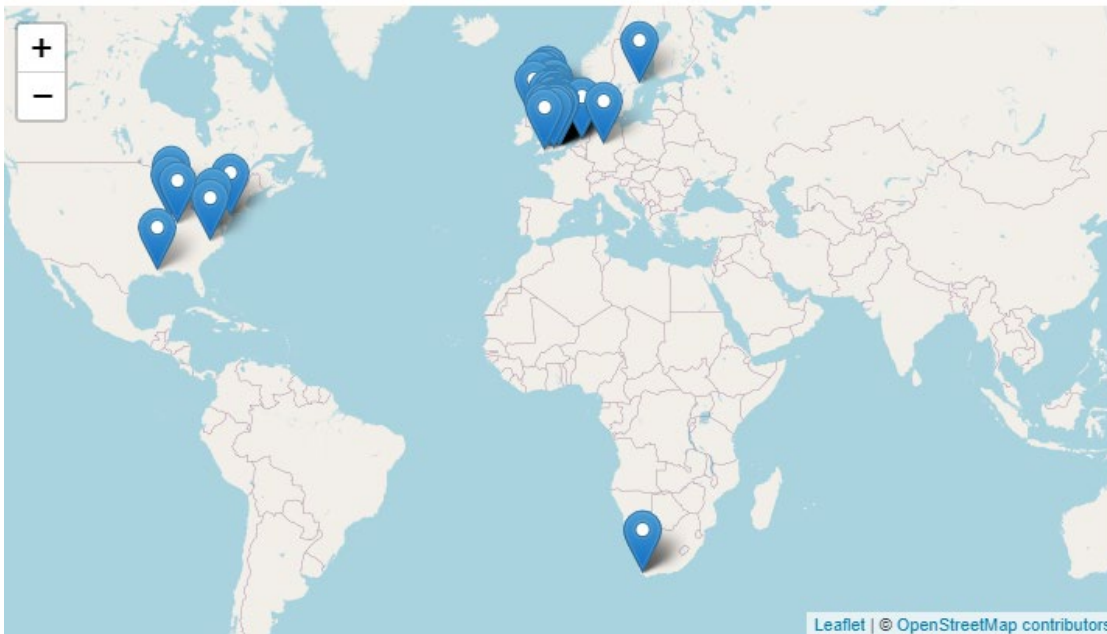


Figure 6.1: Actors in the EOSC ecosystem: roles and interactions

世界に広がるリサーチエンジニア



Society of Research Software Engineering



大学 ICT 推進協議会 提言
「多様な教育研究活動の高度化を支える
大学 ICT 基盤の 集約化・共通化・協働化
～コロナ時代における大学の
デジタルトランスフォーメーションに向けて～」
(2020年12月)

【2030年の大学情報環境】

1. 運営母体としての大学間協働事業体
2. 大学経営における柔軟な情報戦略
3. ICT 人材・キャリアパスの多層化

大学間だけでなくICTに係る民間企業との人材環流が進み、従来にはなかった博士号を有する「**リサーチエンジニア**」と呼ばれる新しい職種が生まれている。各大学のICT基盤を支える人材がより豊富な経験を積み大学間で知見を共有し、国内外の研究開発コミュニティとも繋がった結果、大学全体のICT環境の改善が継続的に推進されている。CIOやリサーチエンジニアだけでなく、利用者支援人材を含むICT環境整備に必要な様々な人材ポートフォリオが整備されている。その結果、大学や企業を渡り歩きながらキャリアアップする組織的に裏打ちされたキャリアパス等、多様な人材が育成されるキャリアパスが形成されている。

4. 国際通用性の担保

データ駆動型科学に向けた 組織的支援

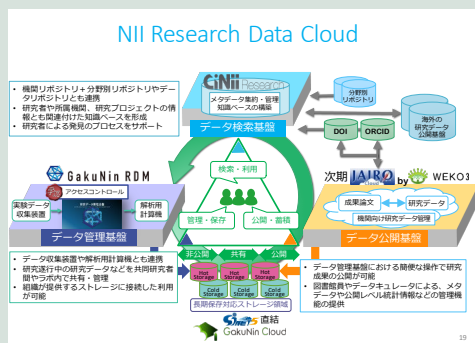


データ駆動型科学の促進...組織的な支援事例①

...研究データ管理・保存・記録環境の提供

組織的支援

- ストレージ
- 研究データ管理基盤
- DMR ツール
- 機関リポジトリ



研究者のメリット

- ✓ セキュアな保存環境
- ✓ バックアップ、セキュリティパッチの手間いらず
- ✓ 容易な研究データ管理環境
- ✓ 半自動の研究データ管理記録環境
- ✓ 永続的な研究データ公開環境

