

NII学術情報基盤オープンフォーラム

コンテンツトラック①「つながるオープンサイエンス基盤：
研究データの管理・公開・検索を考えるアイデアソン」

研究データ管理基盤 GakuNin RDM

2018年6月20日(水)14:00-14:20

国立情報学研究所

込山悠介

komiyama@nii.ac.jp

研究データ管理の必要性

研究推進（オープンサイエンス）

- 時間や資源を節約し効率的に研究
- データセキュリティの強化とデータ消失リスクの最小化

研究倫理（研究公正）

- 研究の再現性の確立
- 研究データや研究記録の正確性、完全性、真正性、信頼性の保障

ポリシー

組織内におけるポリシー

京都大学の研究担当理事裁定制定より

【監督者等の責務】

監督者等は、研究データの保存期間を含む保存計画を作成し、当該監督又は指導する教職員等に提示するとともに、研究データを適切に保存するための環境を整備するものとする。なお、特段の事情がある場合を除き、保存計画により定める**保存期間は、当該論文等の発表後少なくとも10年とし**、これを下回って定めてはならない。

研究費助成機関におけるポリシー

さきがけ、CREST等の公募要領より

データマネジメントプランの作成及び実施について

・・・研究代表者は、研究チームの成果として生じる研究データの保存・管理、公開・非公開、及び公開可能な研究データの運用指針を以下の項目毎にまとめた「データマネジメントプラン」を研究計画書と併せてJSTに提出していただきます。

また、**上記方針に基づいてデータの保存・管理・公開を実施していただきます。**・・・

海外大学の研究データ管理・公開サポート例

- 2011年にエディンバラ大学がイギリスで最初に機関の研究データ管理ポリシーを作成
- 3つのセクションにおいて、研究者の責任、大学の責任、共同の責任について言及
 - **研究者の責任**
 - 全ての新しい研究計画には、データの取得、管理、完全性、機密性、保存、共有、公開について記述された、研究データ管理計画書が含まなければならない。
 - **大学の責任**
 - 大学は、研究プロジェクトの実施中や終了後に、研究データを利用、バックアップ、登録、長期保存できる仕組みとサービスを提供する。
 - **共同の責任**
 - 国際的なデータサービスや研究領域のリポジトリなどに保存されたデータも、大学に登録されるべきである。
 - 助成の条件でない限り、研究データの再利用や公開する独占権を、データの再利用を可能とする権利を残すことなく、商業出版社や代理店に渡すべきではない。

2012年にはRDMロードマップを作成しRDMサービスを立ち上げ開始

イギリスの研究費助成機関のポリシー一覧

● Full Coverage
 ◐ Partial Coverage
 ○ No Coverage

Data Curation CenterによるFunders' data policiesのまとめ
<http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

Research Funders	Policy Coverage		Policy Stipulations					Support Provided			
	Published outputs	Data	Time limits	Data plan	Sharing/ access	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	◐
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	◐
Cancer Research	●	●	●	●	●	●	●	◐	●	○	●
European Commission	●	●	◐	●	◐	◐	◐	●	●	◐	●
Wellcome Trust	●	●	●	●	●	●	●	●	●	●	●

Data plan: 申請時にデータの取得、管理あるいは公開に関する計画を必要としている機関

日本の研究費助成機関でも義務化が始まっている（現在はまだ努力義務）

エジンバラ大学のRDMサービス

4つのサービスで研究データ管理をサポート

研究データ 管理計画

Data management planning



Tools you can use for data management planning activities that you need to perform before you start your project.

Storage & backup solutions



RDM file store to store your active data (data that are actively being used in your current research activities), and tools to assist you in working with your data during your project.

研究データ ストレージ

Sharing & publishing data



Tools and services you can use to describe your data (creating metadata), and deposit and manage your completed research data outputs at the end of your project.

研究データ 公開基盤

Research data management training



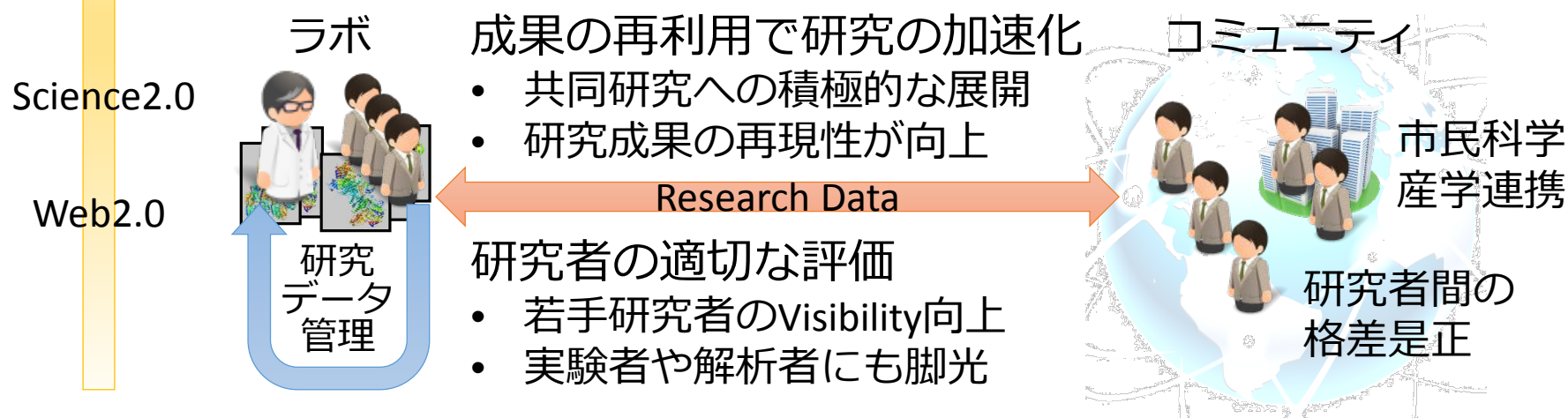
A range of training programmes on research data and research data management (RDM) in the form of an online course (MANTRA), workshops and seminars to help you with managing your research data effectively and efficiently.

