NII学術情報基盤オープンフォーラム

オープンサイエンストラック

研究データ管理基盤 GakuNin RDM

2018年6月20日(水)10:00-12:00
特別会議室（1F）

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れなければならない。
  - **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

http://id.nii.ac.jp/1280/00000019/　より抜粋

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

	Policy Coverage		Policy Stipulations					Support Provided			
Research Funders	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.

研究データ
ストレージ

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
公開基盤

Sharing & publishing data

Edinburgh DataShare

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
管理トレーニング

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

Science2.0

Web2.0

ラボ

成果の再利用で研究の加速化
- 共同研究への積極的な展開
- 研究成果の再現性が向上

Research Data

研究者の適切な評価
- 若手研究者のVisibility向上
- 実験者や解析者にも脚光

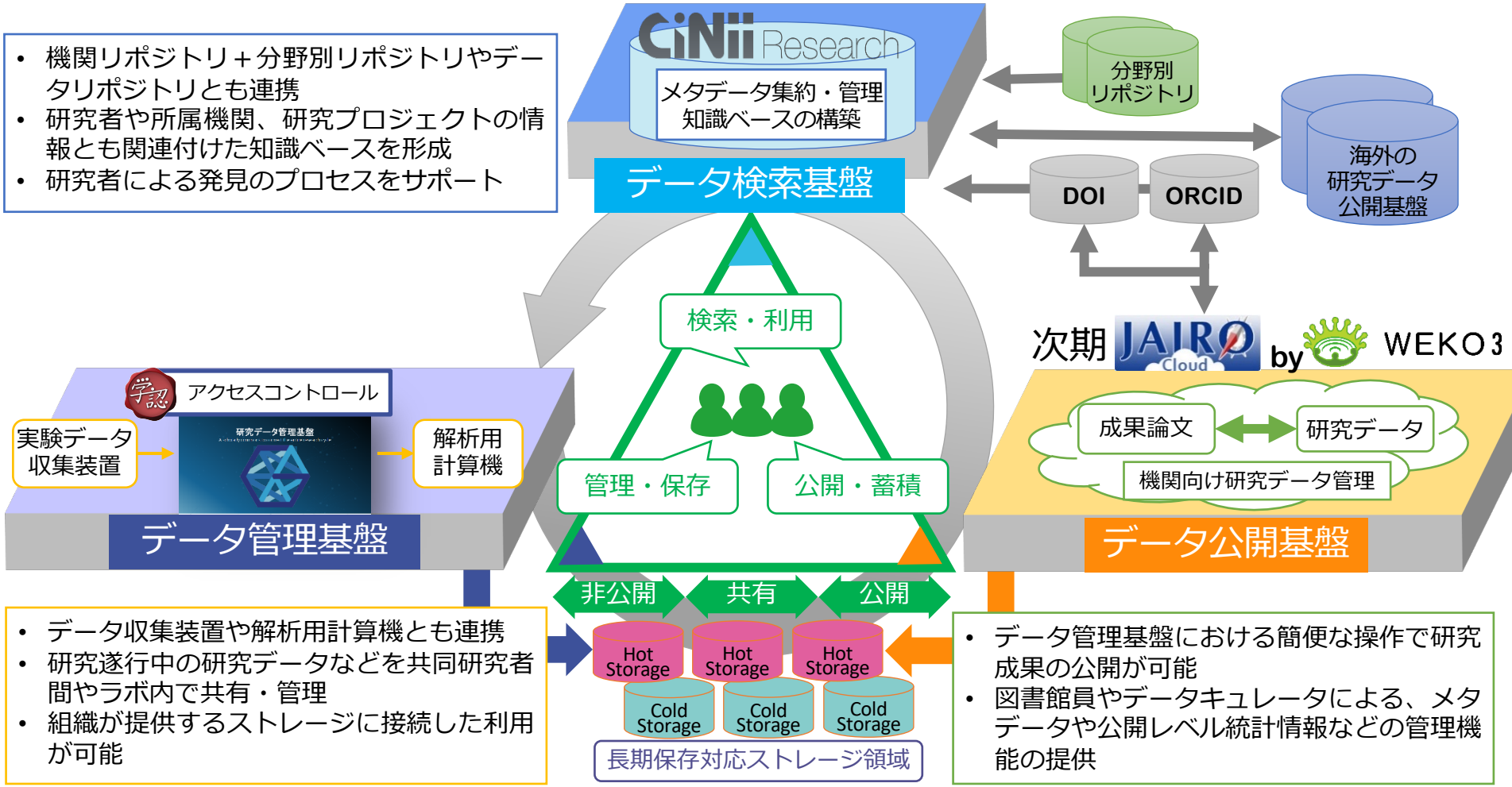
コミュニティ

市民科学
産学連携

研究者間の
格差是正

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

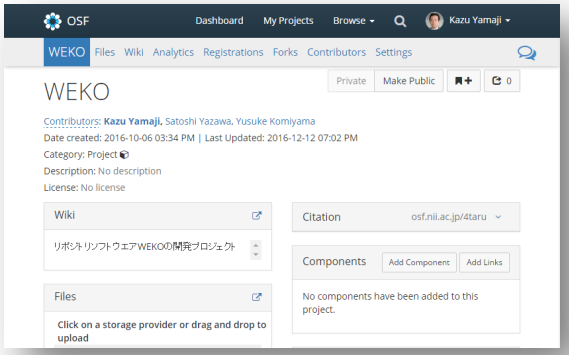


- データ収集装置や解析用計算機とも連携
- 研究遂行中の研究データなどを共同研究者間やラボ内で共有・管理
- 組織が提供するストレージに接続した利用が可能

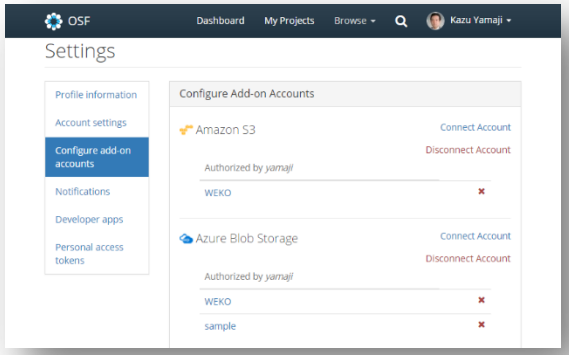
- データ管理基盤における簡便な操作で研究成果の公開が可能
- 図書館員やデータキュレータによる、メタデータや公開レベル統計情報などの管理機能の提供

