

# NIIにおける オープンサイエンスへの 取り組みと課題

平成28年1月22日  
国立情報学研究所

# SINET5 — 学術情報基盤の新展開

平成28年4月より、クラウド、セキュリティ、コンテンツも強化した**SINET5**を運用開始

## 大学などの学術研究・教育活動の連携・推進

### 学術情報の公開・共有

- ◆ 学術情報流通と オープンアクセスの推進
- ◆ 大学の 機関リポジトリ 拡充の推進



### 大学間連携支援

- ◆ 仕様統一したシステムによる 大学間連携、各種資源の相互利用の促進



### クラウド活用支援

- ◆ クラウド利活用促進による 大幅なIT経費削減・研究教育環境の高度化



### セキュリティ強化

- ◆ 暗号技術活用による 情報の保護、安全な認証
- ◆ ネットワーク機能連携による サイバー攻撃対策



### 学術情報ネットワークの構築・運用

- ◆ 国内回線 全国100Gbps化
- ◆ 海外 (米国・欧州・アジア) との高速接続
- ◆ 多様化するニーズに応えるSDNなどの 最新ネットワーク技術の導入



# 議論の場

- 文部科学省学術情報委員会(西尾主査)
- 内閣府 オープンサイエンス推進に関する  
フォローアップ検討会(有川主査、喜連川副)
- 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(土井3部副部長)

# Open Science

STI Horizon 2015 Vol.1 No.1

<http://doi.org/10.15108/stih.00004>

## 国立情報学研究所 所長／東京大学生産技術研究所 喜連川 優 教授インタビュー： シリーズ「オープンサイエンスの展望」の開始によせて

聞き手：科学技術動向研究センター 上席研究官 林 和弘

科学技術イノベーション政策において、イノベーションを生み出す仕組み、環境作りは重要なテーマであり、近年、研究成果の活用・再利用によるイノベーション創出を加速する情報基盤作りとして、オープンアクセス、オープンサイエンスに注目が集まっている。特にオープンサイエンスに関しては、公的資金を得た研究成果をよりオープンにして活用を促進し、新しい科学技術や産業の可能性を構築するという文脈で、内閣府のオープンサイエンスに関する検討会の報告書が今年3月に公開された。そ

### オープンサイエンスが実現する研究の地平線

まず、オープンサイエンスが果たす役割を簡単に考えてみますと、少なくとも論文になった研究については、それに至るまでのデータをオープンにしてシェアすることにより、1. 成果を多くの人がいち早く利用できるため研究開発が加速され、2. 同様に「同じ轆を踏む」「車輪の再発明」といった手戻りを防ぐことができ、3. 研究者自身にとっては、第三者の力も借りながら読りを効率よく見つけることがで

# Role of Open Science

- ① オープン化により情報流通を活性化し学問の進展の飛躍的加速化を目指す。
- ② オープン化により研究の重複を回避することにより研究開発投資の効率化を目指す。
- ③ 研究成果の再現性を担保し研究者を守る
- ④ 科学技術外交

# 背景

- 公的資金を入れた研究成果はオープンが当然ではないか(納税者目線)
- データの取得には多大な努力が必要となることから、研究者にクローズされる傾向が高い。
- 研究費が切れるとデータが散逸する傾向

# 対象

- 論文
- 研究データ
- プログラム
  - 容易に見えるが、現時点でそれほど容易ではない

# 求められる機能

- 情報の発信(検索可能・アクセス可能)・保存
- 永続保存
  - 長期的な保存(Digital Dilemma)
  - エビデンスとして
  - コンプライアンス e-discovery
- 活用
  - Localization(応用依存)



