

# オープンイノベーションに資する オープンサイエンスのあり方

日本学術会議 オープンサイエンスの取組  
に関する検討委員会

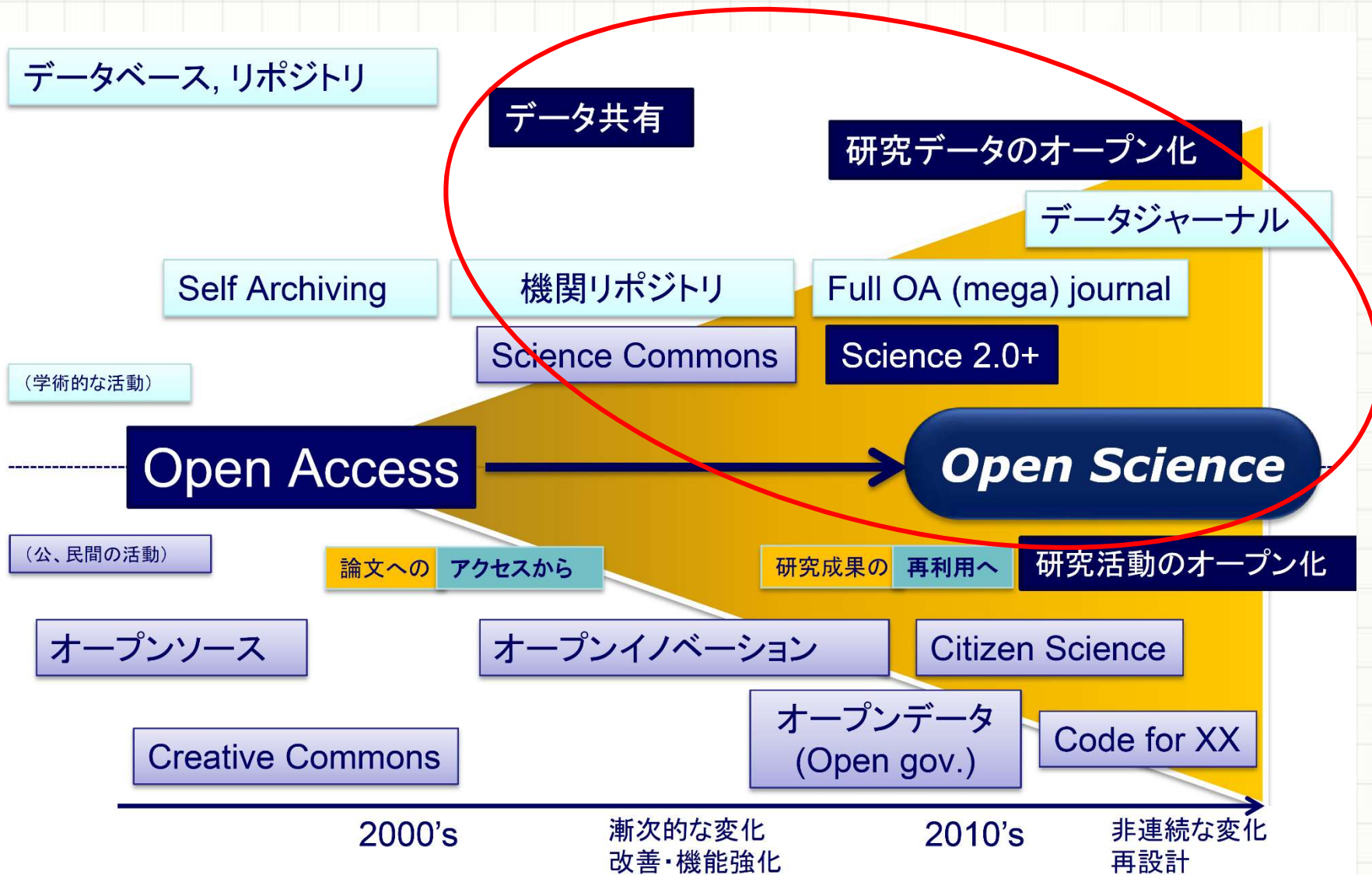
委員長 土井美和子（第三部副部長）

情報通信研究機構

2016年10月18日

文部科学省学術情報委員会

# 対象とするオープンサイエンス



内閣府「国際的動向を踏まえたオープンサイエンスに関する検討会」報告書より引用

# 意見聴取リスト

- 我が国におけるオープンサイエンス推進の在り方について 真子内閣府政策統括官付参事官補佐
- ICSU-WDS等の国際事業と科学データ共有動向 村山NICT室長
- ライフサイエンスにおけるデータ共有の現状と課題 高木東大教授・JST NBDCセンター長
- 材料分野におけるオープンサイエンスについて 藤田NIMS部門長
- 微生物・環境データの統合化と今後の展開 黒川東工大教授
- 天文学における世界規模のデータ共有 大石国立天文台 天文データセンター センター長
- 社会学におけるオープンサイエンスへの取り組み:社会科学のデータアーカイブの構築 北村一橋大教授
- 人文学とオープンサイエンス 青柳文化庁長官
- アカデミア創薬研究とオープンサイエンス 長野医薬品医療機器総合機構理事
- オープンサイエンスに関する日本の最近の活動事例について 林文科省NIST EP上席研究官
- 学術情報のオープン化の推進について(審議まとめ)概要 渡邊文科省学術基盤整備室室長
- オープンサイエンス推進のための研究データ基盤 安達国立情報科学研究所副所長

# 海外動向

- 2008年 品質管理された科学データの長期的な保全と提供を支援するICSU-WDS (World Data System) (世界科学データシステム) 設立。その国際プログラムオフィス (WDS-IPO) は2010年国立研究開発法人情報通信研究機構 (NICT) に設置。
- 2013年 データ共有を円滑に推進するため、国際的なコンソーシアム Research Data Alliance (以下 RDA) 創設。研究者や技術者によるボランティアベースで国際標準の形成を目指す。第7回 RDA 総会は2016年3月にアジア初東京で開催。
- 2015年 ICSU-IAP-ISSC-TWAS のワーキンググループが科学者や研究機関、出版社などオープンデータに責任を持つべきものなど10項目の指針をまとめた。
- オープンサイエンスのための共通基盤整備が欧州 (EUDAT)、米国 (OSF: Open Science Framework,) などで推進。





