

次世代リポジトリシステムと 研究データ管理における 図書館の役割

山地一禎¹、林正治²、林洋平¹、南山泰之³、尾城孝一

1 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

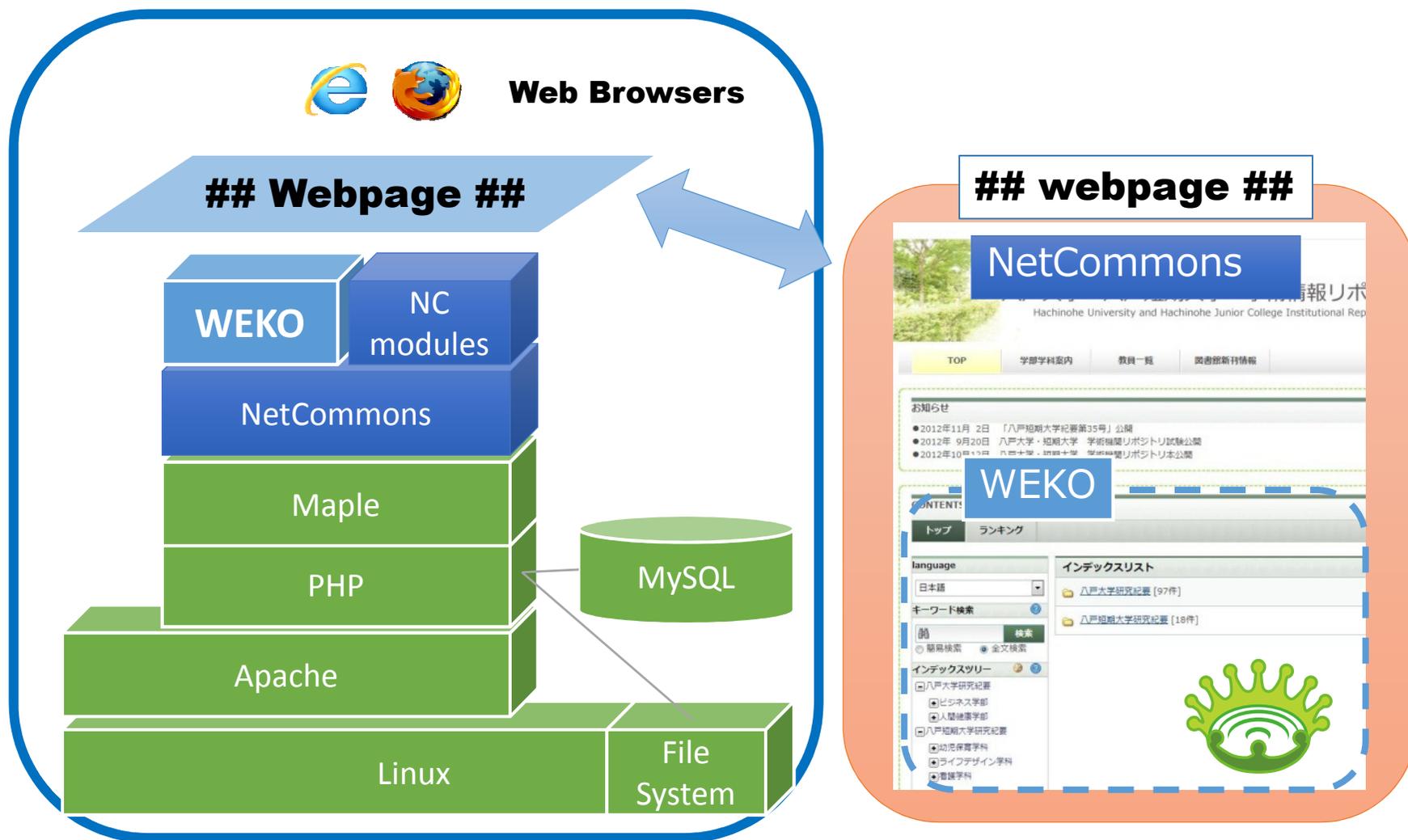
2 国立情報学研究所 サイバーセキュリティ研究開発センター

3 国立極地研究所 情報図書室

話の流れ

- 現状のリポジトリシステム
 - NIIで構築してきたWEKOとJairo Cloudについて
- 解決したい課題
- 次世代リポジトリシステム
- 次世代システムと図書館の役割