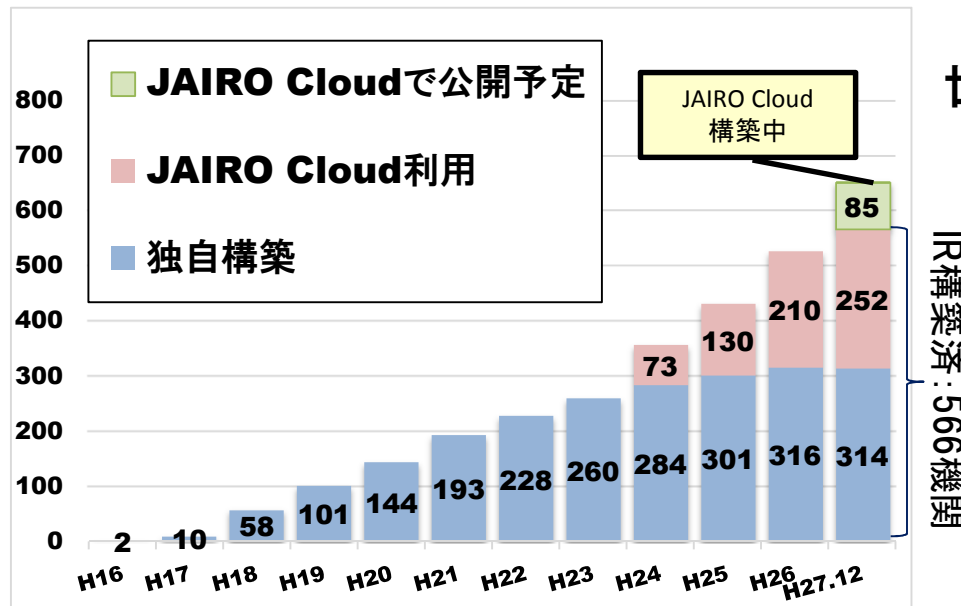
# 論文

# JAIRO Cloud

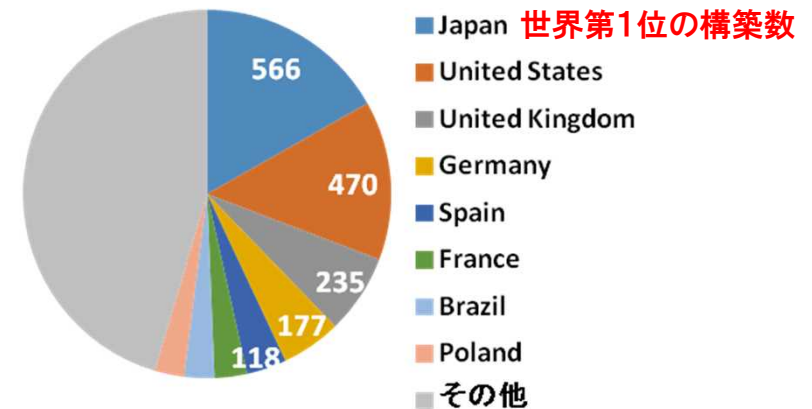
機関リポジトリのクラウドサービス(H24年度～)

- 独自で機関リポジトリの構築・運用が難しい大学等に対して、研究成果の収集・保存・発信を容易にする仕組みを提供
  - 大学でのシステム運用負荷の軽減により機関リポジトリ数の拡大
  - 必要な機能を備えたシステムをクラウド上で提供することで全体の効率化
  - 学術情報のオープンアクセスを推進

機関リポジトリ機関数の推移



世界で構築されている機関リポジトリ: 3,367

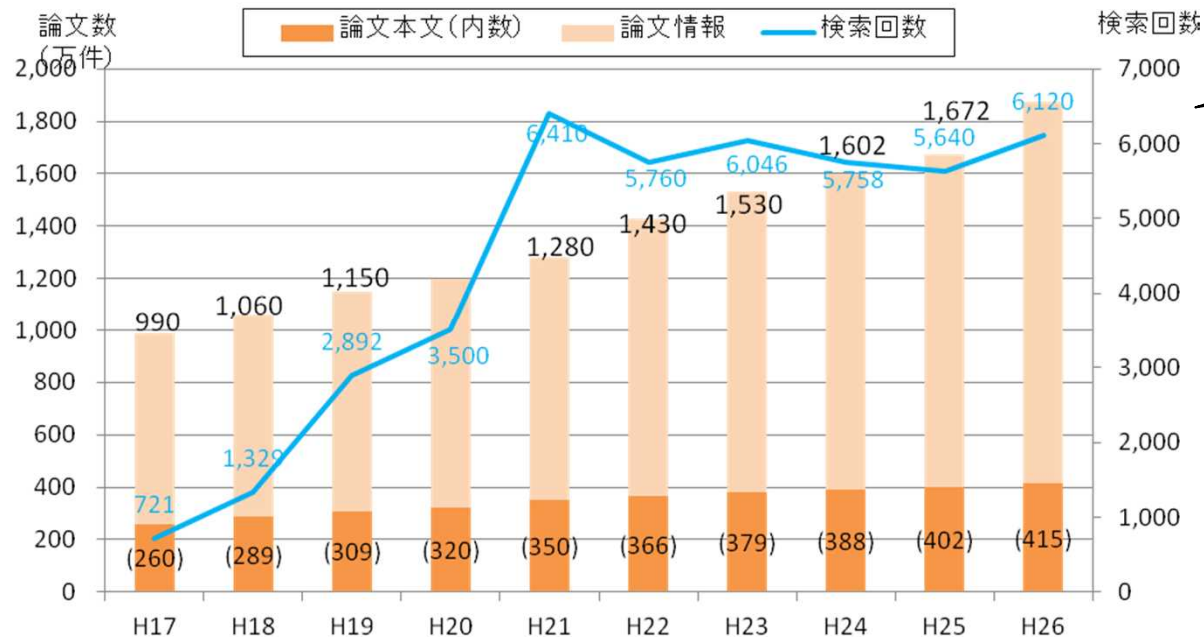


# CiNii

## 日本の学術論文を中心とした論文情報検索サービス

- 1,873万件の膨大な論文データを収録し、うち415万件は本文PDFファイルあり
- 学協会誌や研究紀要の論文本文の閲覧やリンクにより他サイトの本文も参照可能
- 主な収録データベース: 学協会刊行物(国立情報学研究所)、雑誌記事索引データベース(国立国会図書館)、J-STAGE(科学技術振興機構)、機関リポジトリ(各大学等)