データ管理基盤 GakuNin RDM

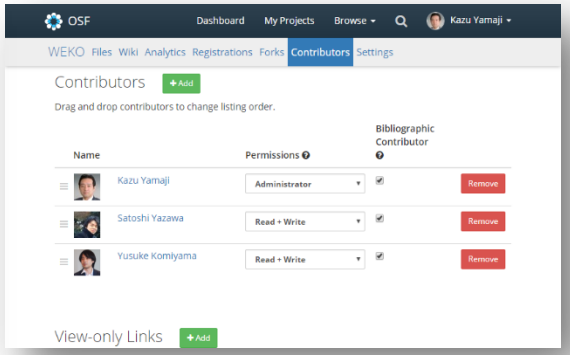
研究プロジェクト単位で
ファイルなどを管理



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

研究データ管理基盤

機関毎に準備

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）

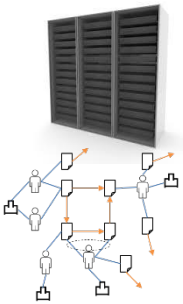
プライベートクラウド
（オンプレミス環境）

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

公開・インターネット

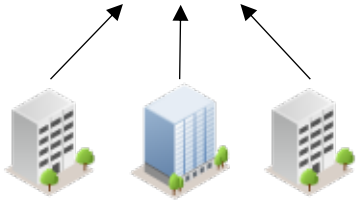
非公開・ラボ



海外サービス連携

データ検索基盤

- ○ 他の論文等とのリンク情報を機械的に付与
- ○ 名寄せ処理などを実施
- ○ 各公開基盤（機関リポジトリ）からの情報を収集



図書館員
キュレータ
などがサポート

データ公開基盤

- ○ 図書館員、キュレータによるメタデータチェック
- ○ DMPとの整合性確認
- ○ エンバーゴの設定
- ○ DOIの付与

従来知見調査

大学等研究チーム

公開

論文・データ登録

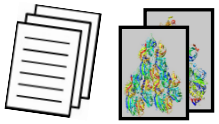
新しい研究着手

実験・解析

新たな研究成果

論文執筆・投稿

論文採録



ファイル＋メタデータ

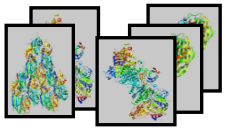
公開OK

非公開希望

情報基盤センター
などが管理

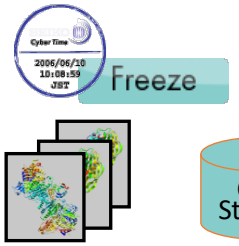
データ管理基盤

- ○ 従来知見（論文や実験データ）の取得、整理
- ○ 独自の実験データや解析結果の管理、共有



- ○ 執筆論文の管理、共有
- ○ 図表等の管理、共有
- ○ エビデンスデータの整理

Hot Storage