## Principles of Open Data

- Responsibilities
  - Scientists
  - Research institutions and universities
  - Publishers
  - Funding agencies
  - Professional associations, scholarly societies and academies
  - Libraries, archives and repositories
- Boundaries of Openness
- Enabling practices
  - Citation and provenance
  - Interoperability
  - Non-restrictive reuse
  - Linkability

<http://www.icsu.org/science-international/accord/open-data-in-a-big-data-world-short>

Science International 18.00h CET, 7 December 2015.

ICSU: the International Council for Science

IAP :the Inter Academy Partnership

ISSC :the International Social Science Council

TWAS : The World Academy of Sciences

# 学協会へのアンケート

- 第1部、第2部、第3部それぞれの関連学協会へのアンケートを実施。
- 回答のあった学協会のうち、過半の学協会が論文や論文にかかわるデータ、データベース等のデジタルデータを既に公開。
- 半数程度の学協会では研究者個人がデータを保有しているが、共通のフォーマット・様式を整えれば価値を生む可能性のあるデータがある。
- 残念ながら、内閣府報告書の存在については、回答のあった学協会ではあまり認識されていなかった。
- 過去5年の間に、オープンサイエンスをテーマとしたイベントの開催事例があるのは、1割程度とあまり多くなかった。
- 半数近い学協会がオープンサイエンスに関する自由意見を寄せており、オープンサイエンスに大いに興味を持っている。

# オープンサイエンスの論点整理 (1/3)

- データ: 現物 (wet data)+デジタルデータ(dry data)  
ex. 創薬等支援技術基盤プラットフォーム(化合物+メタデータ+構造式)

生産 ▶▶ 流通 ▶▶ 活用

- オープンの意味
  - すべてオープンではない、よりオープンに
  - オープン・クローズ戦略とガイドライン
    - 占有期間(embargo)
    - データ すべて/部分的(メタデータetc)/クローズ
    - 解析ツール すべて/部分的/クローズ
    - 論文化/特許化

# オープンサイエンスの論点整理

## (2/3)

- オープン化のインセンティブ(分野別型)
  - 研究成果再利用による迅速化
  - 論文化
  - Vo (Virtual Observatory)/Materials Informatics
  - フィールドサイエンス
  - 研究リソース獲得(例: 東寺百合文書WEB、萬曆版大藏經画像データベース)
- オープン化のインセンティブ(分野横断型)
  - 学際性(例: 古文書+土木データ▶▶地震  
ヒット化合物ライブラリー+創薬スクリーニング▶▶創薬)
  - 社会実装(例: 公的統計+Webデータ▶▶都市計画)



# オープンサイエンスの論点整理 (3/3)

- オープンサイエンスのコスト
  - データ生産コスト
  - データ流通コスト
  - 流通のための標準化にかかる研究者の負担
  - 保存方法(現物(wet data)+デジタルデータ(dry data))
  
- ▶▶活用とバランスのとれたコストか？

# オープンサイエンス分野別俯瞰

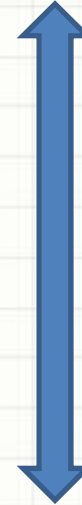
- ホールデータ共有型

- 大型施設などの使用でデータ生産が高コスト
- アウトカムは人類の叡智あるいは環境への貢献
- データを共有、解析プログラムなどは個別ポリシー
- オープンポリシーは占有期間(embargo)あるいは論文化  
ex.天文学

- ブリッジングデータ共有型

- データ解析結果の商用性・活用性が高い
- メタデータや現物などブリッジングに必要なデータや解析プログラムは公開を目指す。ただし公開は提供者に任せられる。
- オープンポリシーは占有期間(embargo)あるいは論文化、特許化  
ex. 材料、創薬

オープン化大



オープン化小

VOを使って生み出された  
査読論文  
1073 (ADS調べ)