研究データ 管理トレーニング

研究データ公開の必要性

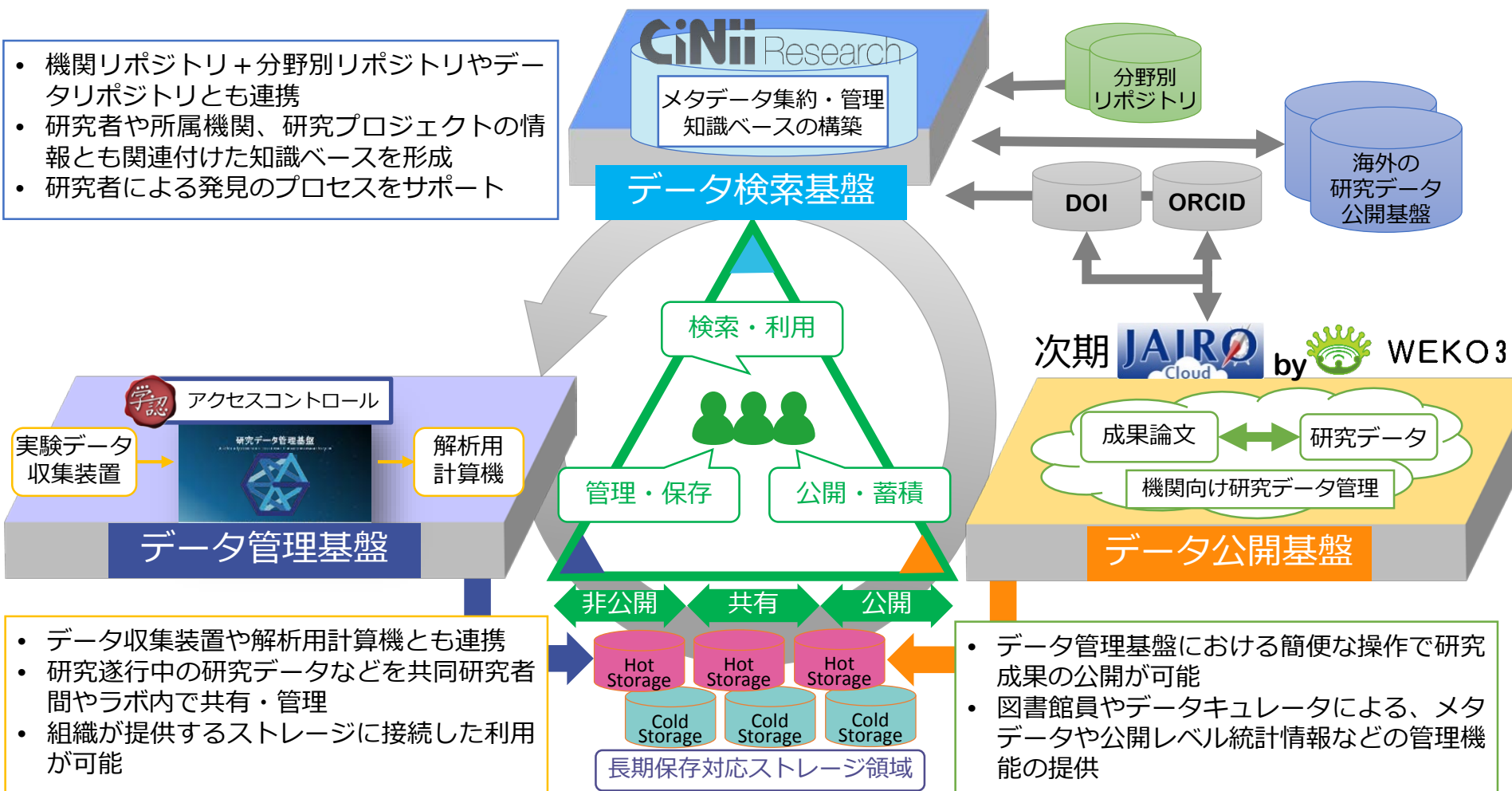
- 2013年6月 G8科学大臣会合の研究データのオープン化声明への調印
 2015年3月 内閣府のオープンサイエンスに関する報告書
 2016年1月 第5期科学技術基本計画のオープンサイエンスへの言及
 2016年7月 日本学術会議のオープンサイエンスに関する提言

オープンサイエンス : オープンアクセス
 : オープンリサーチデータ



オープンサイエンス対応 – 研究データ基盤

- 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

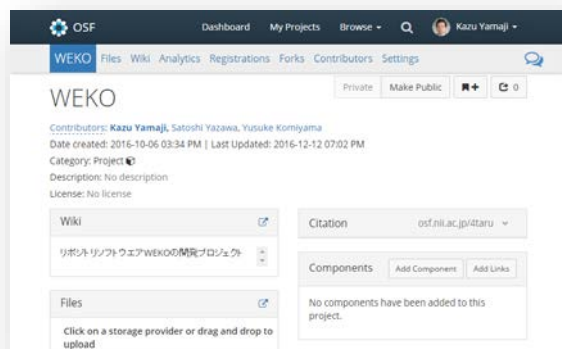


- データ収集装置や解析用計算機とも連携
- 研究遂行中の研究データなどを共同研究者間やラボ内で共有・管理
- 組織が提供するストレージに接続した利用が可能

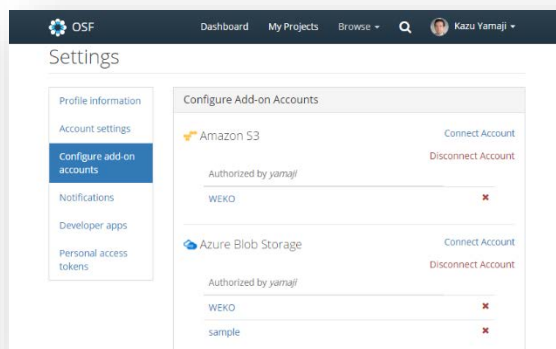
- データ管理基盤における簡便な操作で研究成果の公開が可能
- 図書館員やデータキュレータによる、メタデータや公開レベル統計情報などの管理機能の提供