### 収録件数と検索回数



Pageviewは  
年間2億回

# 識別子 (ID) 連携

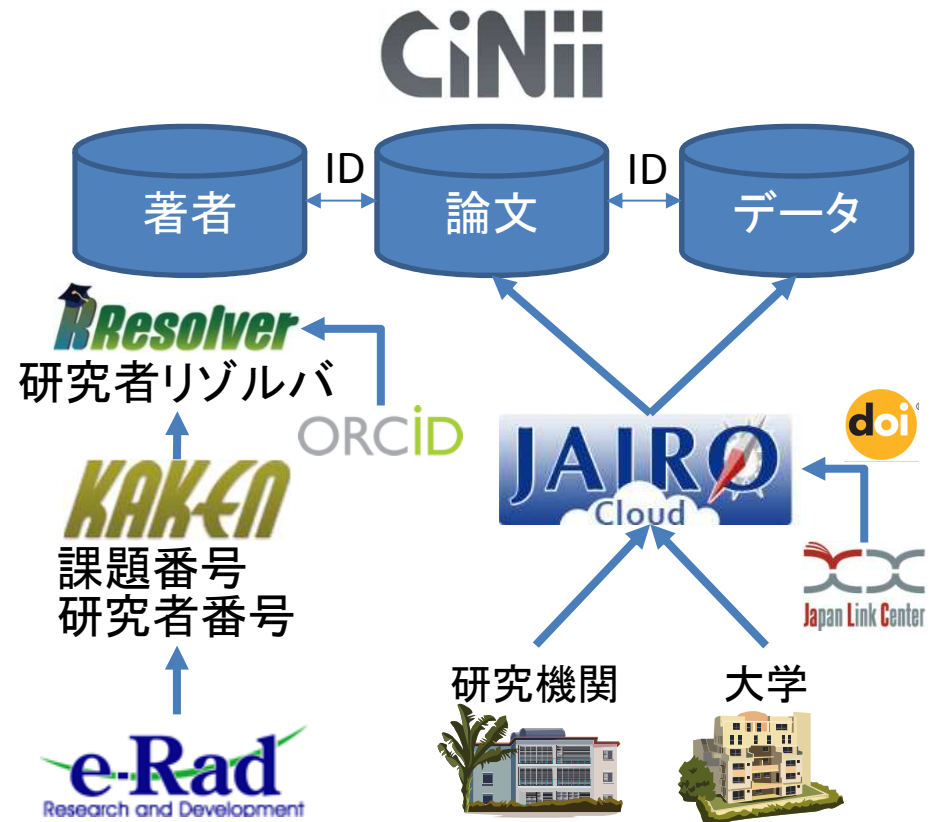
内外の識別子の活用による文献、データ等の有機的な連携基盤

## 現在の取組み

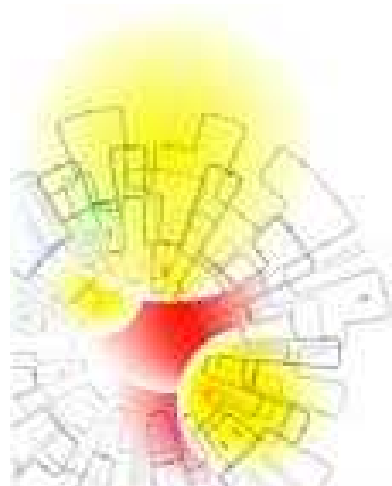
- 機関リポジトリにおけるDOI付与の支援
- JAIROによる研究者識別子の利用促進
- 研究者リゾルバによる研究者識別子の連携基盤提供
- KAKENによる研究者、課題、業績の結びつけ

## 今後の展望・課題

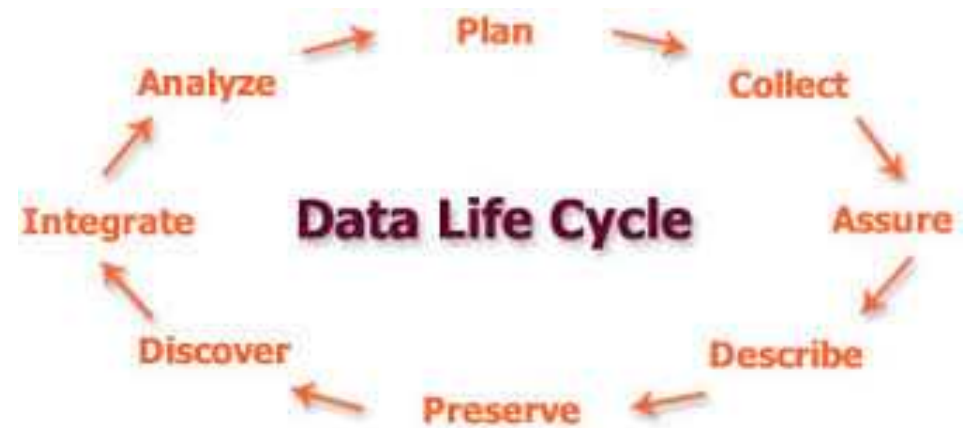
- KAKENおよびJST成果DBの連携による国内の研究成果情報の一元的提供
- 研究者リゾルバをバックエンドにNIIのサービスを各種識別子で統合的に連携
- 内外の標準化団体との連携の強化と標準化への参画
- 国内の大学・研究所等への識別子利用の普及のためのインフラ、ツール提供の強化
- 学認との連携によるデータ品質の向上



# 研究データ



**RESEARCH DATA MANAGEMENT**  
Dealing with Data



# アカデミック クラウドの利用

## ②研究データの保管・管理

- 研究データの保管・管理は、研究データの公開を進めるための前提であり<sup>11</sup>、特に、公的研究資金による研究プロジェクト等の終了後も、研究データが利活用可能な状態で適切に管理されるよう、プロジェクト等の規模（例えば、一定額以上の規模に係るものを対象とする）やその目的及び分野の特性等に応じ、データ管理計画を作成し計画に従った管理を行うことが必要となる。
- 関係機関は、研究データの保管に係る基盤を整備するに当たって、アカデミッククラウドの活用を図ること、及び後述する研究データの公開方法に関連する基盤の整備に努め、これを国が支援していく必要がある。