2002年1月～2015年6月

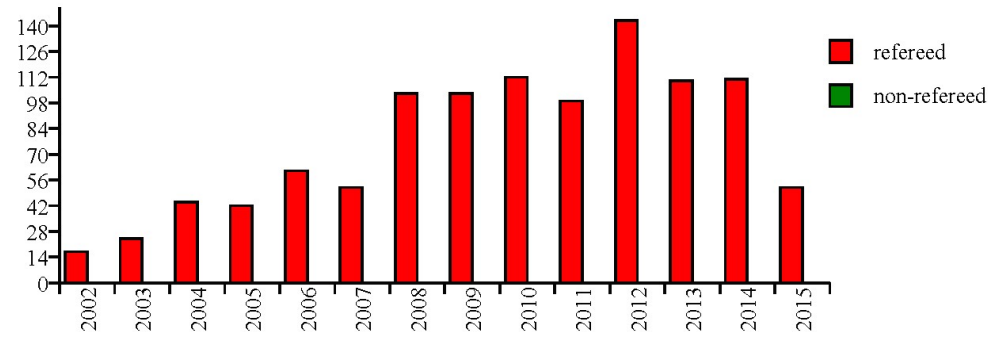
大石国立天文台天文データセンター  
センター長作成

READ10 index

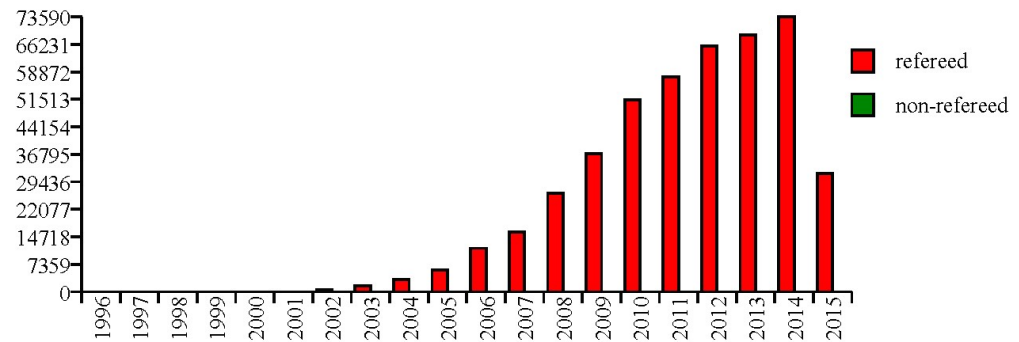
12989

12989

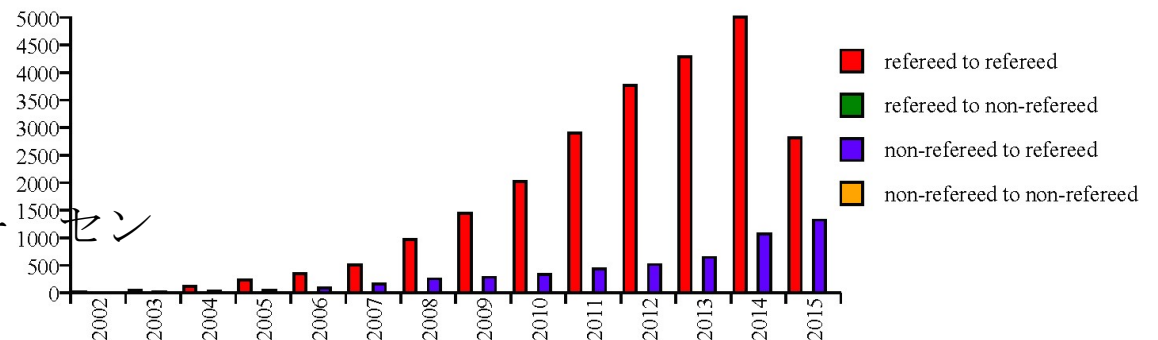
publication histogram



reads histogram

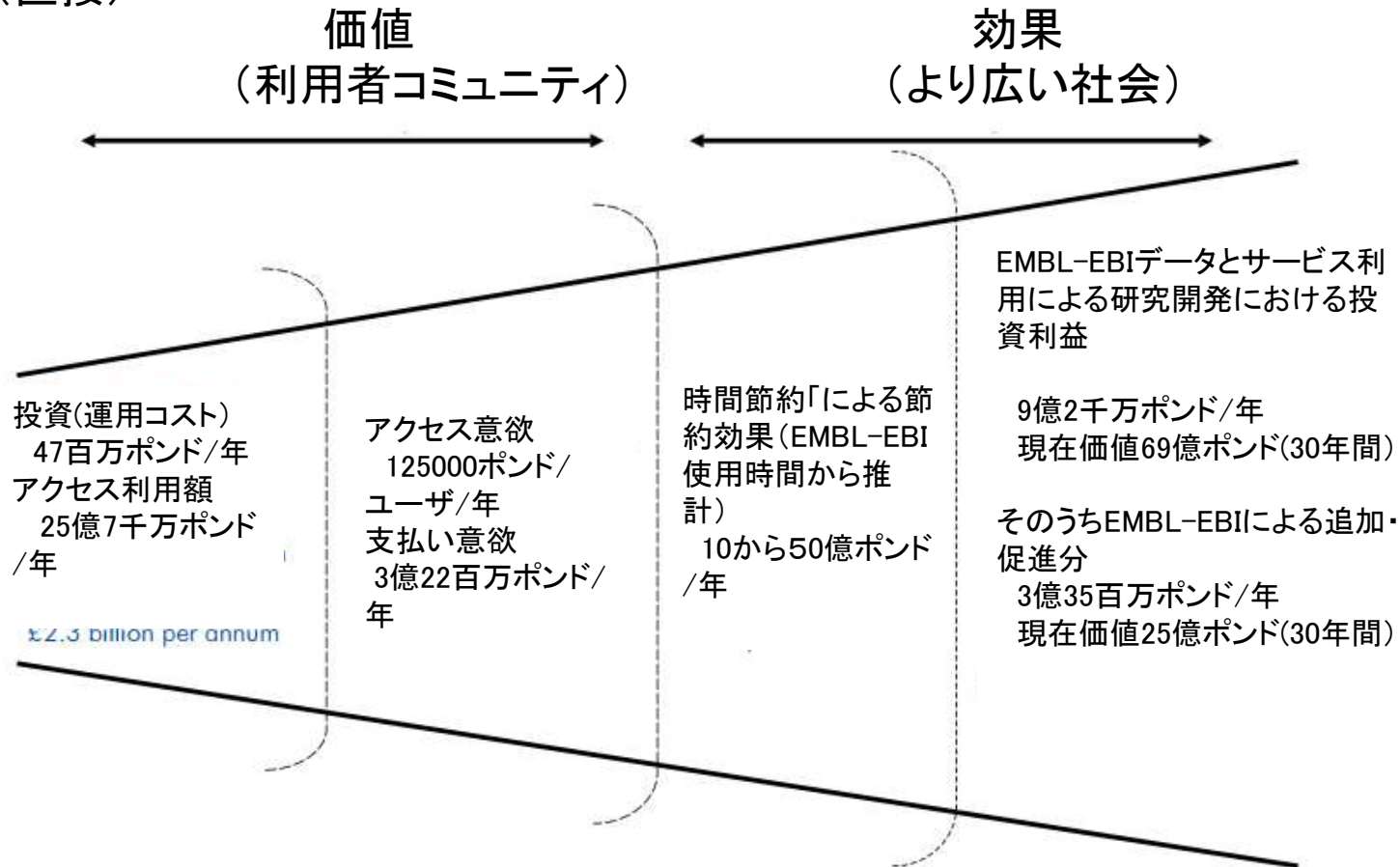


citation histogram



# EMBL-EBLデータとサービスの価値とインパクト

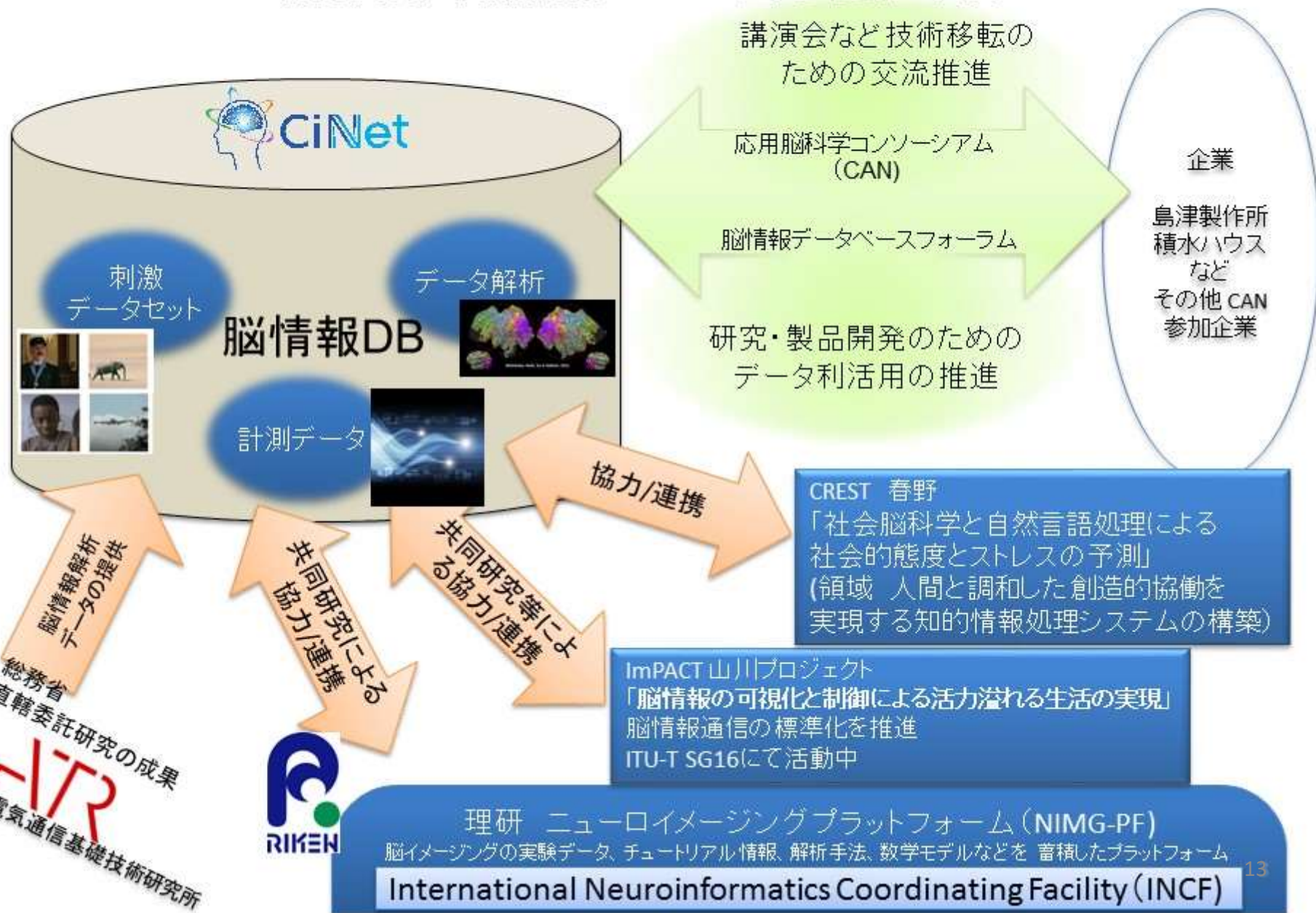
投資と利用価値 (直接)    付随価値(間接)    効率効果(推定)    投資利益(推定)





# 脳情報データベースセンターとして (NICT提供)

研究開発・製品開発コミュニティを創設・中核へ



講演会など技術移転のための交流推進

応用脳科学コンソーシアム (CAN)

脳情報データベースフォーラム

研究・製品開発のためのデータ利活用の推進

企業

島津製作所  
積水ハウス  
など  
その他 CAN  
参加企業

刺激  
データセット

データ解析

脳情報DB

計測データ

協力/連携

脳情報解析  
データの提供

共同研究による  