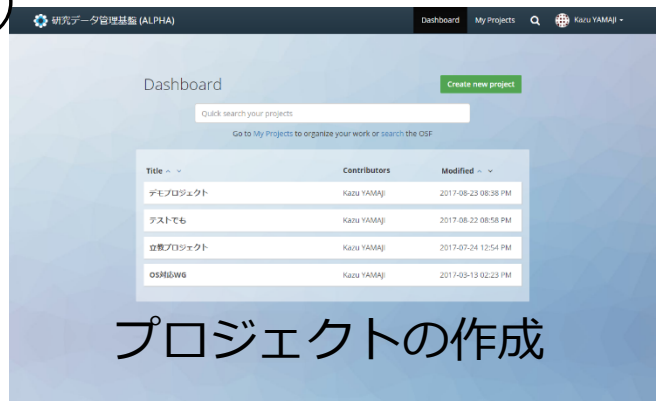
Freeze

Cold Storage

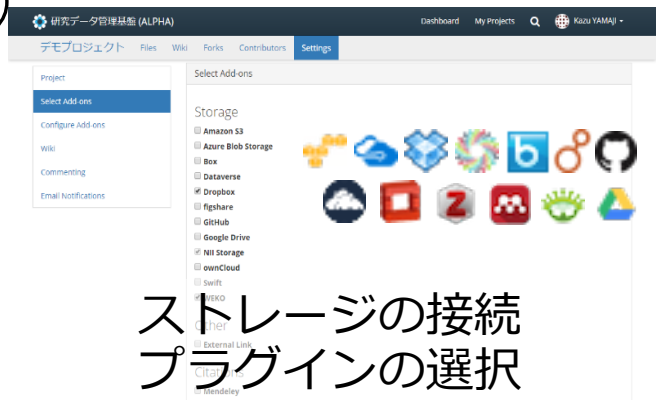
非公開データの長期保存

GakuNin RDM 利用の流れ

①

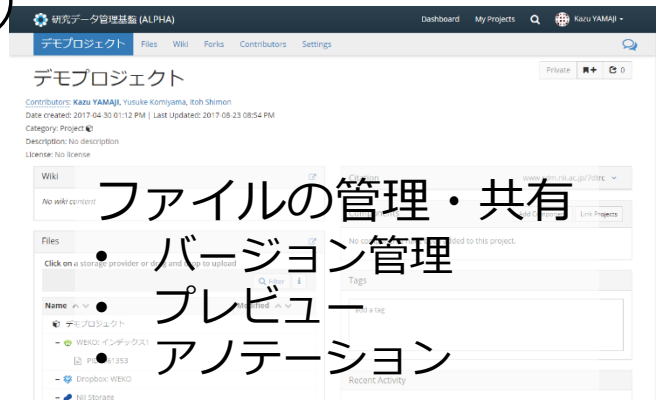


②



③

④



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)
研究データマネージメント部会

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

COS CENTER FOR OPEN SCIENCE

EOSCpilot The European Open Science Cloud for Research Pilot Project

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

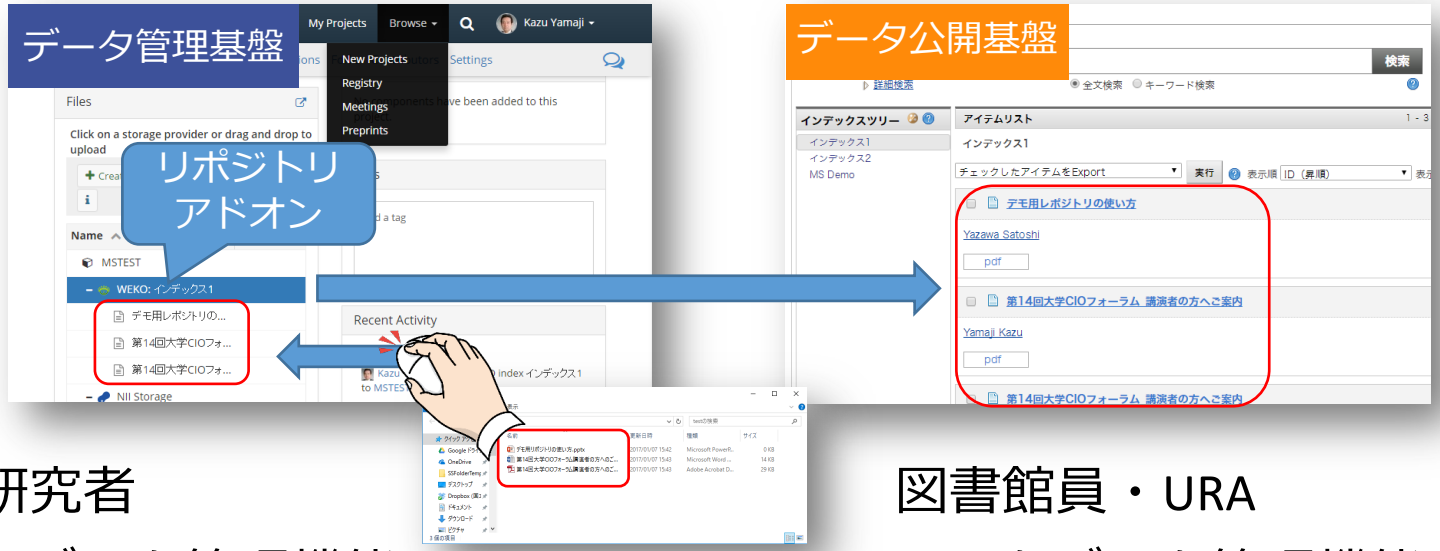
青字：開発中

- 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
  - 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

GakuNin RDM　WEKO3

データ管理基盤とデータ公開基盤の役割



研究者

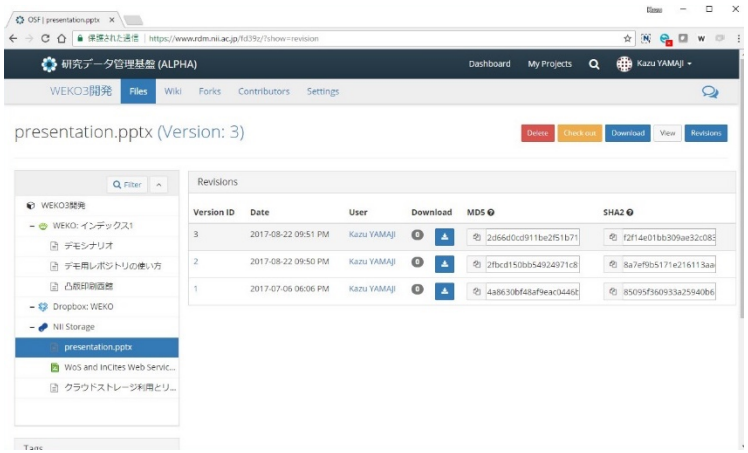
- データ管理機能
- 研究証跡の記録機能
- 公開基盤連携機能
- 非公開データ長期保存機能
- …

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

研究証跡保存機能

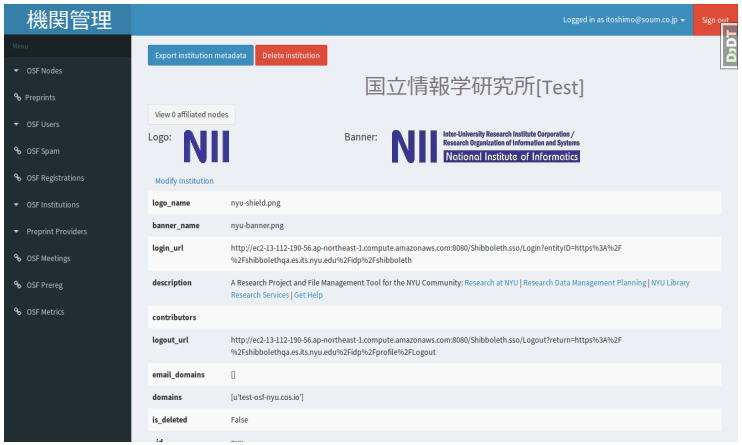


タイムスタンプリクエスト
（データのハッシュ値）

TST (Time Stamp Token)
認証局のデジタル署名

時刻認証局
TSA (Time Stamp Authority)