### 【研究資金配分機関が行うべき取組】

- ・申請者に、必要に応じデータ管理計画の提出を求める。

### 【大学等に期待される取組】

- ・研究者のデータ管理計画の作成と計画に従った管理の実施について支援する。
- ・研究データの保管に係る基盤を整備するに当たって、情報基盤の共有や効率的な整備の観点から、アカデミッククラウドの活用を図る。

### 【NIIが行うべき取組】

- ・アカデミッククラウドの構築に当たり、フォーマットの標準化やシステム開発及び共同調達等について、大学等と連携し進める。

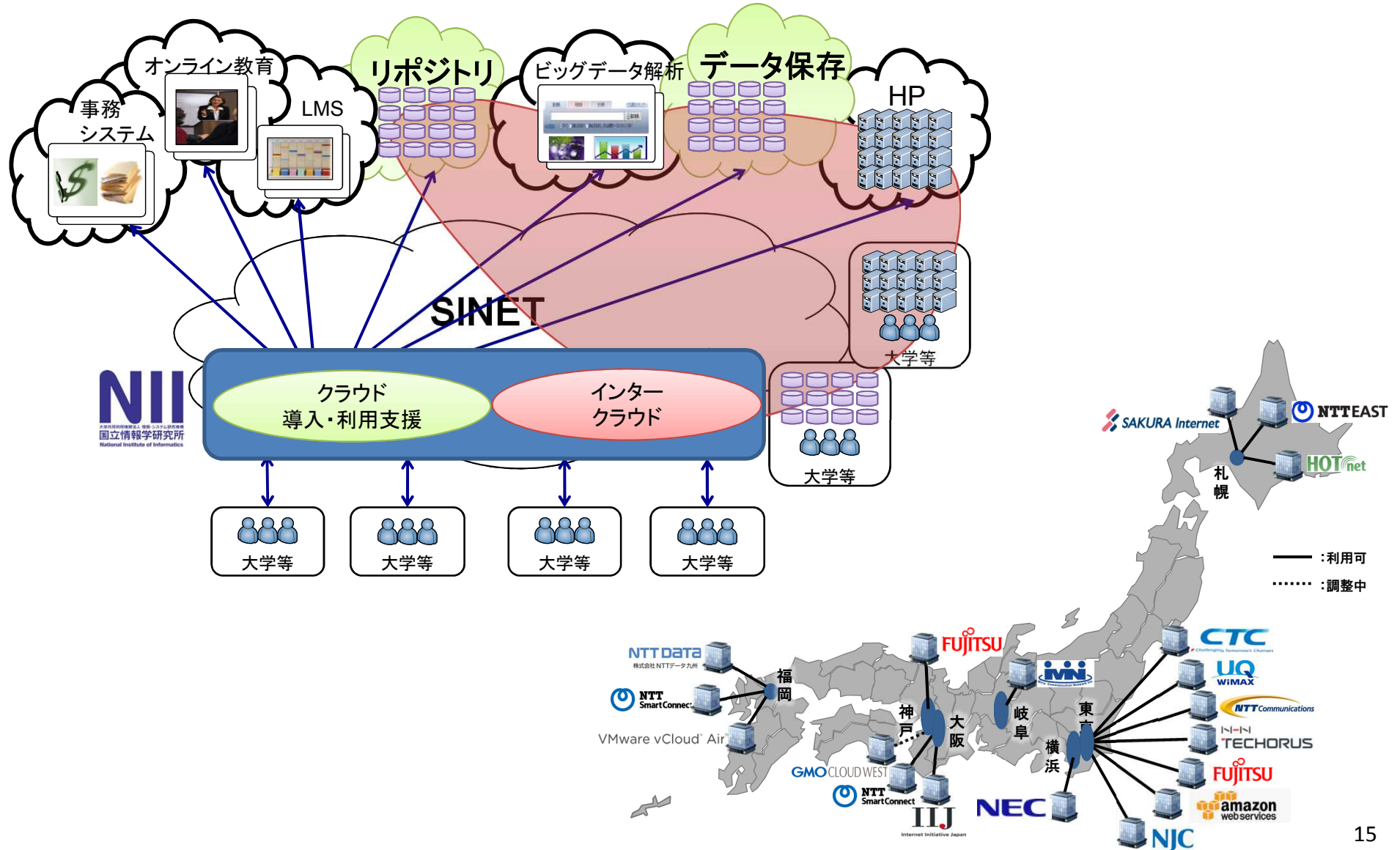
### 【国が行うべき支援】

- ・NIIと大学等が連携して進めるアカデミッククラウドの構築を支援する。

<sup>11</sup> 研究の不正行為を防止する取組に関連して、「研究活動における不正行為への対応等に関するガイドライン」（平成26年8月26日・文部科学大臣決定）に沿って、研究データの保管が各大学等で行われている。なお、この取組は、研究者に研究データを一定期間保存することを義務付けているが、広く公開することは前提にしていない。