協力/連携

共同研究等による  
協力/連携

CREST 春野  
「社会脳科学と自然言語処理による社会的態度とストレスの予測」  
(領域 人間と調和した創造的協働を実現する知的情報処理システムの構築)

ImPACT 山川プロジェクト  
「脳情報の可視化と制御による活力溢れる生活の実現」  
脳情報通信の標準化を推進  
ITU-T SG16にて活動中

理研 ニューロイメージングプラットフォーム (NIMG-PF)  
脳イメージングの実験データ、チュートリアル情報、解析手法、数学モデルなどを蓄積したプラットフォーム  
International Neuroinformatics Coordinating Facility (INCF)

総務省  
直轄委託研究の成果  
ATR  
国際電気通信基礎技術研究所



# 日本学術会議からの提言

(1) 研究分野を超えた研究データの管理およびオープン化を可能とする研究データ基盤の整備

コスト負担やサイバーセキュリティ対策

↓

研究データ基盤：研究コミュニティ毎のオープン・クローズのデータ戦略に基づく研究データのリポジトリの整備・運用、廉価なデータ保存を可能とする仕組みも必要である。

(2) 研究コミュニティでのデータ戦略の確立

対象となるデータの見極め、占有期間の設定、データのオープン範囲の決定、データ解析ツールの包含、などのオープン・クローズ戦略とガイドラインおよびデータの取捨選択の明確な基準は、各研究コミュニティが検討すべき