WEKO2のシステムアーキテクチャ



NetCommonsの特徴

- Public Space (外部向けポータルサイト)
 - Webプログラミングやデザインを知らなくても、簡単に美しいポータル・サイトが構築できる
 - メールマガジンの配信や管理
 - 多様なコンテンツの配信や管理
- Group Space (グループ向けコミュニティサイト)
 - アクセスコントロールを備えた本格的なグループウェア
 - E-ラーニング用モジュールを標準装備
 - 必要な機能を取捨選択し、目的にあった柔軟な情報共有の場を構築できる
- Private Space (個人向けバーチャルオフィス)
 - Web上のファイル格納庫
 - 日誌・備忘録
 - スケジュール管理

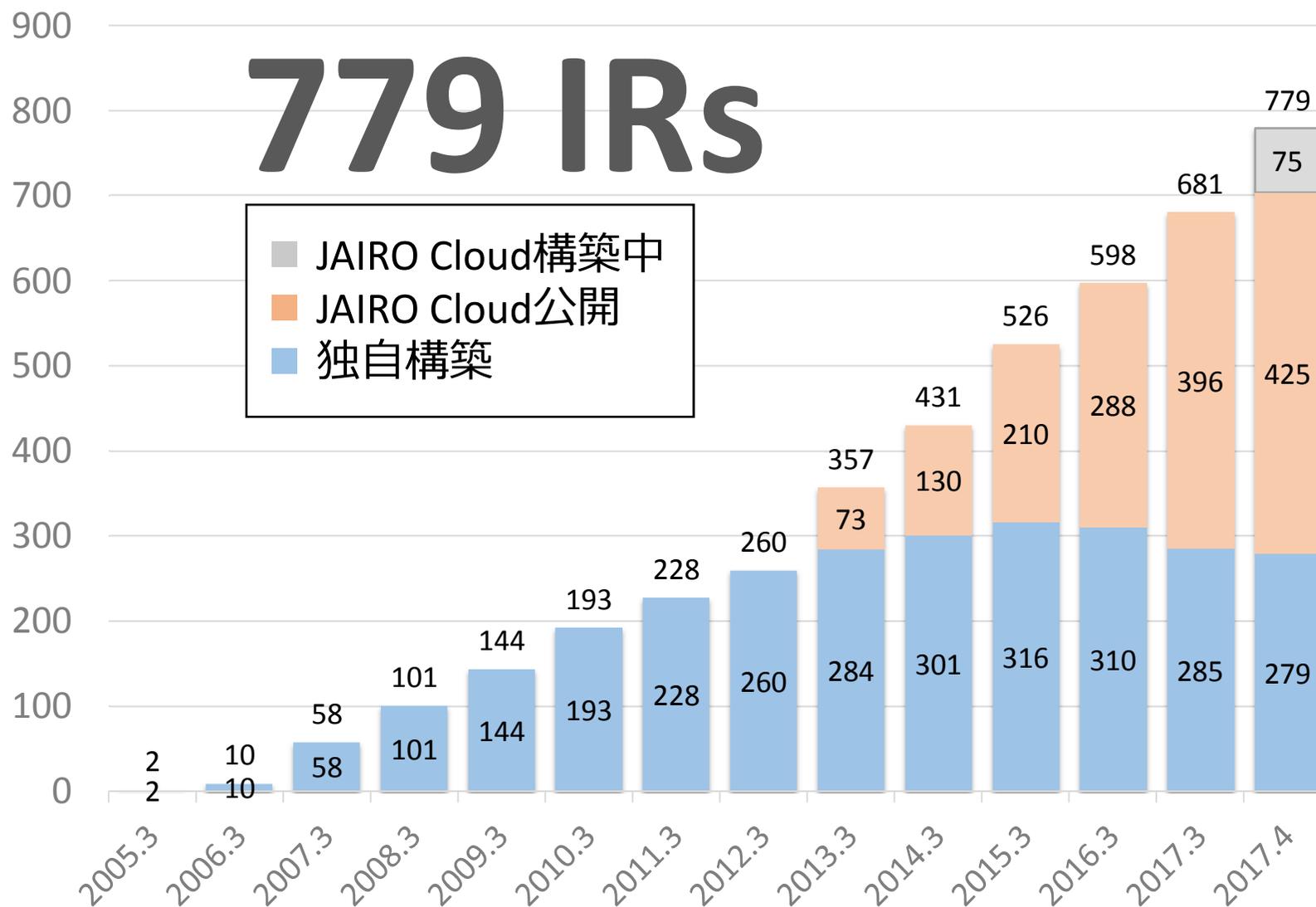
CMS（NC2）ベースで構築した意図

- Ajaxによる直感的で軽快な操作感の提供
- UIからの管理機能の提供（NC2の思想）
- インストールやアップデートの簡便性
- 多様なモジュール（機能）と組み合わせた個性のある機関リポジトリの構築
- Group, Privateスペースを活用した研究利用

CMS（NC2）ベースで構築した意図

- Ajaxによる直感的で軽快な操作感の提供
 - WEKOの強み
 - UIからの管理機能の提供（NC2の思想）
 - WEKOの強み
 - インストールやアップデートの簡便性
 - Jairo Cloudによる位置づけの変化
 - 多様なモジュール（機能）と組み合わせた個性のある機関リポジトリの構築
 - それほど使われているわけではない
 - Group, Privateスペースを活用した研究利用
 - それほど使われているわけではない
- JCの拡大