# アカデミッククラウド

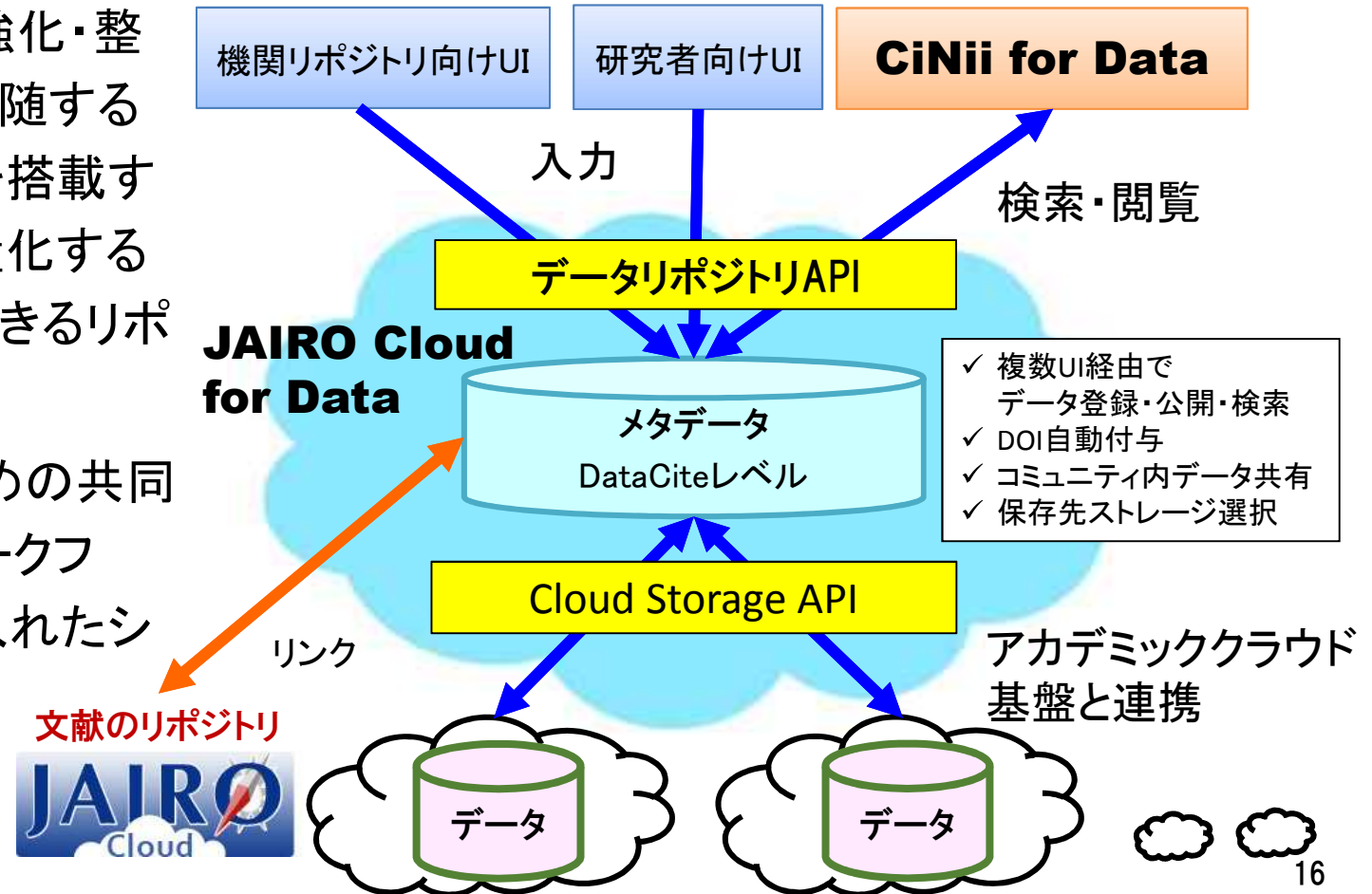
全国規模の高性能ネットワークとクラウドを一体的に利用する基盤



# Data Retrievable Repository

オープンサイエンス推進のための研究データ公開リポジトリ基盤

- JAIRO Cloudを強化・整備し、論文に付随する研究データ等を搭載する機能、大容量化するデータも格納できるリポジトリの提供
- 実証実験のための共同研究と研究ワークフローを視野に入れたシステム開発