(3) データ生産者およびデータ流通者のキャリア設計

# オープンサイエンスを推進する3つの要素

	ボトムアップ: 研究者	トップダウン: 政府・助成機関
データ公開	公開により研究成果の信憑性を向上。従来知見の再利用。	公的研究資金を受けた研究成果の公開。
データ管理	苦勞して取得した研究成果を安全な環境で保全。	研究成果のもととなった研究資料は10年間の保存を原則。
データ検索	研究活動の効率化。分野融合型研究の加速。	市民科学。産学連携。国際的な発信力の強化。

研究データを、

- 公開しなければならないのは研究者の責任。
- 保全する環境を整備するのは研究機関の責任。
- 流通させるサポートをするのは図書館の責任。

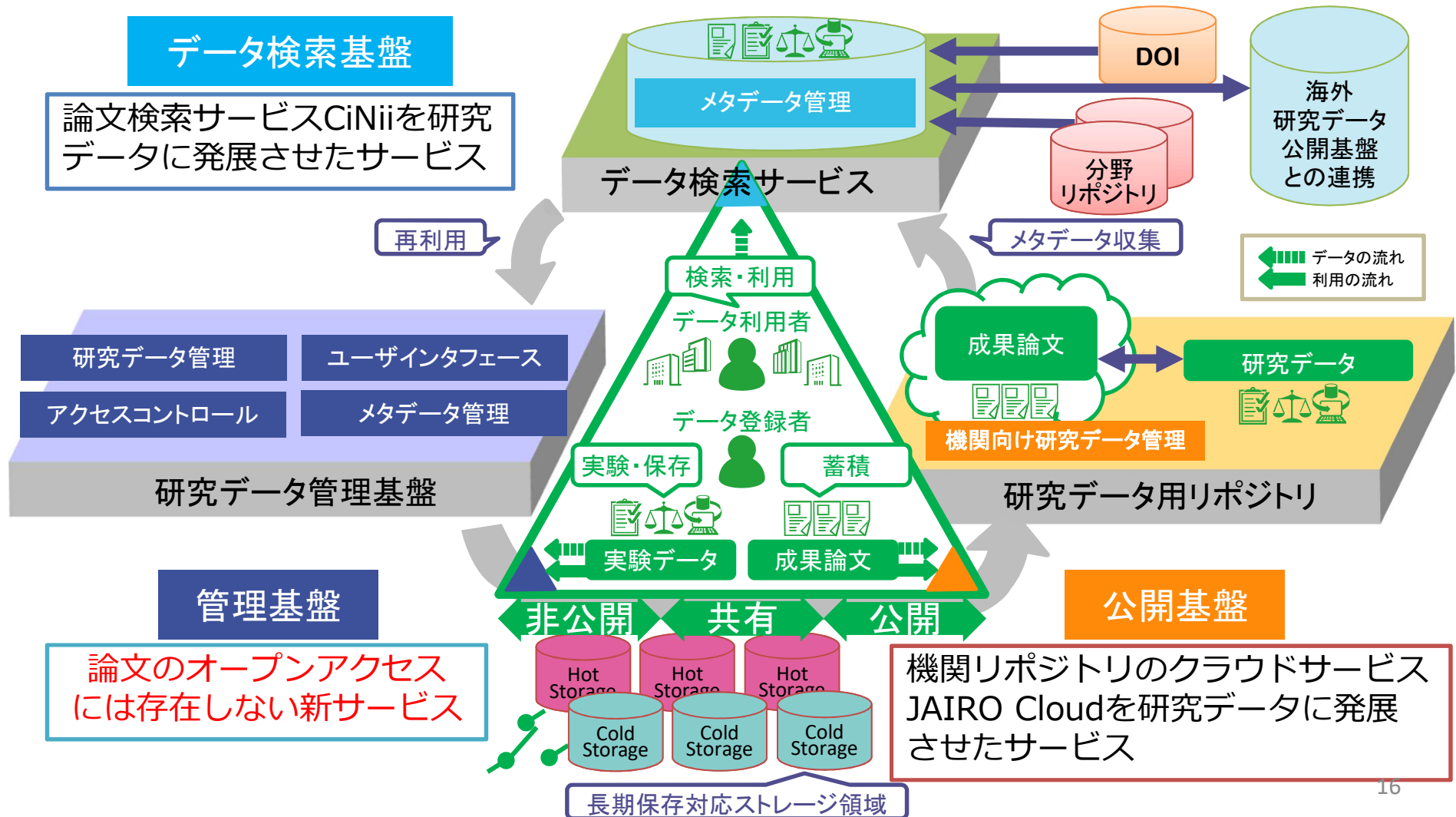
(ICSU-IAP-ISSC-TWAS working group, Open Data in Big Data World, 2015年12月 より改変)

これらのニーズや要求を実現するためのシステム基盤が必要



# 研究データ基盤

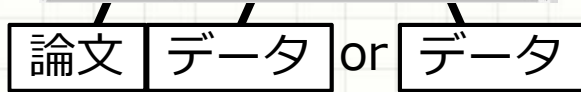
これまでに研究データの共有や公開の機会がなかった分野の研究者も活用できる基盤として学際的なオープンサイエンスの実現を目指す