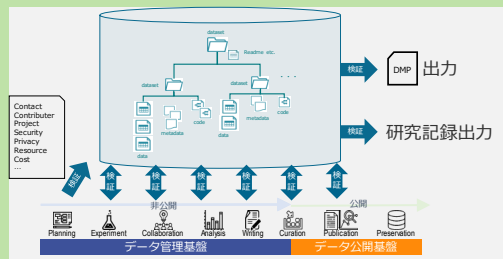


データ駆動型科学の促進...組織的な支援事例②

...解析ツール、データガバナンス機能の提供

組織的支援

- 解析ツール
- 解析ツールと研究データ管理機能の連携
- 研究活動時のデータガバナンス機能
- e-ラボノートブック
- ビジュアル化ツール



研究者のメリット

- ✓ 解析ツールの自己調達不要
- ✓ デジタルスキルの底上げ
- ✓ 解析過程の自動記録と研究データ管理
- ✓ 優れたビジュアル化



データ駆動型科学の促進...組織的な支援事例③

...データ取扱いに関わるコンサルティングの提供(1)

組織的支援

- 全学センターによる支援事例:
モナシュ大学eResearchセンター
 - ✓ 約40名のデータプロフェッショナルを配備
 - ✓ 学内2000の研究プロジェクトにコンサルティング提供
 - ✓ データの取扱いについて個別対応をするうちに、データの管理・保存・伝送・解析・可視化・共有等の技術において分野横断的な共通項を見いだすに至る。

- 大型計算機/情報基盤センターによる支援事例: 豪州複数の大学において実施
 - ✓ 大型計算機の利用について助言や人材派遣
 - ✓ データの管理・保存・伝送等について助言
 - ✓ 学内複数の計算機リソースと大型ユーザのデータファブリックを合理化

研究者のメリット

- ✓ きめ細やかな支援享受
- ✓ デジタルスキルの底上げ
- ✓ **研究のスケールアップ**



eResearchセンターのスコープ

...ケープタウン大学の事例

パートナーとして
知恵を出し合う
ことが大事



Creating a world-class environment for 21st century discovery

UCT eResearch **partners** with research groups to **accelerate** and **transform** research, **connecting** them to the most appropriate hardware, software and services to **sustain** that capability