データ管理基盤 GakuNin RDM

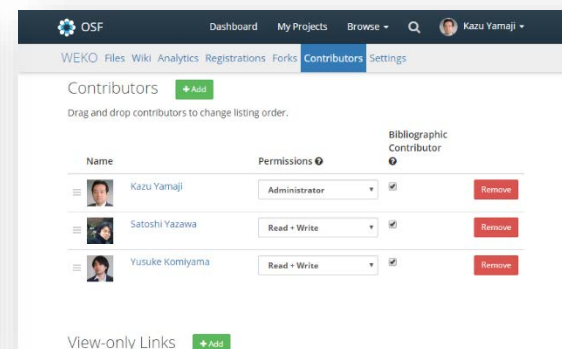
研究プロジェクト単位で
ファイルなどを管理



豊富なアドオンで必要な
クラウドストレージを接続



学認と連携しvo(仮想組織)
メンバーでファイルを共有



フロントエンドはNIIが提供

機関毎に準備

研究データ管理基盤

クラウドストレージ



NIIストレージ

大学・研究機関毎の既存のクラウドストレージの
事情に合わせてプラグインをカスタマイズ

NII提供の最小限の
デフォルト領域



パブリッククラウド
(プロバイダーDC)

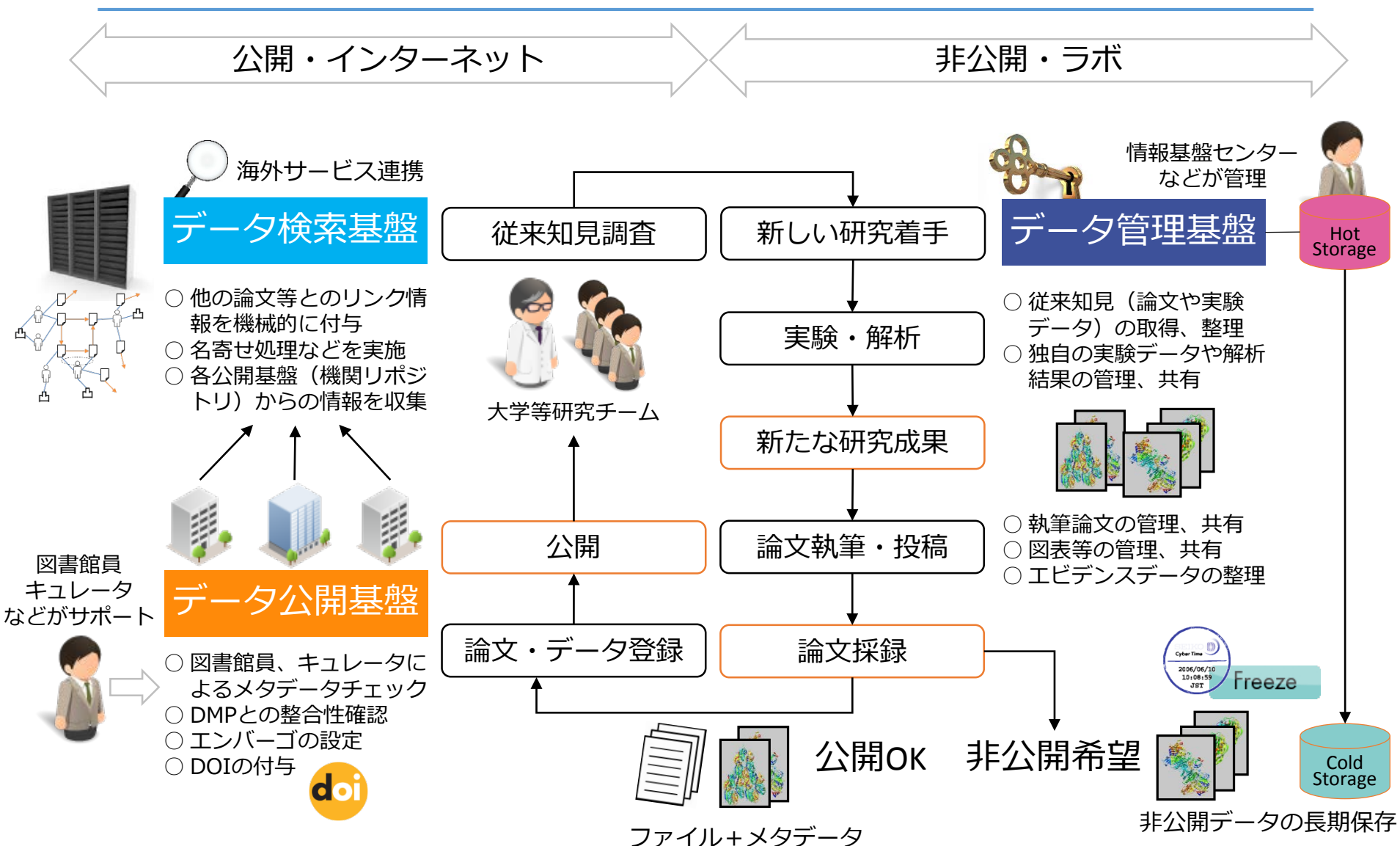


プライベートクラウド
(オンプレミス環境)