# 研究データ基盤の特徴

**新規** データ管理基盤



**新規**



**新規**



**【既存】** 通常の研究データベース整備事業は、公開準備が整った研究データを公開するPFを提供。

**【新規】** データ管理基盤は、研究データの日々の管理や公開が容易にできる汎用的な機能を提供。個人やラボ、共同研究者間でのクローズドな研究データ管理が主目的。既存のDBが存在する分野の研究者にとっても有用な基盤。

**【既存】** 通常の研究データベース整備事業は、特定の分野の研究コミュニティを対象に設計。

**【新規】** データ公開基盤は、研究データ公開のための汎用的なリポジトリとして整備。大学や研究所機関や、新しい分野における利用を対象。既存のDBからの移行も可能な柔軟性を実現。オープンサイエンスの裾野を広げるために必須な基盤。

**【既存】** これまでにも既存のDBを横断的に検索できるサービスは存在したが、特定の分野を対象。

**【新規】** データ検索基盤は、分野を超越した研究データの検索サービス。論文や研究者、研究プロジェクトとの関連情報も対象。国際的な研究データ検索基盤とも連携し、国や分野を問わない研究データの発見とアクセスを実現。

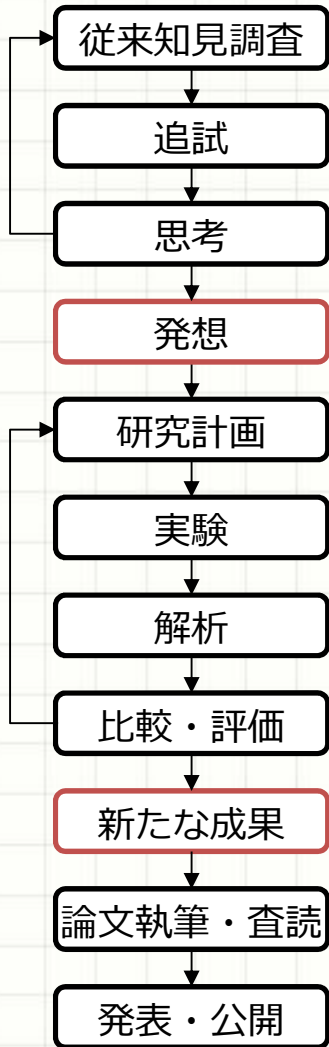
# 研究データ基盤の役割

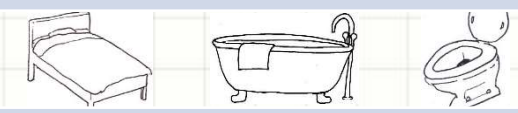
研究データ基盤  
がなければ

- 研究の鈍化
- セキュリティ対応
- 知財の海外流出  
などへの懸念大

- 論文から推測して試行錯誤
- メール添付 USBドライブ
- SNS、メール
- HDDが壊れた
- 無償共有サイト SNS、メール
- 論文から読取
- 無償共有サイト SNS、メール
- 無償動画サイト ブログ

## 研究フロー



研究者のニーズ	研究データ基盤
論文などの従来知見を分野を限定せず <b>横断的に検索</b> したい。	データ検索基盤
論文のもととなる <b>実験データ</b> や <b>実験方法</b> についても詳細を知りたい。	データ検索基盤
共同研究者間で、 <b>研究の素材</b> を簡単に <b>共有</b> しコミュニケーションをとれる環境が欲しい。	データ管理基盤
	
公募申請ドキュメントを <b>バージョン管理</b> しながら共同研究者で共有できる環境が欲しい。	データ管理基盤
セキュリティ対策やプライバシー保護にも対応して <b>研究データ</b> や <b>ラボノート</b> を <b>安全に保存</b> できる環境が欲しい。	データ管理基盤
<b>解析プログラム</b> や <b>データの処理結果</b> を共同研究者で安心かつ簡便に <b>共有</b> できる環境が欲しい。	データ管理基盤
他の研究成果と比較するための <b>従来知見の解析データ</b> が欲しい。	データ検索基盤
論文の共著者と、分担執筆のための <b>バージョン管理</b> 、図表や <b>エビデンス</b> となる <b>データ</b> を一括して <b>管理</b> できる環境が欲しい。	データ管理基盤
論文とその根拠となる <b>エビデンスデータ</b> も <b>公開</b> し、研究成果を研究コミュニティに広めたい。	データ公開基盤