日本における機関リポジトリ数



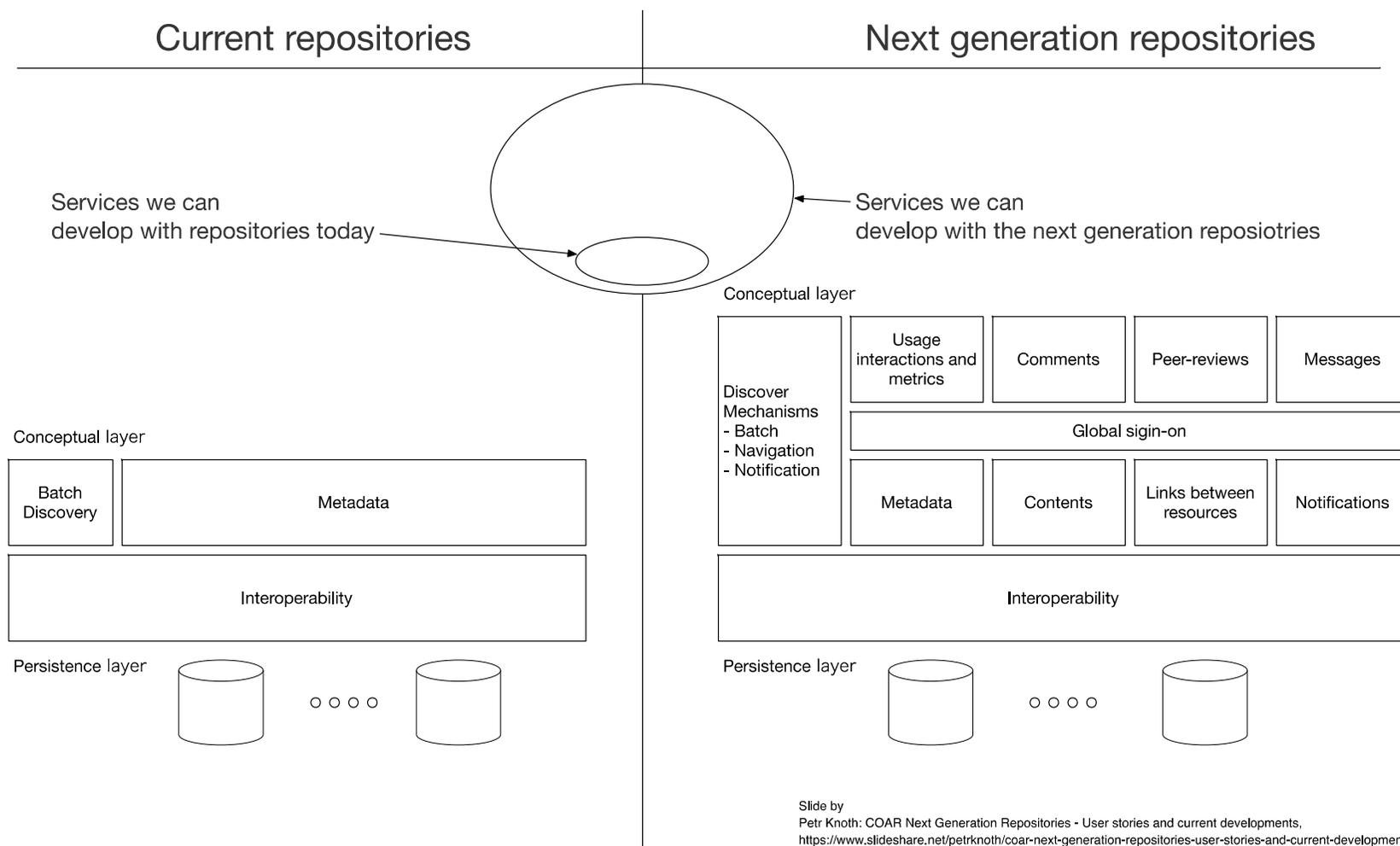
機関リポジトリとは

- Clifford Lynch (2003)
 - a set of services that a **university offers to the members of its community** for the management and dissemination of digital materials created by the institution and its community members
- JISC (2016)
 - A repository is a set of services[1] that a research organisation[2] offers[3] to the members of its community[4] for the management and dissemination[5] of digital materials[6] created by its community members
 - 1～6についてより詳細な定義（引用）を提供
- Herbert Van de Sompel (2016)
 - The purpose/mission of repositories is no longer well defined and has IMO to quite an extent drifted since their original inception from being about "all kinds of digital materials created by an institution's staff" to "formally published materials created by an institution's staff". In this drift lies (IMO) one of the major problems of many current IRs: they don't provide a service to their local community because scholars don't care about stuffing their papers in the IR once they were able to get it published in some journal. Under the right conditions (global visibility), they might care about stuffing other materials there; materials they currently drop in e.g. Figshare.