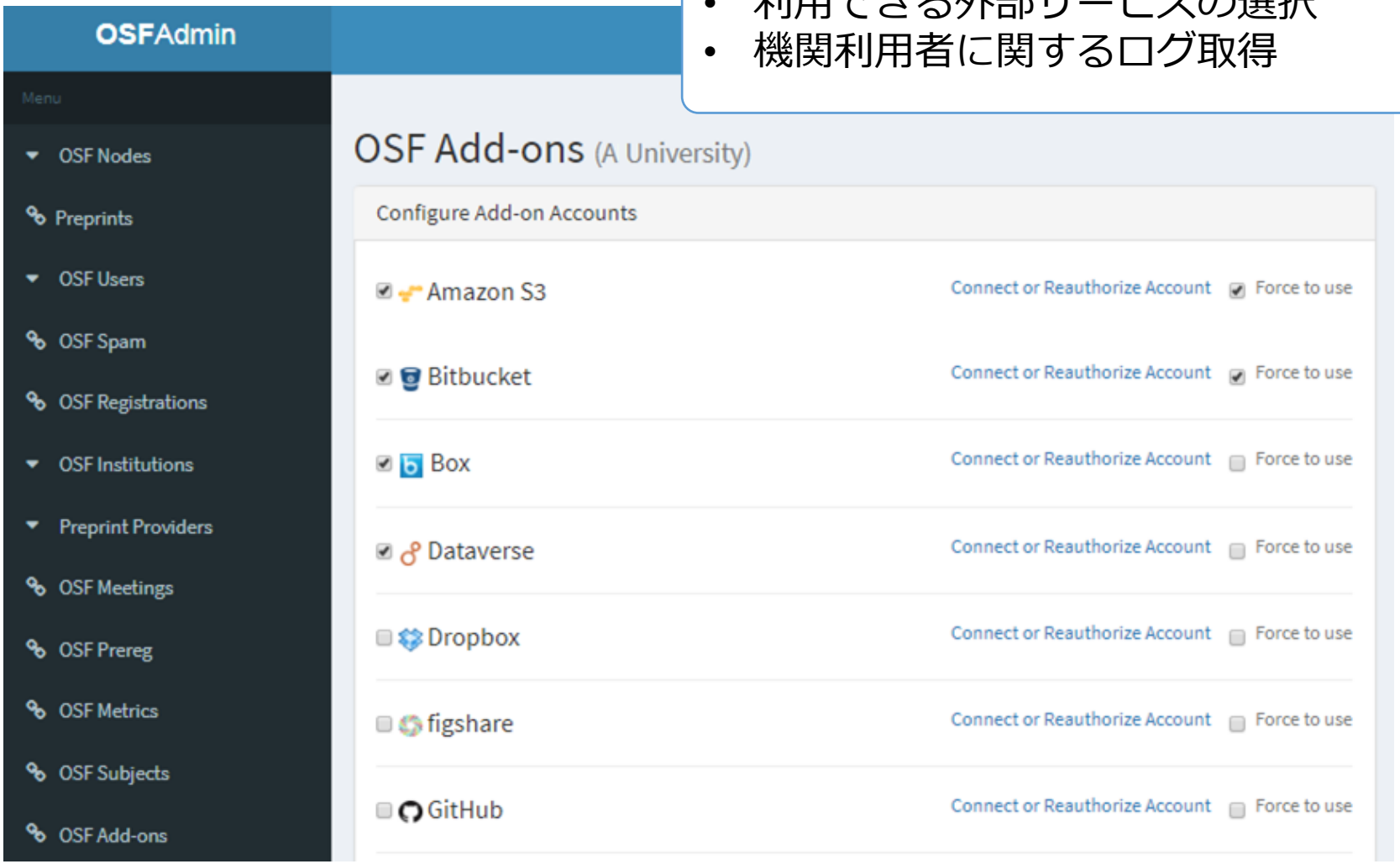
プロジェクトログ



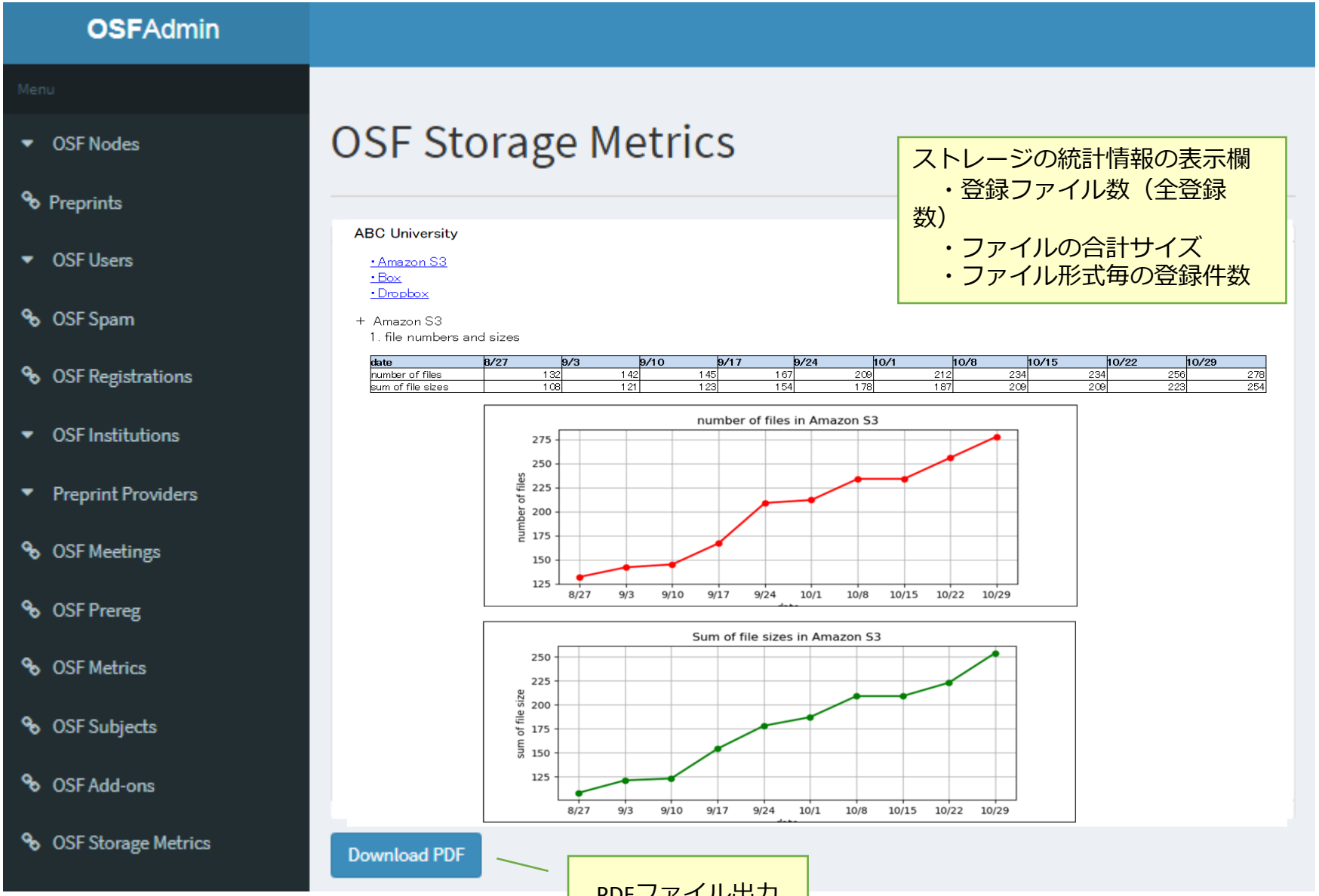
機関ログ

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



統計情報

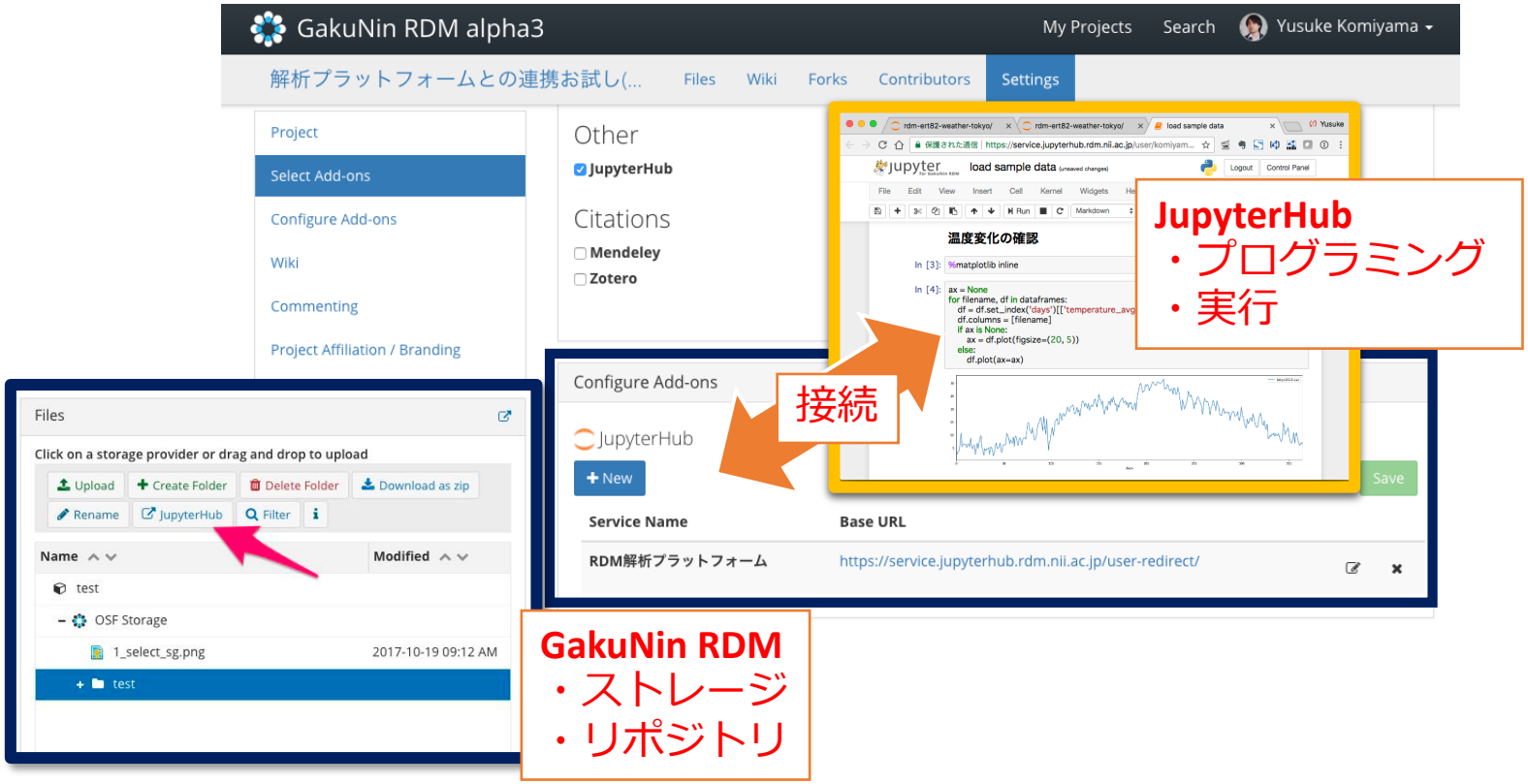