# 海外の状況と日本の強み

- 欧米の現状

- オープンサイエンスに関するポリシーの制定、DMPの普及
- 必要な基盤の先行開発
- 主要な分野でのケーススタディ

従来の個々のインフラ系のプロジェクトが乱立しており連携はこれから日本より先行しているが広範な分野・研究者への普及はこれから

- 日本の強みは共通基盤の整備・普及力

- 日本の機関リポジトリの普及は世界でも注目
  - 各国で日本のような基盤整備への展開を模索

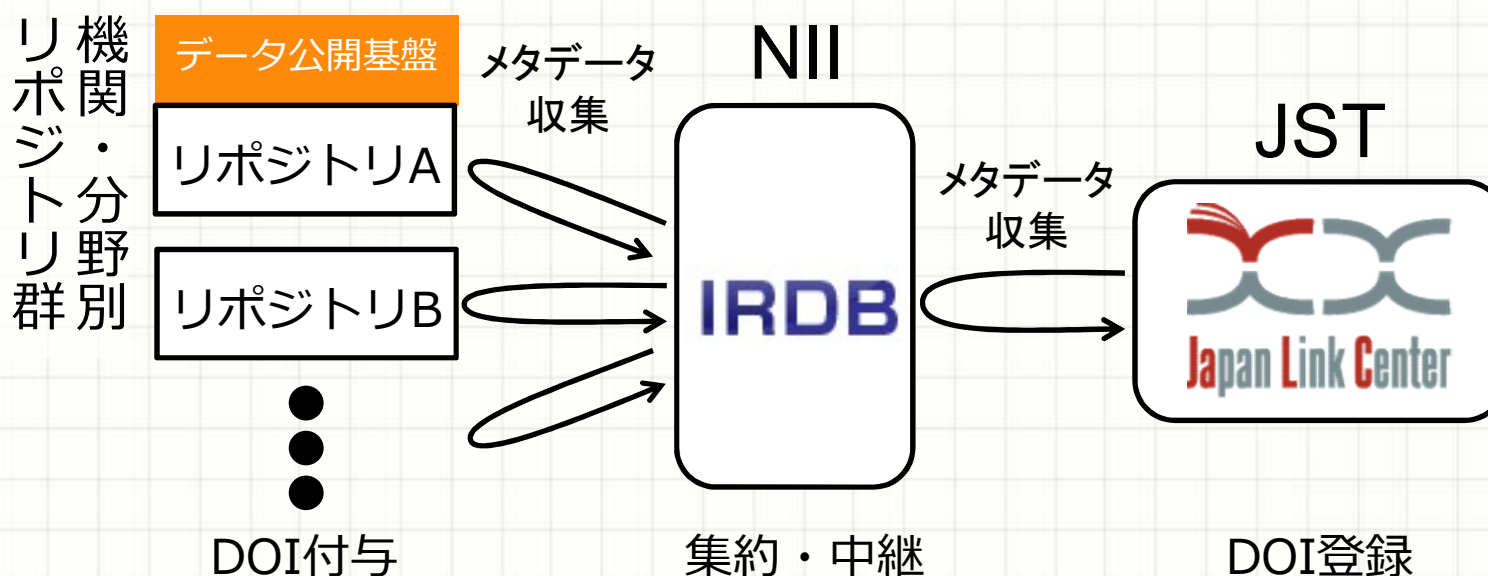
欧米で開発された基盤の活用 + JAIRO Cloudの成功パターン  
研究データ基盤の共通化により研究分野間の連携を一気に促進

# 補足資料



# 研究データ基盤とJaLCの関係

- 研究データ基盤はデータ公開基盤でDOIを付与
- JaLCはデータ公開基盤で付与されたDOIを登録



1. JaLCの正会員であるNIIは、データ公開基盤利用機関を準会員としてとりまとめる。
2. NIIは、準会員をJaLCに申請し、JaLCは準会員にDOIのPrefixを割り当てる。
3. DOIのSuffixは、データ公開基盤が自動的に(手動割振機能もあり)に付与する。
4. NIIは付与されたDOIを集約するIRDBを運用し、JaLCの登録システムに中継する。

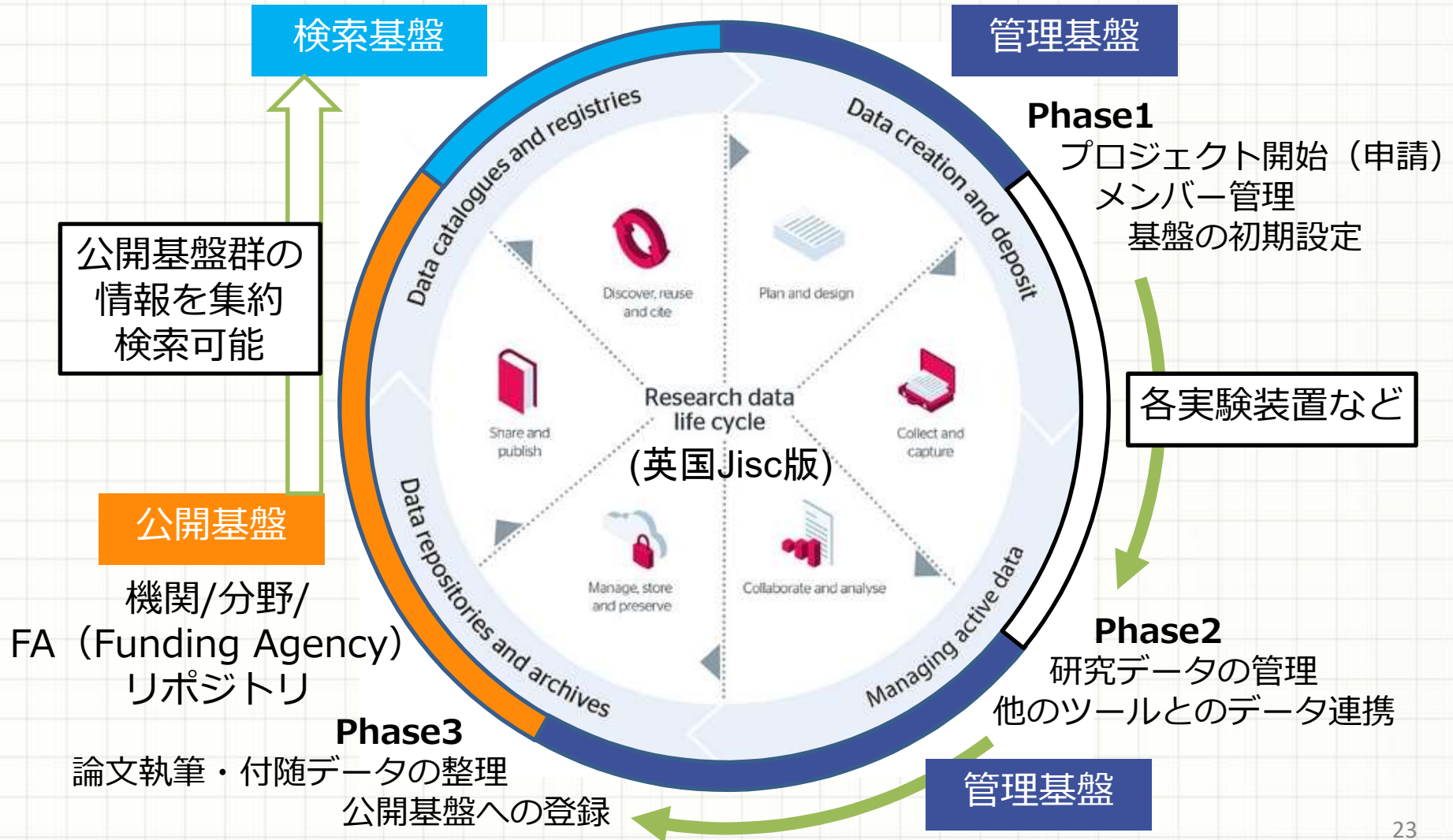
# 欧州の抱える課題

- オープンサイエンスを実現するために、従来からあるインフラ系の取り組みを連携させるEuropean Open Science Cloudという枠組みを模索
  - GÉANT(ネットワーク、セキュリティ、認証)
  - EUDAT, EGI, PRACE(コンピューティング、データ管理),
  - OpenAIRE(研究成果検索、関連付け、評価)
  - C.f. 米国は個別の活動が主体。機関リポジトリの活動さえもバラバラ
- オープンサイエンスの実現には先端的な研究・研究者だけではなく、ロングテールのサポートが不可欠であることは共通の認識
  - そのためには、まず、各インフラ系プロジェクトの連携が不可欠