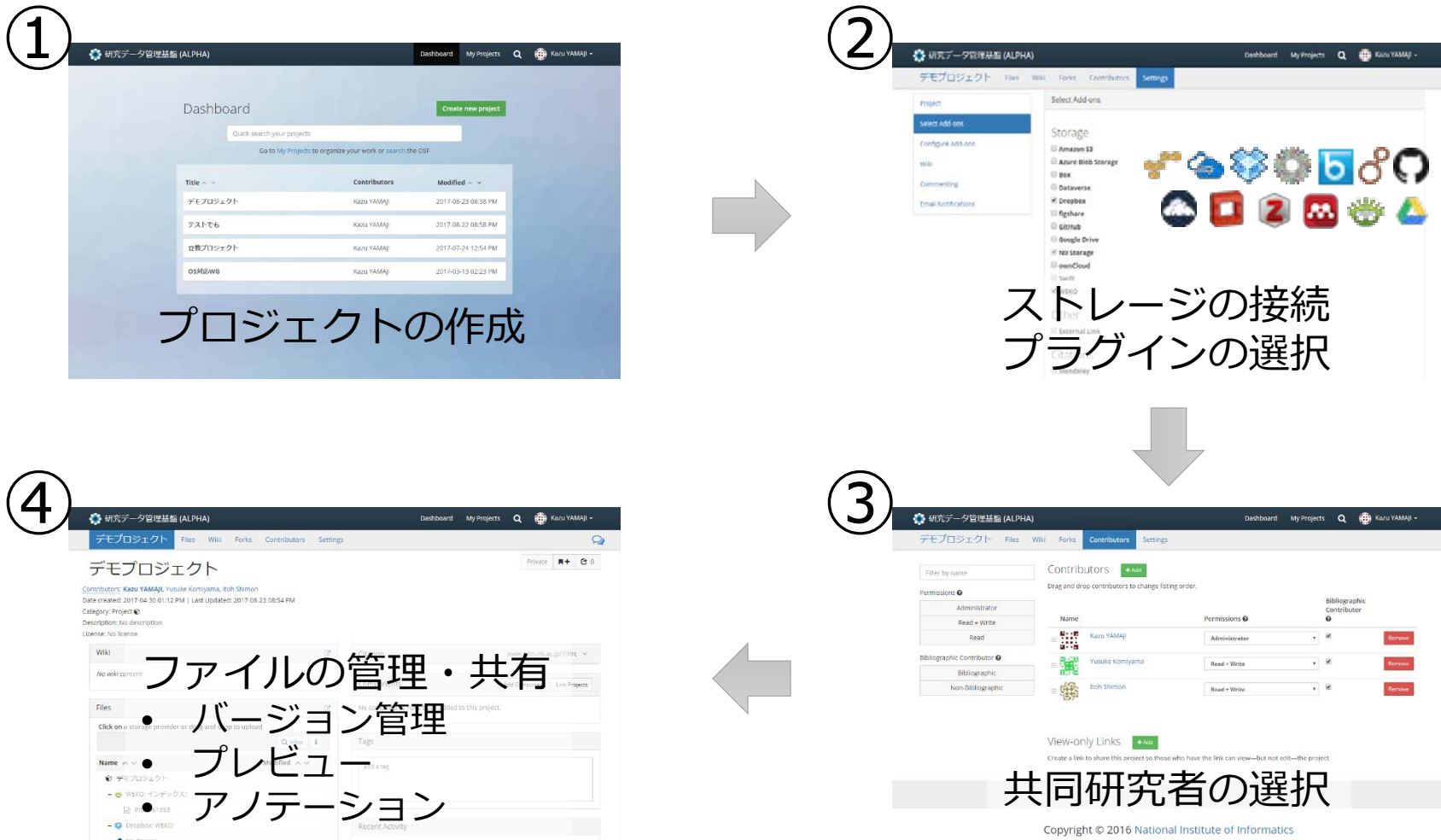


パブリッククラウド
(プロバイダーDC)

研究データ基盤による研究活動のサポート



GakuNin RDM 利用の流れ



国内外との研究開発連携

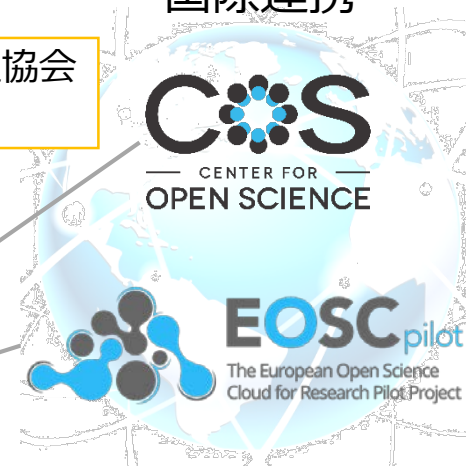
基盤センターとの連携
(機能要件、機関ポリシー)

大学ICT推進協議会 (AXIES)
研究データマネジメント部会

図書館との連携
(RDMトレーニング)

オープンアクセスリポジトリ推進協会
(JPCOAR) 研究データTF

国際連携



GakuNin RDM



オープンサイエンス基盤研究センター

学術ネットワーク
研究開発センター

- セキュアNW
- サービスデプロイ

クラウド基盤
研究開発センター

- 大学ストレージ調達
- データ解析基盤

学術認証推進室

- 認証連携基盤の普及
- VOサービスの普及

GakuNin RDMと連携可能な外部サービス

• クラウドストレージ

• パブリック

-  • Amazon S3
-  • S3互換ストレージ
-  • Azure Blob Storage
-  • Box
-  • Dropbox
-  • Google Drive
-  • One Drive

• プライベート

-  • ownCloud
-  • NextCloud
-  • OpenStack Swift
 - API ver2, ver3

• データ解析

青字：開発中

-  • JupyterHub
-  • [Galaxy](#)
- 図表・スライド共有
 -  • Figshare
- ソースコードレポジトリ
 -  • GitHub
 -  • BitBucket
 -  • GitLab
- 機関レポジトリ
 -  • WEKO2
 -  • Dataverse
 -  • SWORD互換
- 文献管理ツール
 -  • Mendeley
 - Zotero

研究者がよく利用するストレージ、データ解析、レポジトリ、文献管理ツールと連携

データ管理基盤とデータ公開基盤の役割



研究者

- データ管理機能
- 研究証跡の記録機能
- 公開基盤連携機能
- 非公開データ長期保存機能
- ...

図書館員・URA

- メタデータ管理機能
- データ公開機能
- 管理基盤連携機能
- DOI機能
- ...

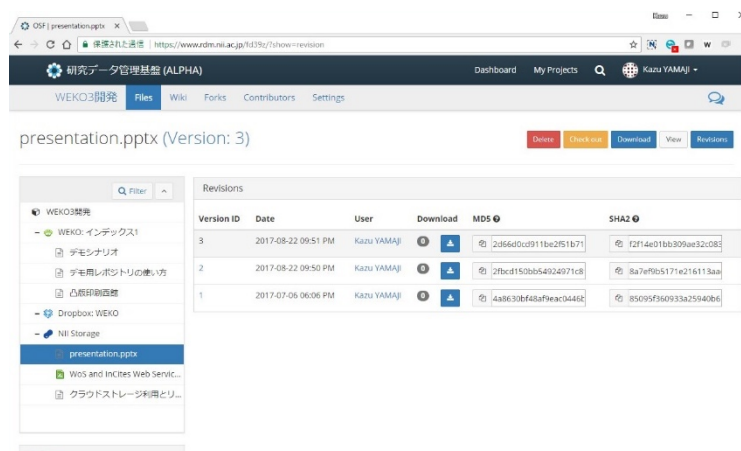
大学における業務と整合するように機能を整備

タイムスタンプリクエスト (データのハッシュ値)