Collect



Compute



Comprehend



Collaborate



Communicate



Customise

UCT eResearch

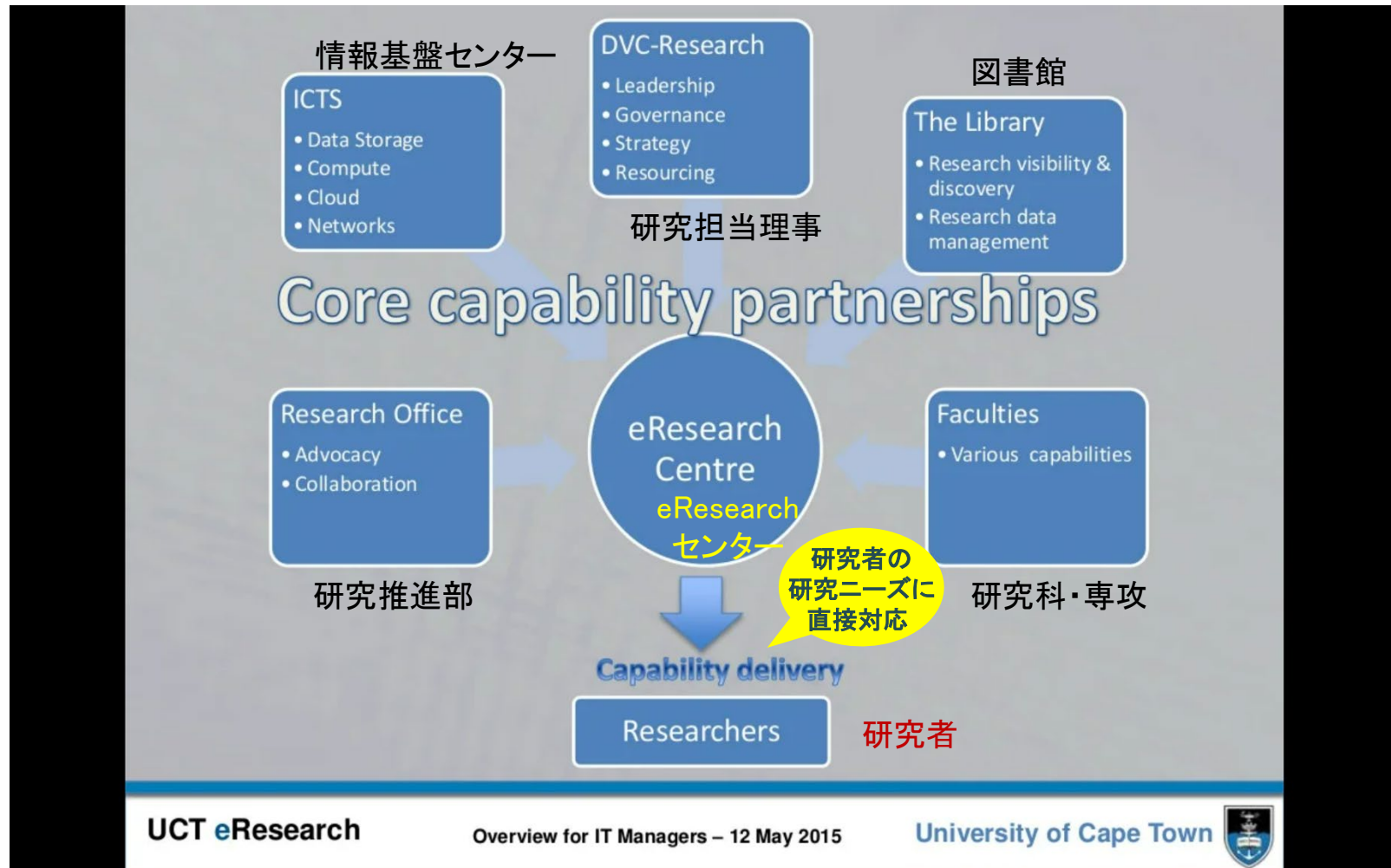
Overview for IT Managers – 12 May 2015

University of Cape Town



eResearchセンターの学内体制

...ケープタウン大学の事例



データ駆動型科学の促進...組織的な支援事例④

...データ取扱いに関わるコンサルティングの提供(2)