しかし

- FP7のころからの歴史的な背景が長く、個々の取り組みが先行してきたため、順調には進んでいない状況
  - まずは認証機能の共通化を進めようとしている段階
  - 連携を目指しDigital Infrastructure for Researchの第1回目の会議開催(9月)

# 一般的な研究データライフサイクルと 研究データ基盤との関係



# Phase1 管理基盤によるプロジェクト開始

## — 研究過程におけるデータ管理基盤の使い方の例 —

データ管理基盤以外の他のサービスとも連携

- MLサービス
- Wikiサービス
- スケジュール調整サービス
- ファイル転送サービス



**データ管理基盤**

機関リポジトリ推進委員会 > 研究データTF

Contributors: Eriko Amano, Hayahiko Ozono, Yasuyuki Minamiyama, Misumi Taro, yu.rishibono, Kazu Yamaji, Shota Maeda, Hiroshi Maruyama, Tadaoike Taguchi  
 Date created: 2016-05-28 08:53 AM | Last Updated: 2016-08-29 01:30 PM  
 Category: Project  
 Description: no description

Wiki: 元のプロジェクトがスペースと統合し、その中で、プラットフォーム一つ作りました。まずはdropboxなどの連携をします。(南山)

Citation: osf.io/ud55v

Components: No components have been added to this project.

Tags: add a tag

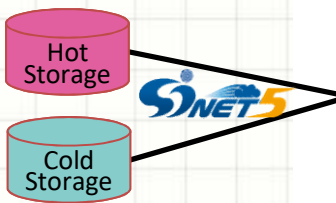
Files: Click on a storage provider or drag and drop to upload

Name	Modified
機関リポジトリ推進委員会 > 研究データTF	
- Dropbox: OSF	
- test	
- OSF Storage	
- ROM_training_tool JP	

Recent Activity: Eriko Amano added file ROM\_training\_tool JP (研究データTF) to OSF Storage in 機関リポジトリ推進委員会 > 研究データTF. 2016-08-29 01:30 PM



**情報基盤センター**  
クラウドストレージ等の  
契約・提供



1. プロジェクトの発案
2. 初期メンバーでVOを作成
3. ML等で初期的な議論
4. メンバーの拡充

管理基盤上で  
プロジェクト領域を作成

5. 申請書類等の共有
6. DMPの作成
7. 申請書類の完成

応募

8. ヒアリング資料等の共有

採択

9. プロジェクト開始

※外部資金プロジェクトだけではなく、ラボの管理にも活用可能



# Phase2 管理基盤によるデータの管理

## — 研究過程におけるデータ管理基盤の使い方の例 —

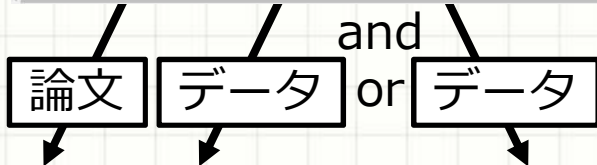
各種クラウドタイプのストレージ接続だけでなく、外部の文献管理ツール（Mendeley）やソースコードリポジトリ（GitHub）、データ解析環境（R）やそのノートツール（Jupyter Notebook）などとも連携し研究者の利便性を向上

Name	Size	Version	Downlo...	Modified
Replication Studies				
OSF Storage				
Analysis notes.txt	3.2 kB	7	0	2016-06-13 05:22 PM
MarkdownNotes.txt	110 B	1	0	2016-04-13 05:33 PM
METHOD_to_select_papers.txt	1.8 kB	1	94	2013-12-11 04:48 PM
papers_and_keywords.xlsx	12.5 kB	1	55	2014-01-13 04:28 PM
Replication_Study_functions.R	8.9 kB	11	2588	2016-07-21 05:50 PM
Statistics calculators.xlsx	52.4 kB	1	24	2015-02-01 07:49 PM
Studv 1: Polisen et al. 2010. Nature				

- 共同研究者間でオリジナルデータからの派生データや差分データを体系的に管理
- データや処理方法に関するアノテーションやメタデータを管理
- ホットストレージとコールドストレージを使い分ける機能により肥大化するストレージ領域を効率的に管理

# Phase3 論文執筆・付随データの公開

## — 研究過程におけるデータ管理基盤の使い方の例 —



機関リポジトリ



分野/FAリポジトリ

公開・非公開例

- ・ 非公開データはエンバーゴを設定
- ・ 非公開データはメタデータのみを登録
- ・ 非公開実データは管理基盤上でフリーズ

1. 論文原稿のバージョン管理
2. Mendeley等の外部ツール (add-on) を利用し引用情報等の管理
3. 論文と紐づく根拠データの管理・整理
4. 図表等の管理

投稿

5. 査読の返事を編集・共有

採録

6. FAのポリシーに応じて公開基盤に登録
7. 出版社のOAポリシー、DMP等に従い論文・付随データを公開基盤に登録
8. 図書館員やキュレータによる統制語の付与、メタデータ記法の標準化
9. DOIの付与

公開

※OAの手段や査読より手順は異なる