date	8/27	9/3	9/10	9/17	9/24	10/1	10/8	10/15	10/22	10/29
number of files	132	142	145	167	209	212	234	234	256	278
sum of file sizes	108	121	123	154	178	187	209	209	223	254

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

PDFファイル出力

データ解析ソフトウェア連携



- 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

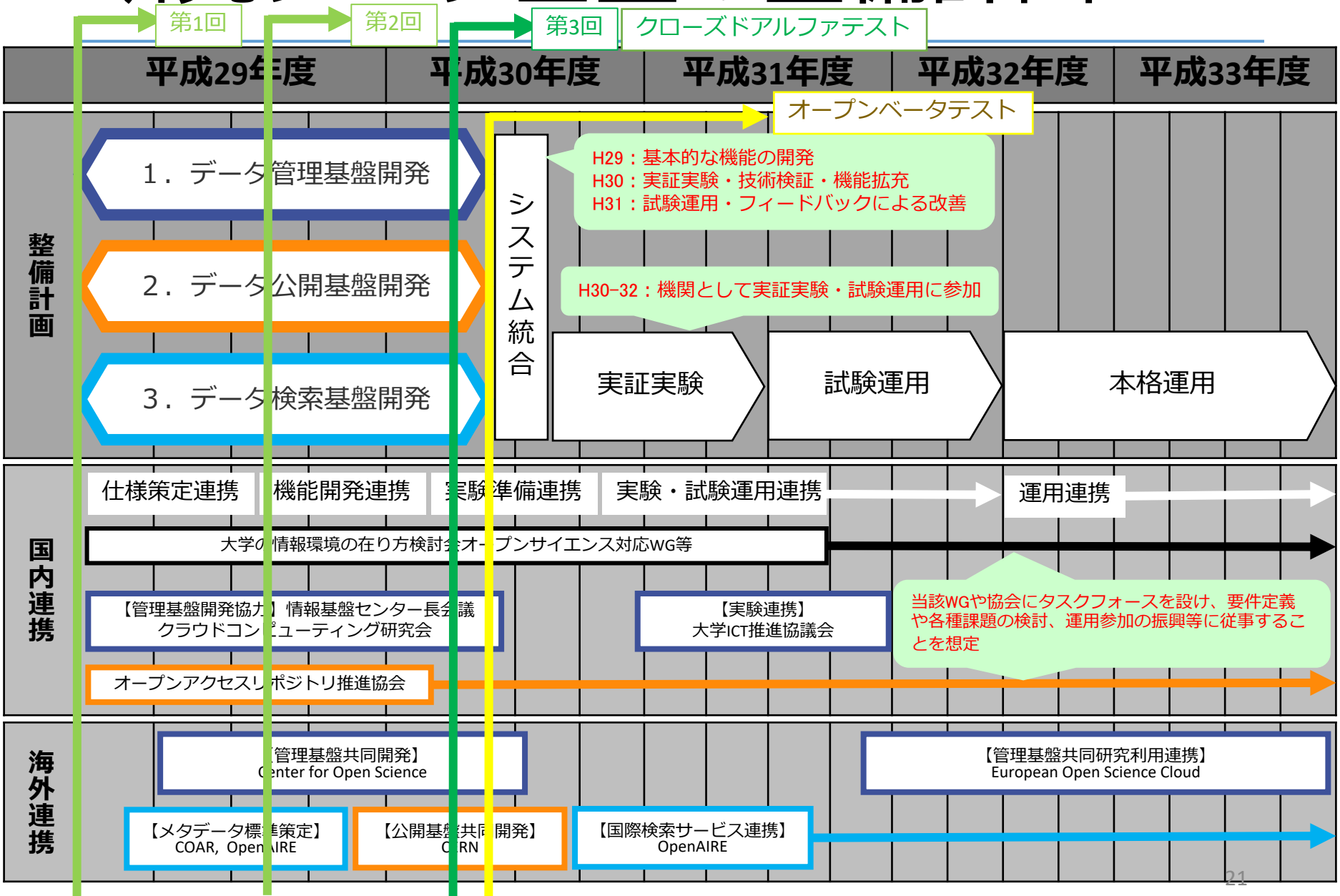
第2回機能評価試験

- **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3回機能評価試験

- **第3回機能評価試験の目的**
  - 第3回の機能評価試験（クローズドアルファテスト3）では、中規模の大学、多様な分野の研究機関のシステム導入担当者、開発者を対象に、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30年6月4日（木）～ 6月29日（金）　開催中
- **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関連部局・部署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。