Next Generation Repository

- COAR内に設置された次世代リポジトリに必要とされる要件を考えるWG
- 基本的な考え方の方向性
 - リポジトリを学術情報流通システム全体のランドスケープの中にきちんと位置づけることにより、現在の学術情報流通システムの変容を促したい。
 - そのためには、世界中に分散したリポジトリを連携させ、そのグローバルなネットワークの基盤の上に、さまざまな付加価値サービス（査読、注釈、コメント等）のレイヤを展開する必要がある。
 - こうしたグローバルなリポジトリのネットワークを発展させるには、人だけでなくマシン・ユーザを想定した最新の標準的なウェブ技術を採用した次世代リポジトリが必要となる。

Petr Knoth: COAR Next Generation Repositories - User stories and current developmentsより



Slide by
Petr Knoth: COAR Next Generation Repositories - User stories and current developments,
<https://www.slideshare.net/petrknoth/coar-next-generation-repositories-user-stories-and-current-developments>

旧：メタデータと本文ファイルを公開するのが主目的
新：公開したメタデータと本文に価値を与える機能群を提供するのが主目的

オープンアクセスからオープンサイエンス

- 2013年6月
 - G8科学大臣会合における研究データのオープン化を確約する共同声明 ⇒ 日本調印
- 2015年3月
 - 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 ⇒ 2015年度以降はフォローアップ検討会を実施
- 2016年1月
 - 第5期科学技術基本計画 ⇒ 知の基盤の強化に向けてオープンサイエンスを推進
- 2016年2月
 - 科学技術・学術審議会 学術分科会 学術情報委員会「学術情報のオープン化の推進について（審議まとめ）」
- 2016年7月
 - 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「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に関する提言」
- 2017年6月
 - 科学技術イノベーション総合戦略2017「オープンサイエンスの推進」