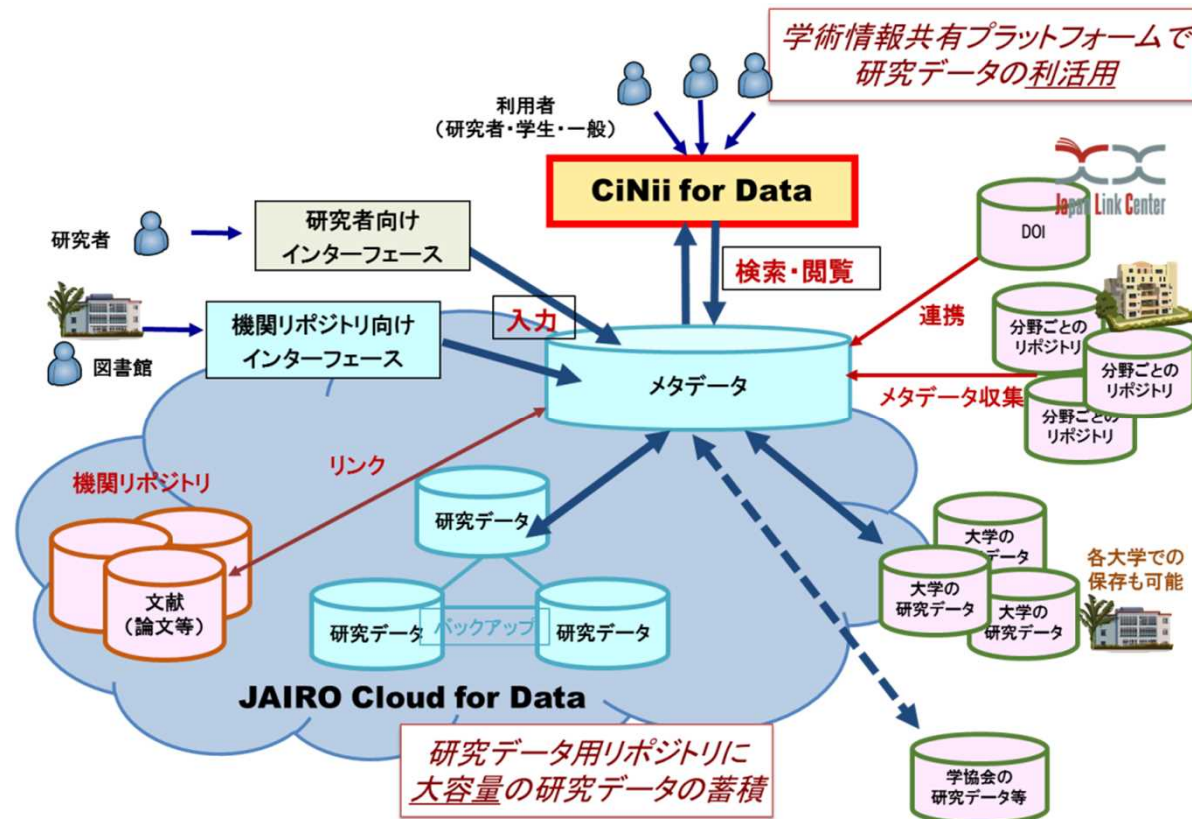




# CiNii for Data

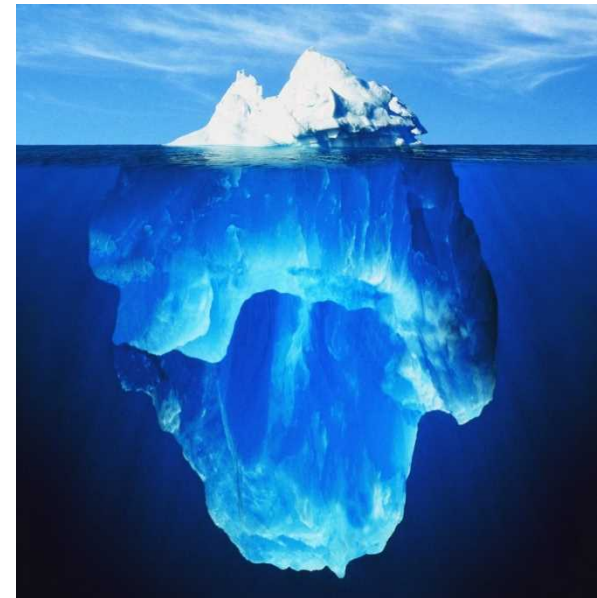
## 研究データと論文情報・研究者情報を統合した総合的発見サービス

- 論文情報と機関リポジトリの全コンテンツ、各種研究データを統合した総合的発見サービス CiNii for Dataのプロトタイプ構築および試行提供
- 研究データ利活用のユースケース調査とシステム設計へのフィードバック
- 研究データの発見・利活用を促進するための名寄せ技術の研究開発



# 永続保存サービス

- コンプライアンススペース on Cold Storage over Academic Cloud
- 普通のストレージ
  - 3700 - 4500万円/PB/年
- 永続ストレージ(Cold Storage)
  - 140 - 1600万円/PB/年 (4時間)



# Vertical Research Data Sharing

- 応用依存の利活用基盤の構築
  - 古典籍
  - デジタルシルクロード
  - DIAS
  - ゲノム解析

# 事例 古典籍

## NIIの学術情報基盤を活用した大規模な日本語の古典籍データセットの公開と共同利用

- 国文研と研究連携協定を締結し、国文研が作成したデータセットの公開を担当
- SINETの高速ネットワークを活用して350点の古典籍データ(目録と画像約63,000枚, 合計約140GB)を配信
- 公開後2か月あまりで延べ約13,000点のダウンロード(ロボット及び重複アクセスを除く)
- データセットの活用を図るためにアイデアソンを実施(2015年12月18日)