時刻認証局

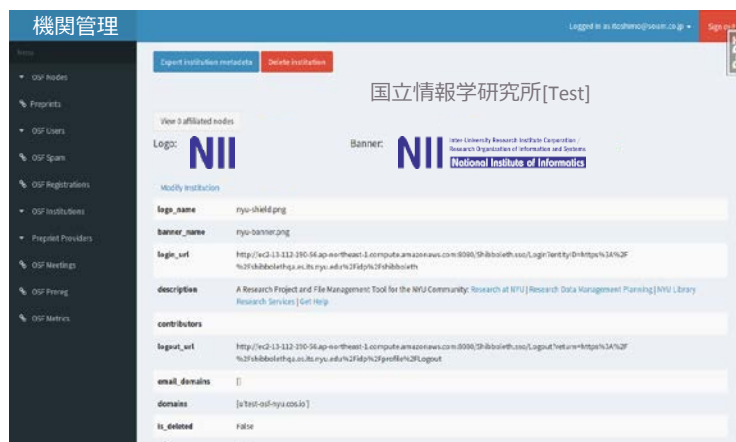
TSA (Time Stamp Authority)



TST (Time Stamp Token)
 認証局のデジタル署名

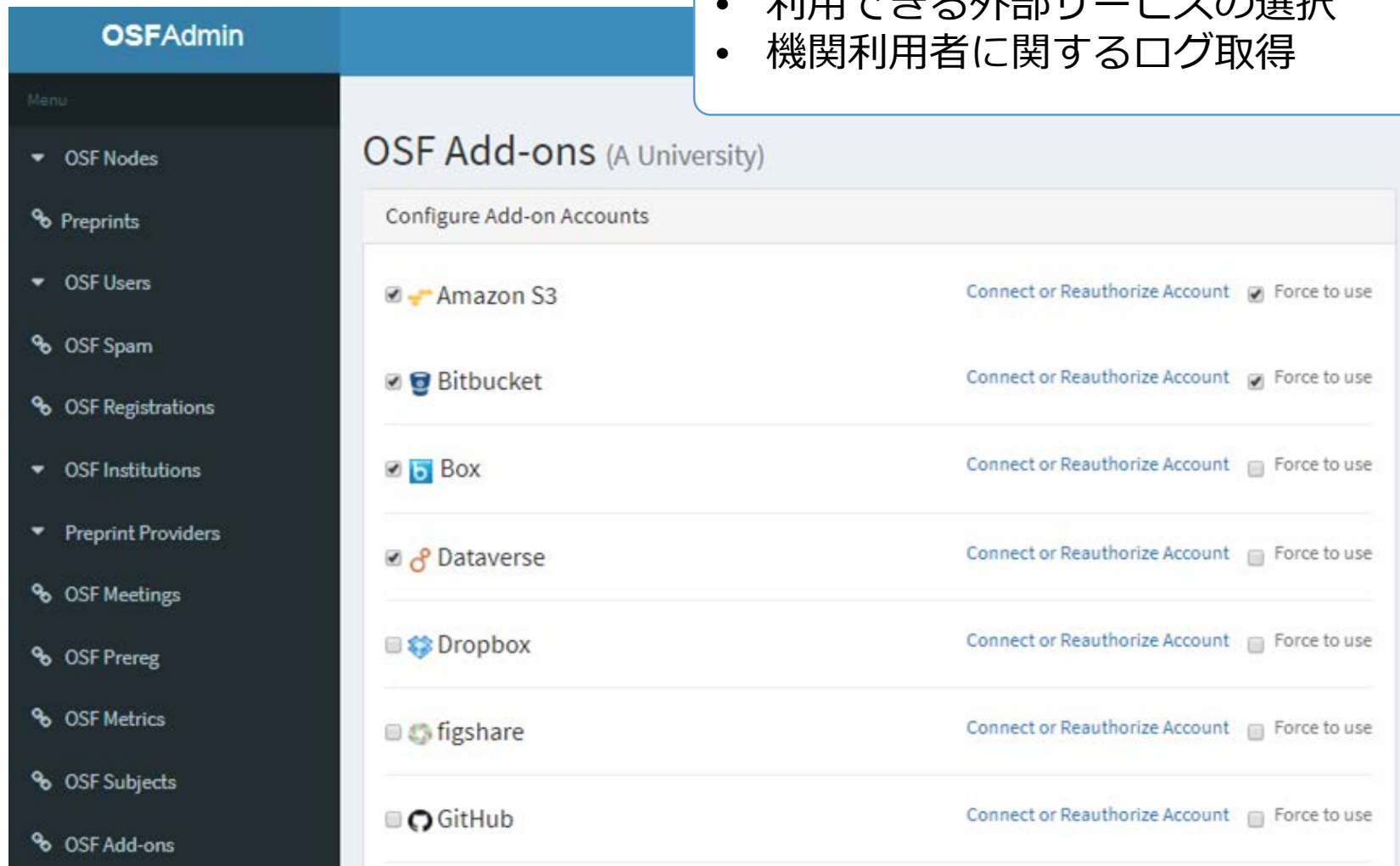
プロジェクトログ

機関ログ

[illegible]

機関向け管理機能

- 機関契約のクラウドストレージを接続
- 利用できる外部サービスの選択
- 機関利用者に関するログ取得



OSFAdmin

Menu

- ▼ OSF Nodes
- 🔗 Preprints
- ▼ OSF Users
- 🔗 OSF Spam
- 🔗 OSF Registrations
- ▼ OSF Institutions
- ▼ Preprint Providers
- 🔗 OSF Meetings
- 🔗 OSF Prereg
- 🔗 OSF Metrics
- 🔗 OSF Subjects
- 🔗 OSF Add-ons

OSF Add-ons (A University)

Configure Add-on Accounts

☒  Amazon S3	Connect or Reauthorize Account ☒ Force to use
☒  Bitbucket	Connect or Reauthorize Account ☒ Force to use
☒  Box	Connect or Reauthorize Account ☐ Force to use
☒  Dataverse	Connect or Reauthorize Account ☐ Force to use
☐  Dropbox	Connect or Reauthorize Account ☐ Force to use
☐  figshare	Connect or Reauthorize Account ☐ Force to use
☐  GitHub	Connect or Reauthorize Account ☐ Force to use