オープンサイエンスの実現に向け研究データ管理が重要視される時代

研究データ管理

- 一般的にある研究プロジェクトにおいて使用された、あるいは生成された情報を、どのように組織化、構造化、保管、管理していくのかを指す言葉
- 研究データ管理には次のことが含まれる：
 - どのようにデータを取り扱っていくかの計画（Data Management Plan）策定
 - 研究プロジェクトが行われている間の、日々の情報の取り扱い方
 - 長期的にデータをどのように取り扱うか（研究プロジェクトが終わった後はそのデータをどうするか）

CMS (NC2) ベースで構築した意図

- Ajaxによる直感的で軽快な操作感の提供
 - WEKOの強み
- UIからの管理機能の提供 (NC2の思想)
 - WEKOの強み
- インストールやアップデートの簡便性
 - Jairo Cloudによる位置づけの変化
- 多様なモジュール (機能) と組み合わせた個性のある機関リポジトリの構築
 - **それほど使われているわけではない** ←コミュニティウエア
- Group, Privateスペースを活用した研究利用
 - **それほど使われているわけではない** ←研究データ管理

JCの拡大

OS時代のNGR ⇒ 更に機能強化したシステムが必要

話の流れ

- 現状のリポジトリシステム
 - NIIで構築してきたWEKOとJairo Cloudについて
- 解決したい課題
- **次世代リポジトリシステム**
- 次世代システムと図書館の役割

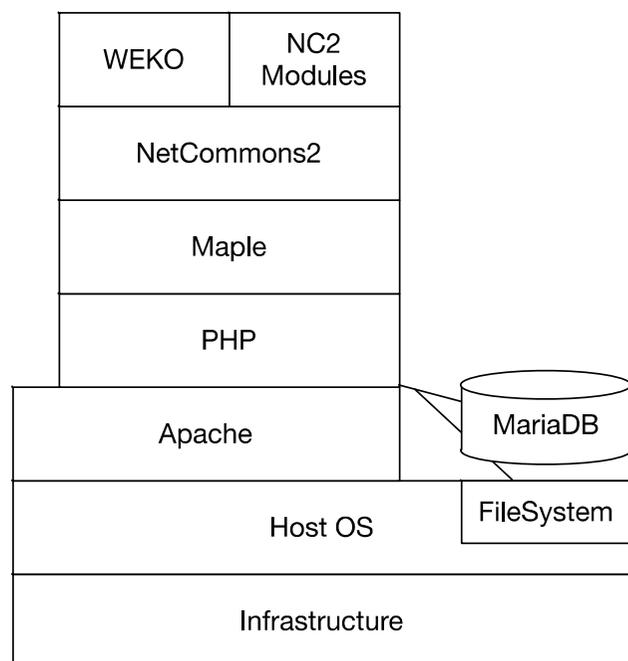
既存のリポジトリの比較調査

name	Lang.	Framework	CMS	DB	Metadata Store	Full text Search	OR mapping	Message queue	Message broker	etc
DSpace	Java	Cocoon	-	PostgreSQL	PostgreSQL	Solr	Hibernate	-	-	Configurable Workflow
Eprints	Perl	-	-	MySQL	MySQL	-	-	-	-	-
HUBzero	PHP	Joomla!	Joomla!	MySQL	MySQL	-	-	-	-	simulation/ modeling tool
Zenodo	Python	Invenio	-	PostgreSQL	PostgreSQL (JSON)	elasticsearch	SQLAlchemy	Celery	Rabbitmq	-
invenio	python	flask	-	PostgreSQL	PostgreSQL (JSON)	elasticsearch	SQLAlchemy	Celery	Rabbitmq	-
Hyrax	Ruby + Java	Samvera+ Fedora	-	-	Fedora (JDBC)	Solr	ActiveFedora	Sidekiq	Redis	-
Hyku	Ruby + Java	Hyrax	-	PostgreSQL	Fedora (JDBC)	Solr	ActiveFedora	Sidekiq	Redis	IIIF
Islandora	PHP+ Java	Drupal+ Fedora	Drupal	-	Fedora (JDBC)	Solr	ActiveFedora	-	-	-
fedora	java	modeshape	-	JDBC	JDBC	-	-	-	-	API-X
dataverse	Java	-	-	PostgreSQL	PostgreSQL	Solr	-	-	-	R package
WEKO2	PHP	Maple	NC2	MariaDB	MariaDB	MariaDB (Mroonga)	-	-	-	-