大学Aの
コア
ファシリティ



大学Bの
コア
ファシリティ



大学Cの
コア
ファシリティ



大学Dの
コア
ファシリティ



大学のコアファシリティに、データの取扱いに関わるコンサルティングの提供

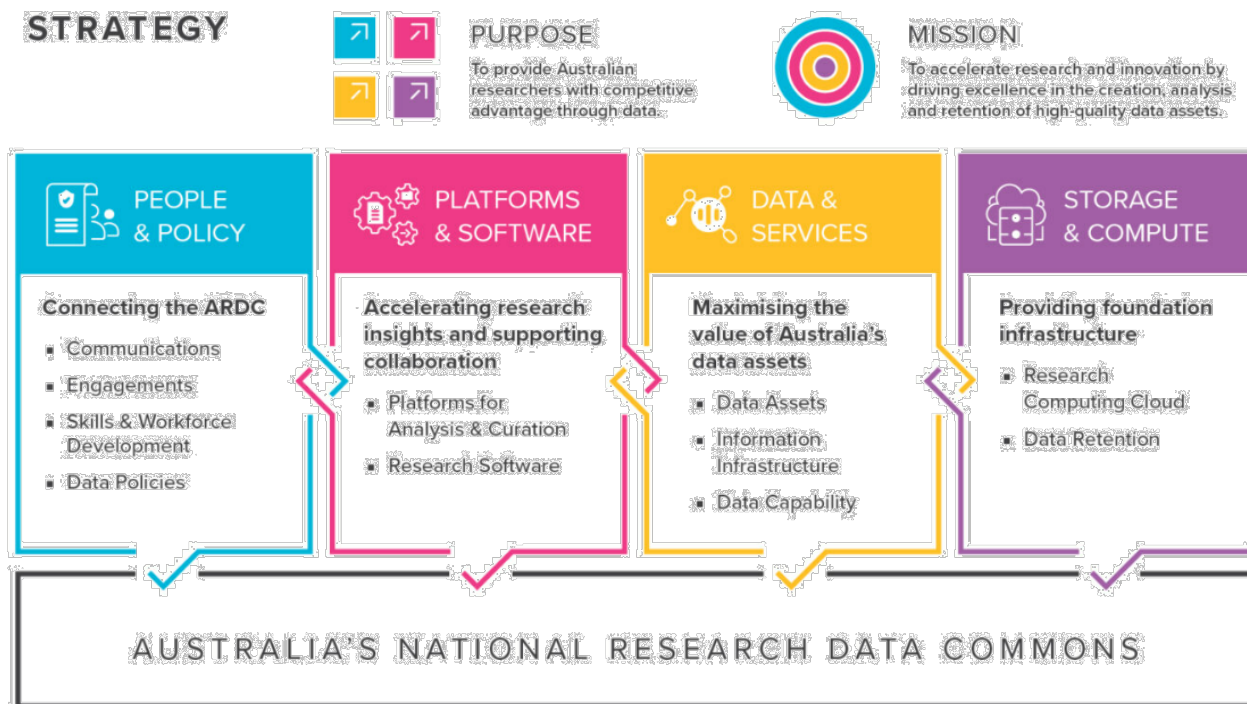
事例: Microscopy Australia

1. 全国大学の共用研究機器としての顕微鏡類とユーザのマッチングサービス
2. 顕微鏡類を扱う全国大学のコアファシリティに対して、データの取扱いに関わるガイドライン策定、コンサルティング、人材派遣。

豪州研究データコモンズ

Australian Research Data Commons (ARDC)

- ARDC: データに関連する標準やスキル、研究基盤を整備することにより、豪州研究コミュニティと産業界の研究競争力につなげることを目的として、2018年に設置される。
- 豪州協働型研究基盤戦略(NCRIS)の一部として位置付けられ、2021年度の予算は3910万豪ドル。
- 学术界と産業界にサービスを提供し、豪州の研究データ及びe-リサーチの国際的優位性を形成する。



EOSCと違って
一国型モデルだから
日本も参考に
しやすい



EOSC戦略的研究イノベーション課題 (SRIA, 2021)