研究データ基盤の整備計画



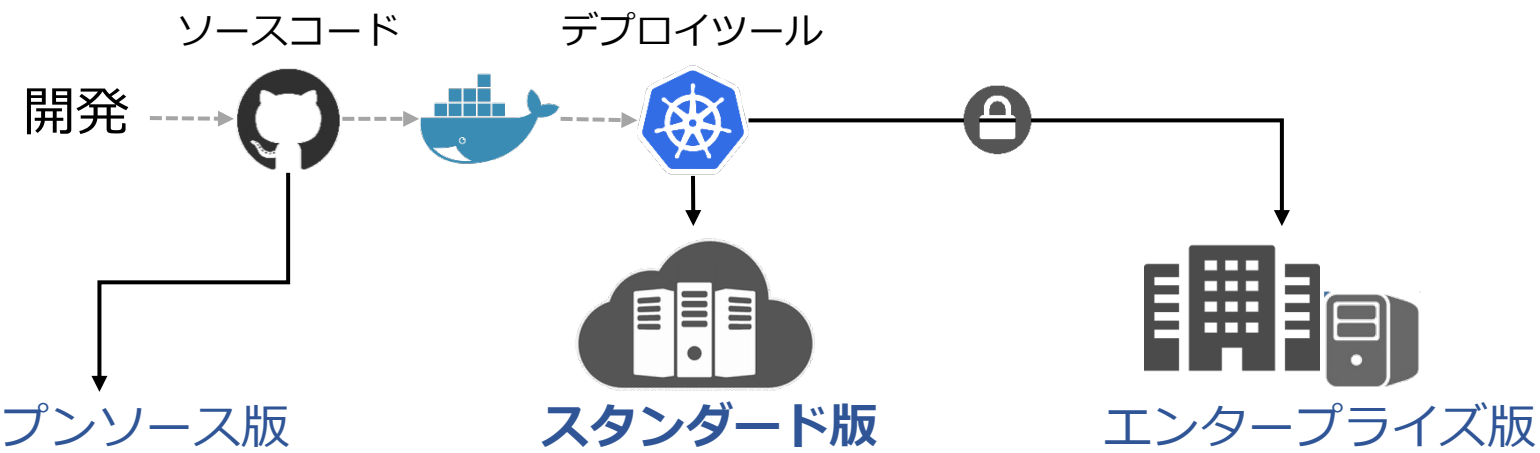
GakuNin RDM試験参加の方法

- 第3回機能評価試験（2018年6月29日終了）
  - GakuNin RDMユーザサポートページ
    - https://goo.gl/7kUKQd
  - 「クローズドアルファテスト３に参加希望」と書いて
    - GakuNin RDM事務局
    - rdm_support@nii.ac.jp

GakuNin RDM試験参加の方法

- 公開実証試験の先行申込
  - 今夏（第２四半期）ご案内予定
  - １年以上のロングランのテスト
  - 「オープンベータテストに参加希望」と書いて
    - GakuNin RDM事務局
    - rdm_support@nii.ac.jp

GakuNin RDMのサービス提供方法



オープンソース版	スタンダード版	エンタープライズ版
• 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

機関の利用モデルに沿った展開

部局や機関ごとの利用フローに合わせてGakuNin RDM利用のテンプレートや接続機能をカスタマイズ可能

GakuNin RDM

新しい研究着手

実験・解析

新たな研究成果

論文執筆・投稿

論文採録

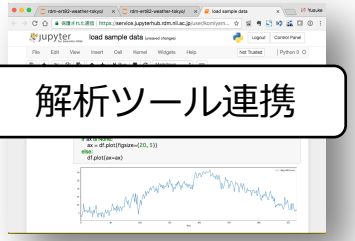
論文・データ登録

WEKO3
公開

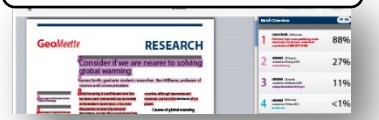
Aラボの利用パターン

研究データ管理計画

解析ツール連携



剽窃チェック



B機関の利用パターン

投稿論文
管理アドオン

業績DB連携

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。