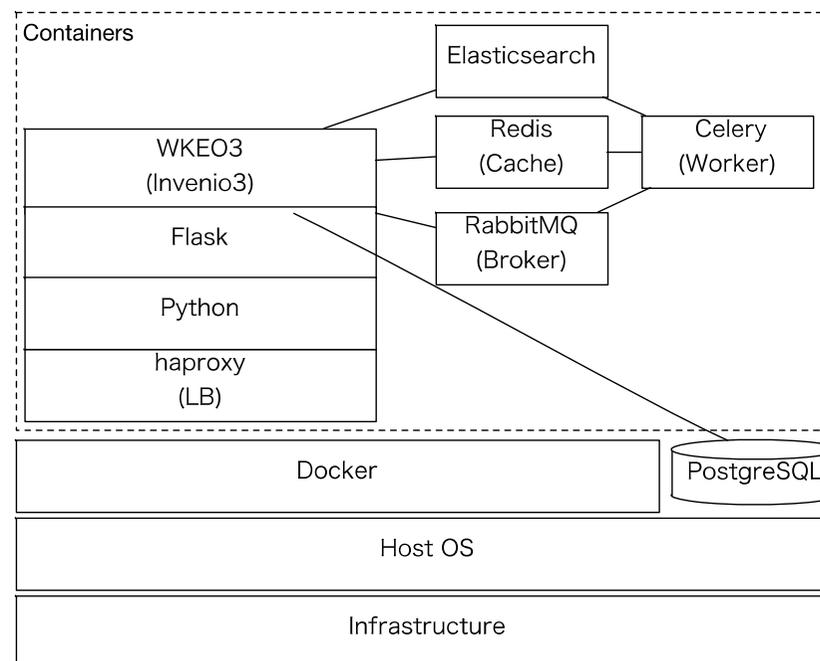
リポジトリシステムのマイクロサービス化

WEKO3のアーキテクチャイメージ

WEKO2

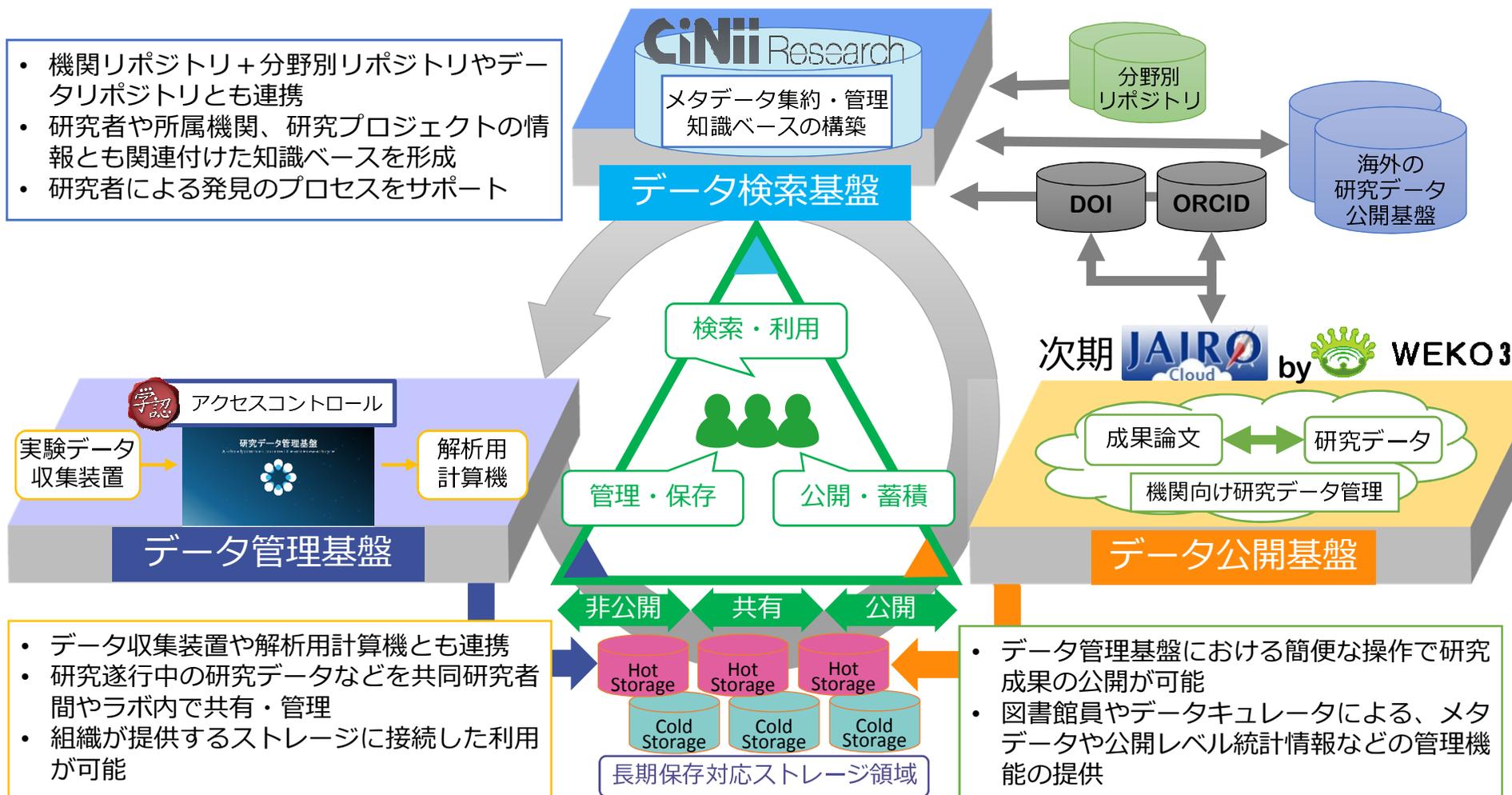


WEKO3



マイクロサービス化によりシステム構成の柔軟性を確保

NIIが構築中の研究データ基盤 データ管理機能と公開機能の分離