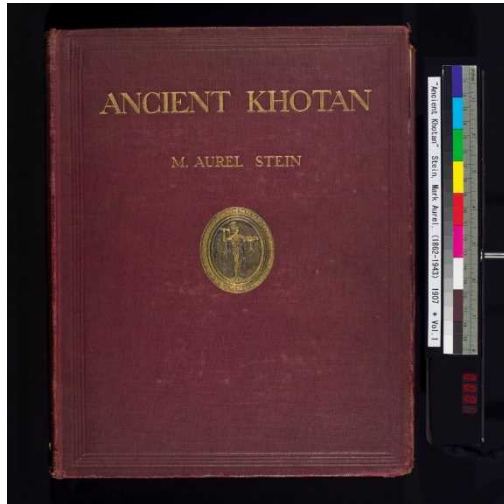
### ○データセットのダウンロード先

国立情報学研究所「情報学研究データリポジトリIDR」

URL : <http://www.nii.ac.jp/dsc/idr/nijl/nijl.html>



# 事例 デジタルシルクロード



東洋学の分野で世界的に著名な研究図書館「東洋文庫」のシルクロード貴重書をデジタル化



ウェブサイトでオープン化(アクセス数200万~600万/年)



古地図の解析に基づき、行方不明だった遺跡を再発見。



開発したモバイルアプリを活用し、古写真の撮影場所を特定。



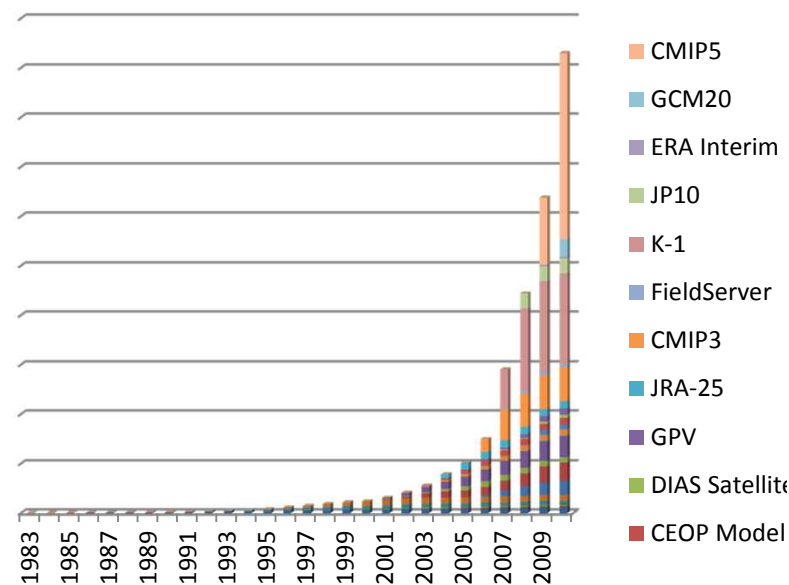
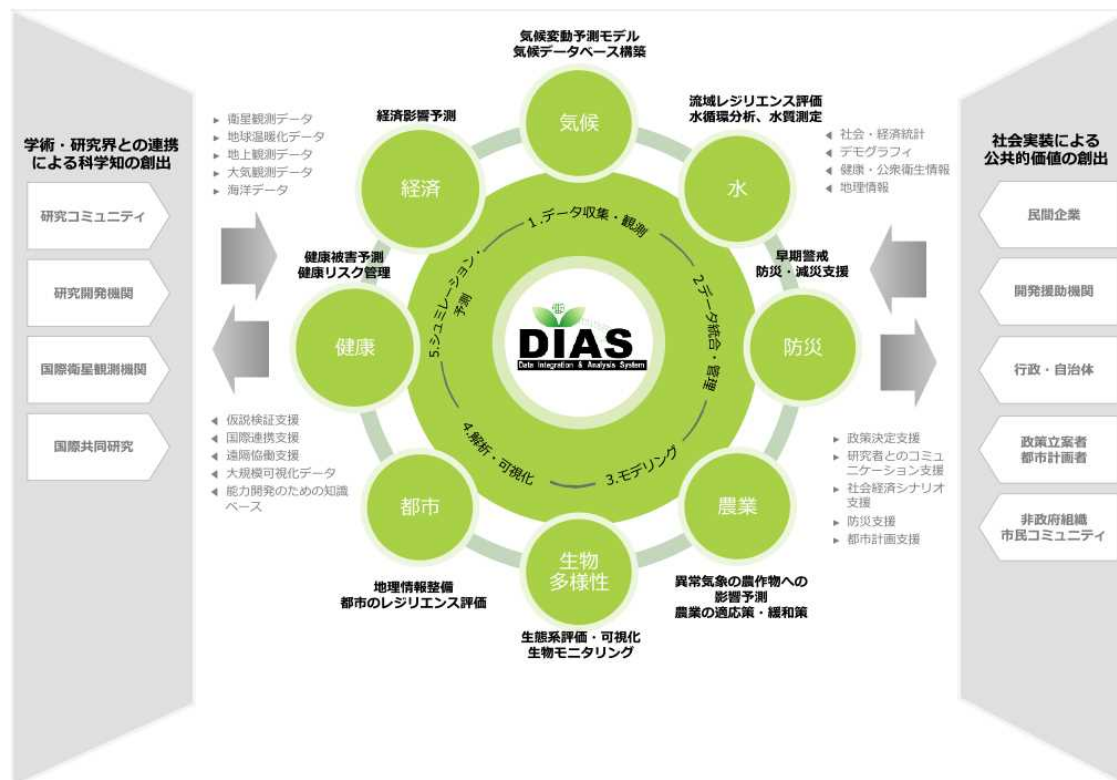
東洋文庫ミュージアムの来館者サービスとして活用。

## 境界を越えるオープンサイエンスへ

- 国を越えた利用(特にヨーロッパで人気)
- 分野を越えた利用(歴史学から人文学へ)
- 学術を越えた利用(図書館・市民・子ども)