OSFAdmin

Menu

▼ OSF Nodes

🔗 Preprints

▼ OSF Users

🔗 OSF Spam

🔗 OSF Registrations

▼ OSF Institutions

▼ Preprint Providers

🔗 OSF Meetings

🔗 OSF Prereg

🔗 OSF Metrics

🔗 OSF Subjects

🔗 OSF Add-ons

🔗 OSF Storage Metrics

OSF Storage Metrics

ABC University

• [Amazon S3](#)
• [Box](#)
• [Dropbox](#)

+ Amazon S3

1. file numbers and sizes

date	8/27	9/3	9/10	9/17	9/24	10/1	10/8	10/15	10/22	10/29
number of files	132	142	145	167	209	212	234	234	256	278
sum of file sizes	108	121	123	154	178	187	209	209	223	254



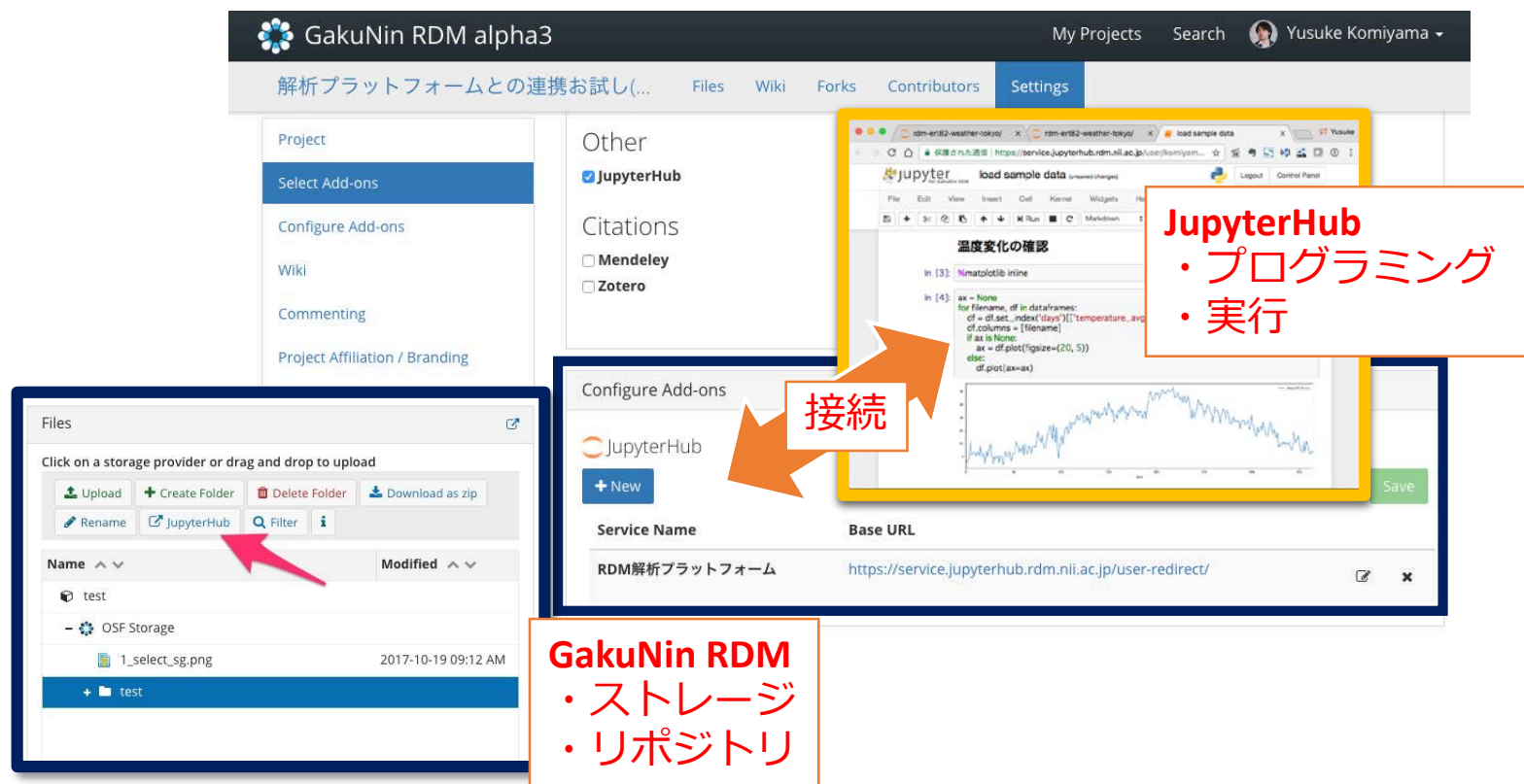
Download PDF

PDFファイル出力

ストレージの統計情報の表示欄

- ・登録ファイル数（全登録数）
- ・ファイルの合計サイズ
- ・ファイル形式毎の登録件数

データ解析ソフトウェア連携



JupyterHub

- ・プログラミング
- ・実行

接続

GakuNin RDM

- ・ストレージ
- ・リポジトリ

- データ解析プラットフォームJupyterHub用のGakuNin RDMアドオンを開発
- 連携によりシステム間でのデータ入出力が容易に
- 双方のアプリケーションでの認証は学認に対応

第1回NII 研究データ管理基盤機能評価試験

• 目的

- 国立情報学研究所（NII）で開発した研究データ管理基盤（以下、管理基盤という）サービスの機能評価試験（クローズドアルファテスト）を行った。

• 期間

- 平成29年2月23日（木）～3月17日（金）

• 方法

- 商用パブリッククラウド上に設けた試験用管理基盤に、Web ブラウザから管理基盤サービスにログインし、利用マニュアルに基づく操作を行い、システムの使用感や改善点について報告する。

• 試験者

- 大学の情報環境のあり方検討委員会
- オープンサイエンス対応 WG
- サブワーキングメンバー

第1回機能評価試験の試験者の代表

参加機関名	部局名
北海道大学	情報基盤センター
東北大学	サイバーサイエンスセンター
名古屋工業大学	情報工学専攻
京都大学	情報環境機構
大阪大学	サイバーメディアセンター
九州大学	附属図書館
九州大学	情報基盤研究開発センター
国立環境研究所	地球環境研究センター

情報基盤センター 6、図書館 1、国研 1
7機関8組織

第2回NII 研究データ管理基盤機能評価試験

- **第2回機能評価試験の目的**

- 第2回の機能評価試験（クローズドアルファテスト2）では、多様な学術分野の研究所、研究センターまたは大学研究室を対象に、GakuNin RDMに対する要望や研究ソフトウェアとの連携の可能性、共同研究での利用ユースケースにおける有意性や懸念事項について検証した。