European Open Science Cloud Objectives Tree

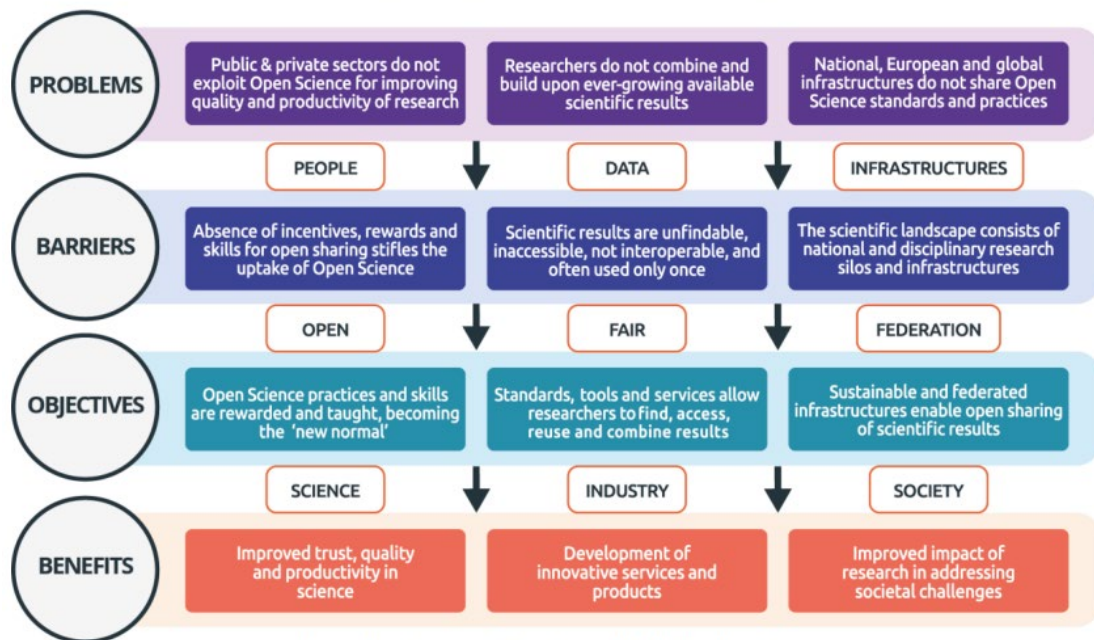


Figure 0.1: European Open Science Cloud Objectives Tree



データとインフラ整備だけでなく、ポリシーやガイドライン、人やコミュニティ形成に注力していることにポイントあり！



Communities

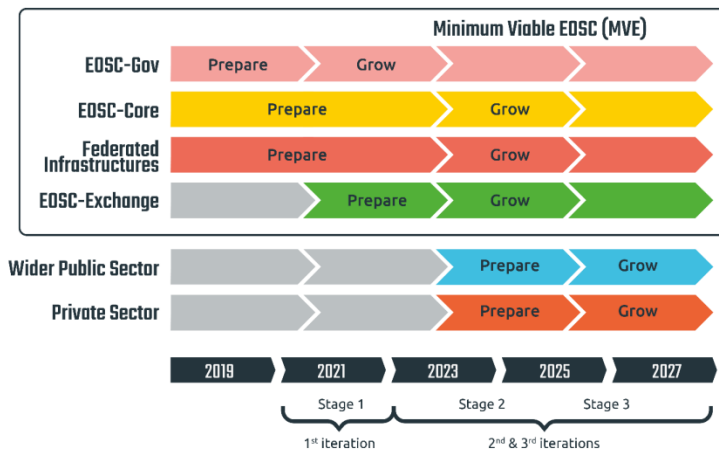


Figure 4.2: Schematic representation of timelines of EOSC iterations



Funders, Organisations and Ministries

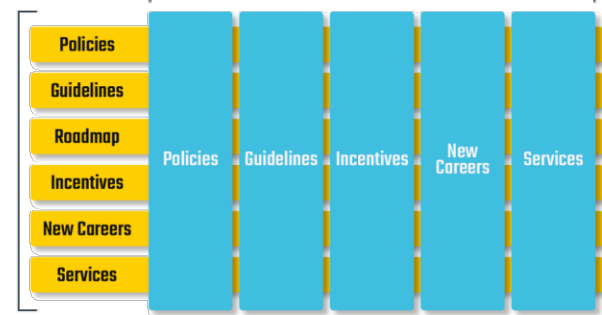
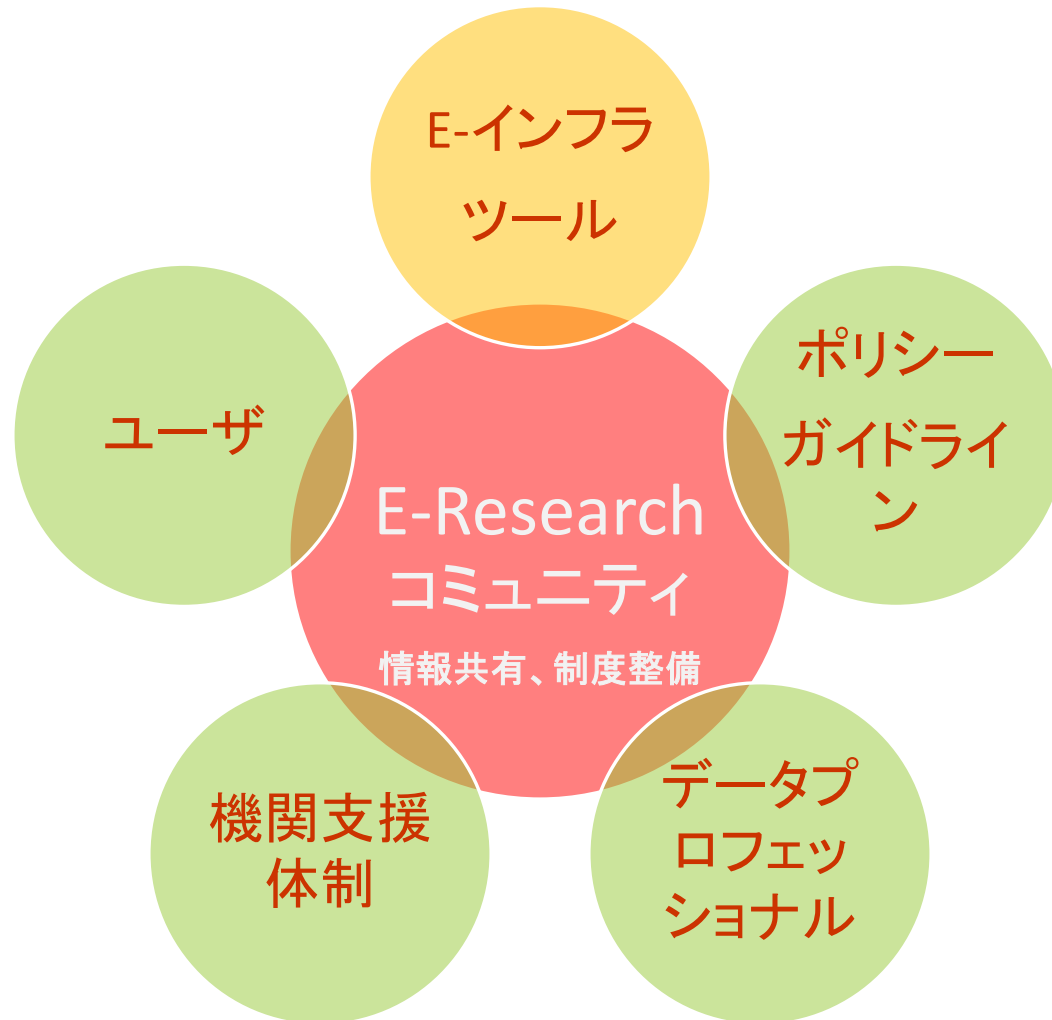


Figure 1.5: Open Science at the crossroads between communities and funders, organisations, and ministries

データ駆動型科学とオープンサイエンスを 推進するために



コミュニティとして
制度整備できると
いいね！



海外大学事情 mihoチャンネル 配信中

<https://rcos.nii.ac.jp/miho/>

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