# 事例 DIAS (25PB)

- システム資源の半分以上がNIIに存在
- データの品質保証等、先進的取り組み



# 事例 ゲノム解析(遺伝研)

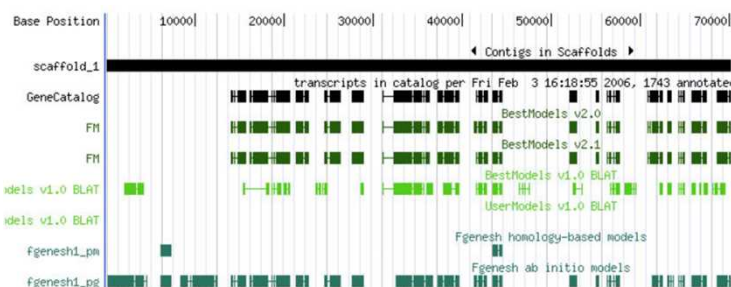
ゲノム解析ワークフロー支援の基盤開発(CREST採択課題)

DNA Sequencing



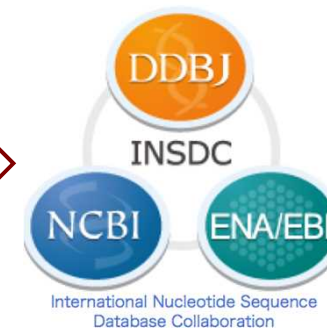
観測  
シーケンサ

Assembly, Analysis and Annotation  
(identification of functional regions,  
e.g. genes, transcription starting sites, ...)



計算  
スパコン・クラウド

Submission to DDBJ

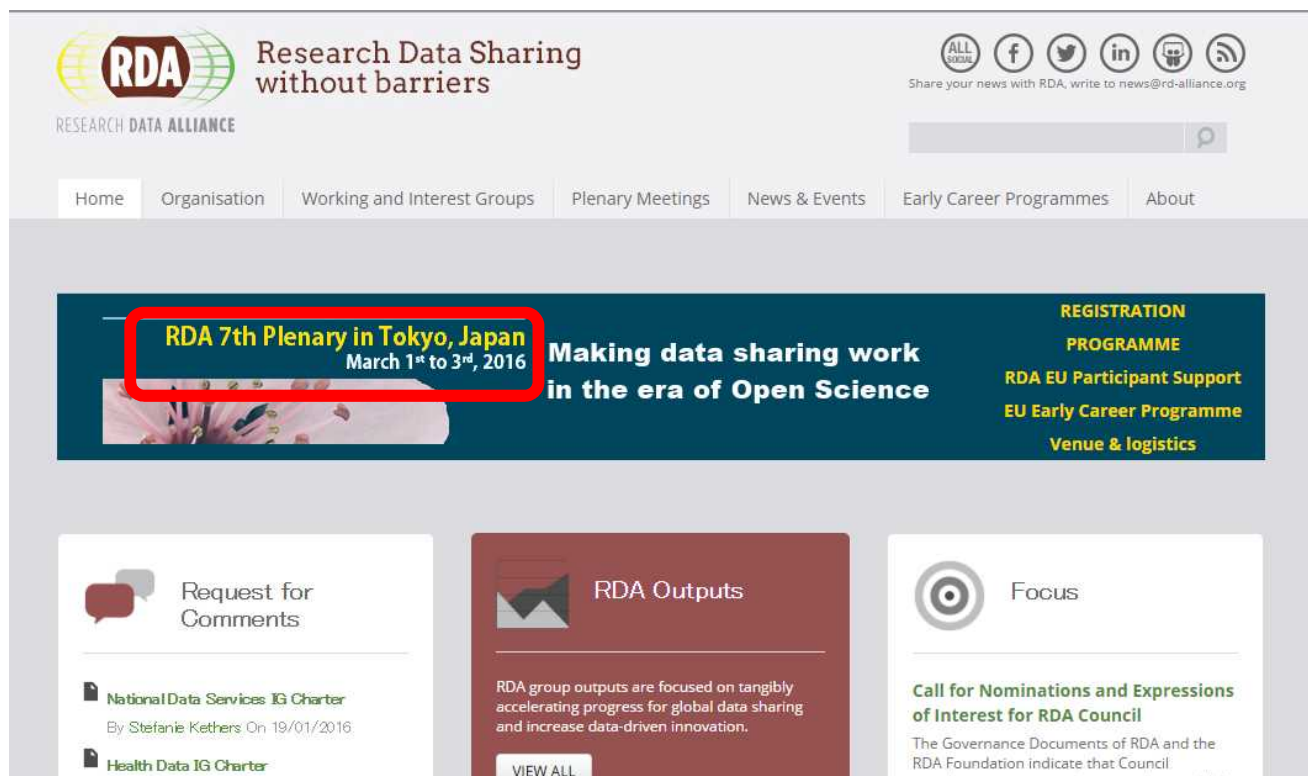


公開  
国際塩基配列DB

現状:ゲノム解析の定型的な3つのワークフローをつなぐ作業は手作業  
→将来:シーケンサでデータを読み取ったら、自動的に解析からDB登録まで可能

国立遺伝学研究所とNIIでインタークラウドを共同研究・開発中

# おわりに



- NIIは大学の皆様と一緒に環境を構築できれぽとと考えております。