- **期間**

- 平成29 年10月12日（木）～ 11月12日（日）

- **実験方法**

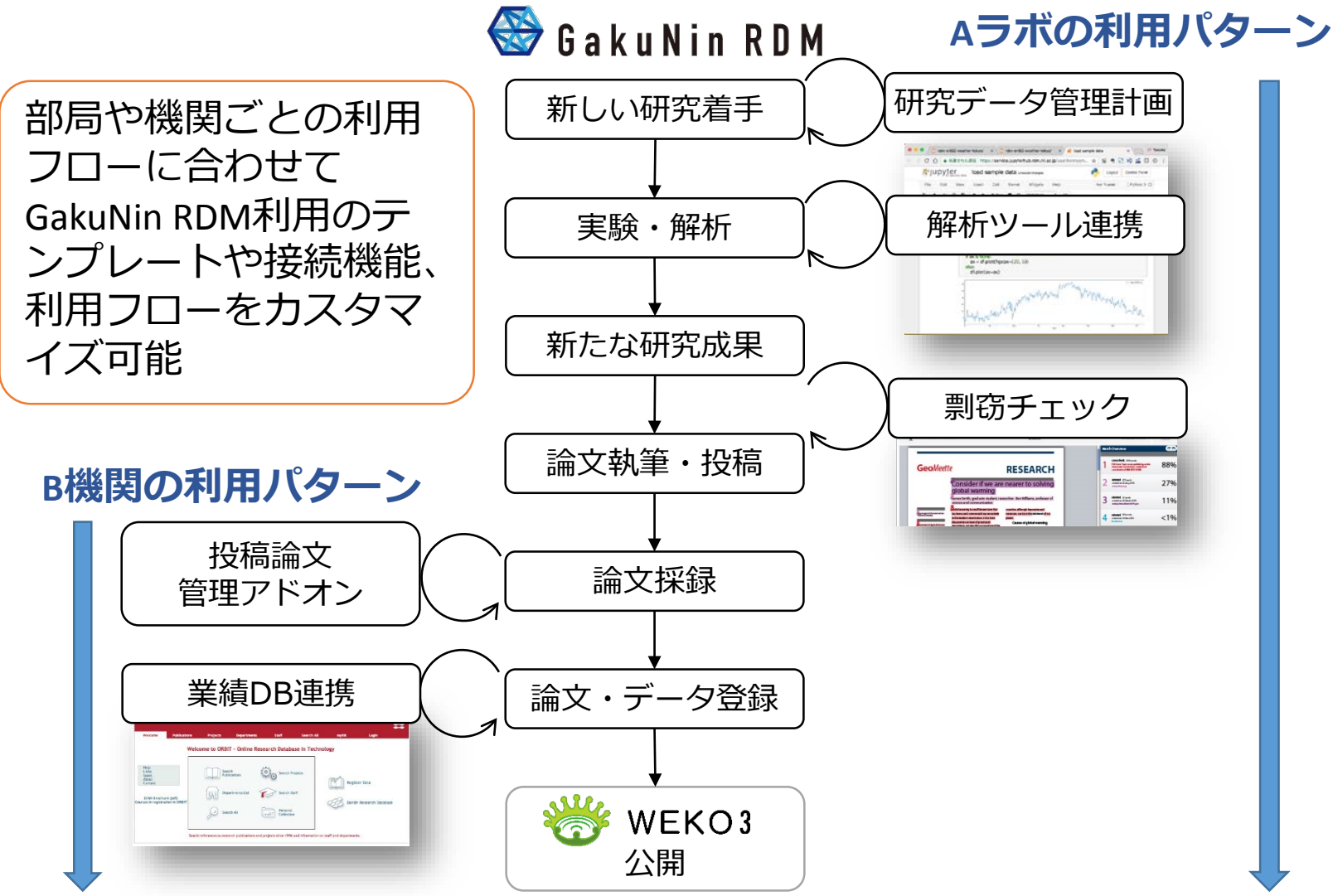
- 商用パブリッククラウドAmazon Web Services (AWS)上で構築した評価試験用のGakuNin RDMに対し、試験者をWeb ブラウザから本サービスにログインさせ、利用マニュアルに基づく一連の操作および研究室での自由な利用を行い、システムの使用感や改善点についてアンケートに回答し報告する。

第2回機能評価試験の試験者の代表

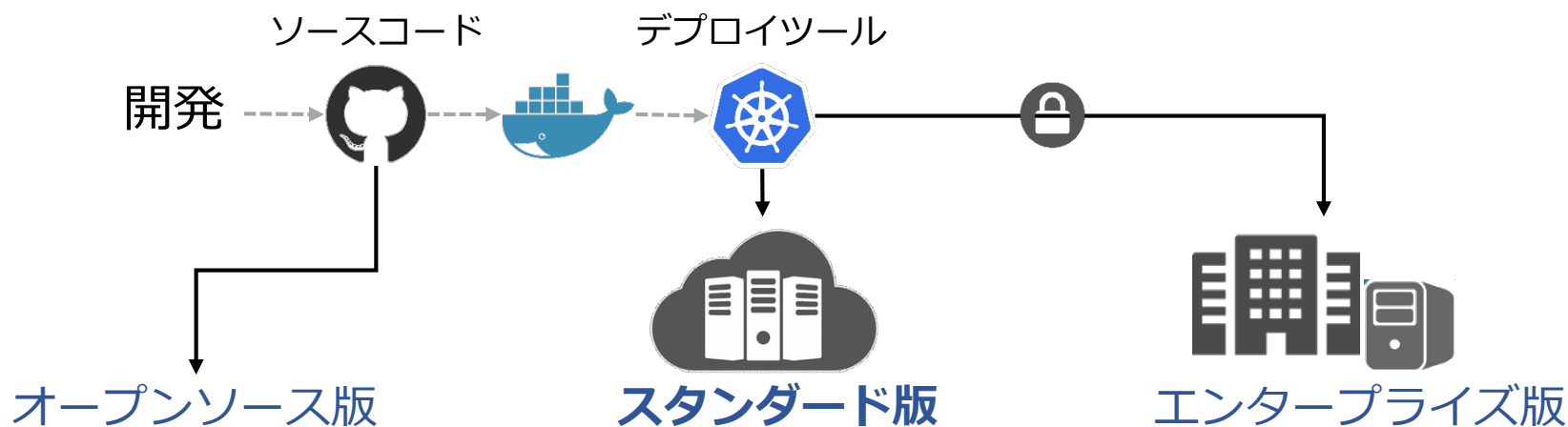
所属	分野
東京大学 医科学研究所	タンパク質科学
東京大学 先端科学技術研究センター	ゲノム科学
会津大学 先端情報科学研究センター	宇宙情報科学
福島県立医科大学 附属病院 臨床研究センター	臨床薬理学
慶応義塾大学 文学部	図書館情報学
東京大学 附属図書館 アジア研究図書館	宗教学
物質・材料研究機構 統合型材料開発・情報基盤部門	材料科学
理化学研究所 情報基盤センター	計算機科学
東京大学 情報基盤センター	*情報基盤センター
名古屋大学 情報基盤センター	*情報基盤センター