Research Center for Open Science and Data Platform

ENHANCED BY Google

日本語 English

TOP RCOSについて プロジェクト オープンサイエンスとは 各種資料 RCOS日記

TOP > RCOS日記 - mihoチャンネル

mihoチャンネル

世界の高等教育や学術情報流通に関わる最新の動きについて、各種のネット上の情報源や自身で取材してきた内容をもとにご紹介します。あくまでも研究者としての情報発信で、RCOSあるいはNIIの公式見解ではありません。情報の出典は可能な限り明示してありますので、内容の正確性についてはそちらをご確認ください。

(船守 美穂)

2020.05.29	大学の管理・運営	コロナ下の米国大学 (3) : オンライン授業続行を宣伝する大学
2020.05.28	大学の管理・運営	コロナ下の米国大学 (2) : 「9月は授業をキャンパスで実施」の発表相次ぐ
2020.05.27	大学の管理・運営	コロナ下の米国大学 (1) : 米国の2020年度大学進学人数、2割減か？
2020.04.22	学術情報流通	ネイチャー誌、プランSに含まれる見込み
2020.03.13	研究評価	中国、研究評価におけるSCI論文と関連指標の使用を規制
2019.12.21	国内政治と学術界	復旦大学、大学憲章から「思想の自由」の文言を削除

カテゴリ別

- OER
- オンライン教育
- オープンアクセス
- オープンサイエンス
- ジェンダー
- データサイエンス

連絡先

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
船守美穂 (funamori@nii.ac.jp)