8分野10組織、ユニークユーザ数40名

機関の利用モデルに沿った展開



GakuNin RDMの提供方法



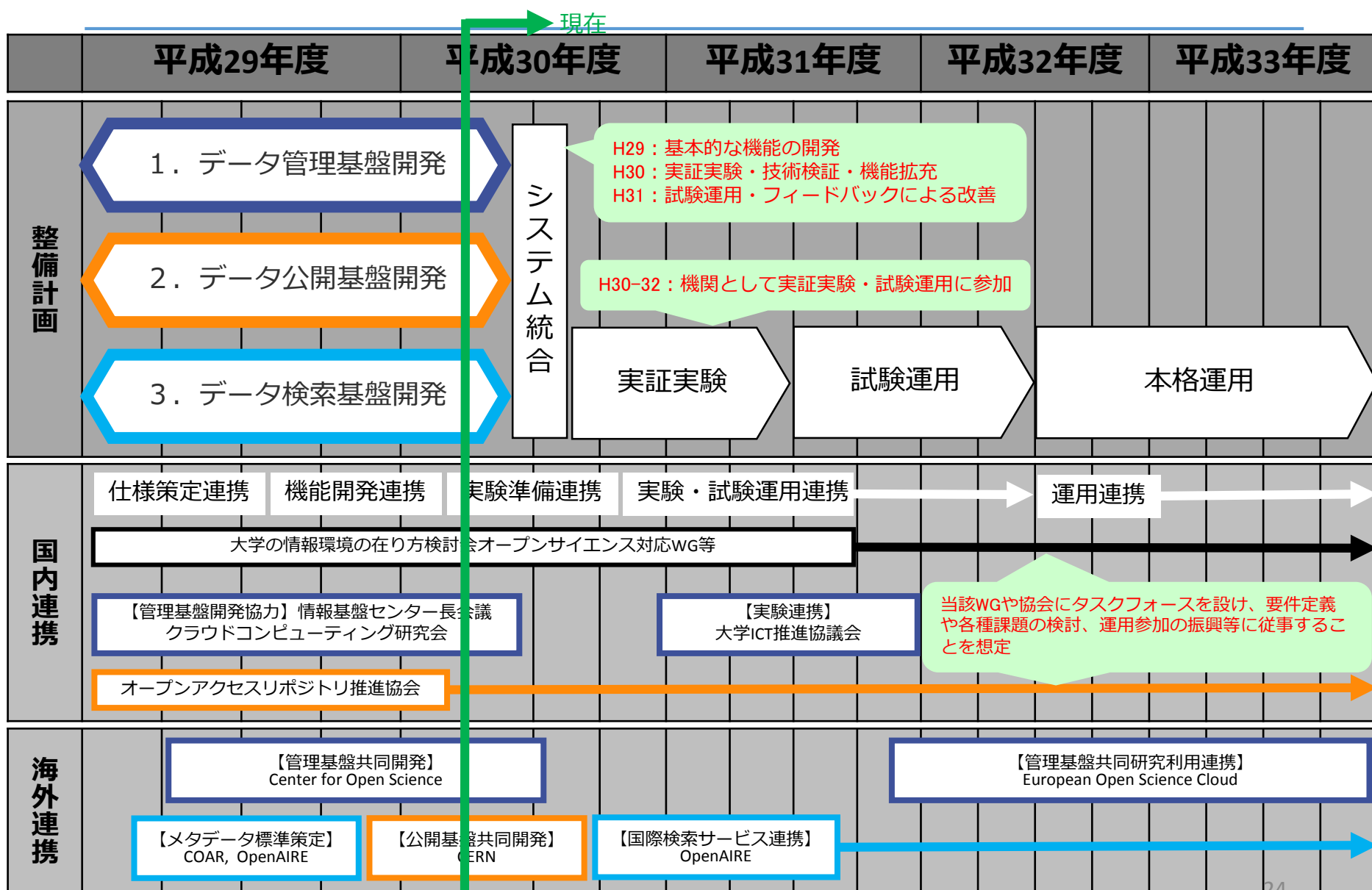
- 機関が独自にGakuNin RDMを運用したい場合
- ソースコードとドキュメントを提供
- 運用のノウハウを共有

- パブリッククラウド上のSaaSとしてNIIが提供
- 機関が用意するストレージを接続して利用可能
- 学認による認証

- 機関のポリシー上スタンダード版ができない場合
- デプロイコードの提供、あるいはNIIが直接デプロイ
- 機関のIdPによる認証

機関のポリシー等の制約などにも対応できるようにいくつかの運用オプションを検討中

研究データ基盤の整備計画



まとめ

- オープンサイエンス基盤（2020年度以降に本稼働予定）
 - 研究データ検索基盤（CiNii Research）
 - 研究データ公開基盤（WEKO3）
 - 研究データ管理基盤（GakuNin RDM）
- 管理基盤のクローズドテスト（第三回） **6月29日まで実施中**
大学の情報基盤センターに向けて、本年度開発予定の GakuNin RDM の研究管理機能や外部ツール連携などを中心にご利用いただき、機関導入の観点からレビュー。意見を集約し次期四半期の開発にフィードバックします。

2018年度第2四半期からも中長期的なオープンベータテストも予定しています、研究データ管理基盤 GakuNin RDM の検証実験に参加いただけますよう、よろしくお願いいたします。