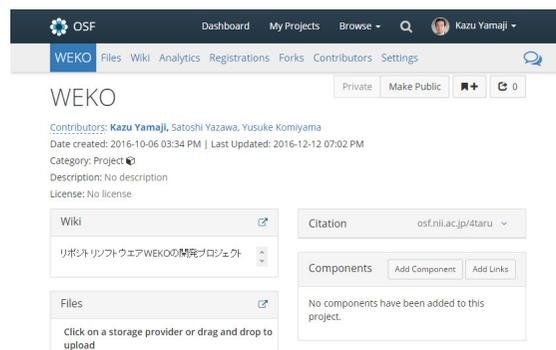


GakuNin RDM

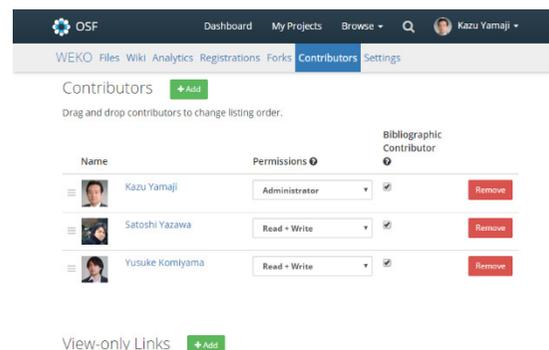


Copyright © 2016 National Institute of Informatics

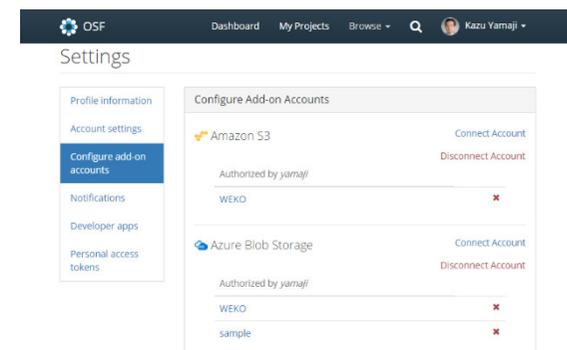
研究プロジェクト単位で
ファイルなどを管理



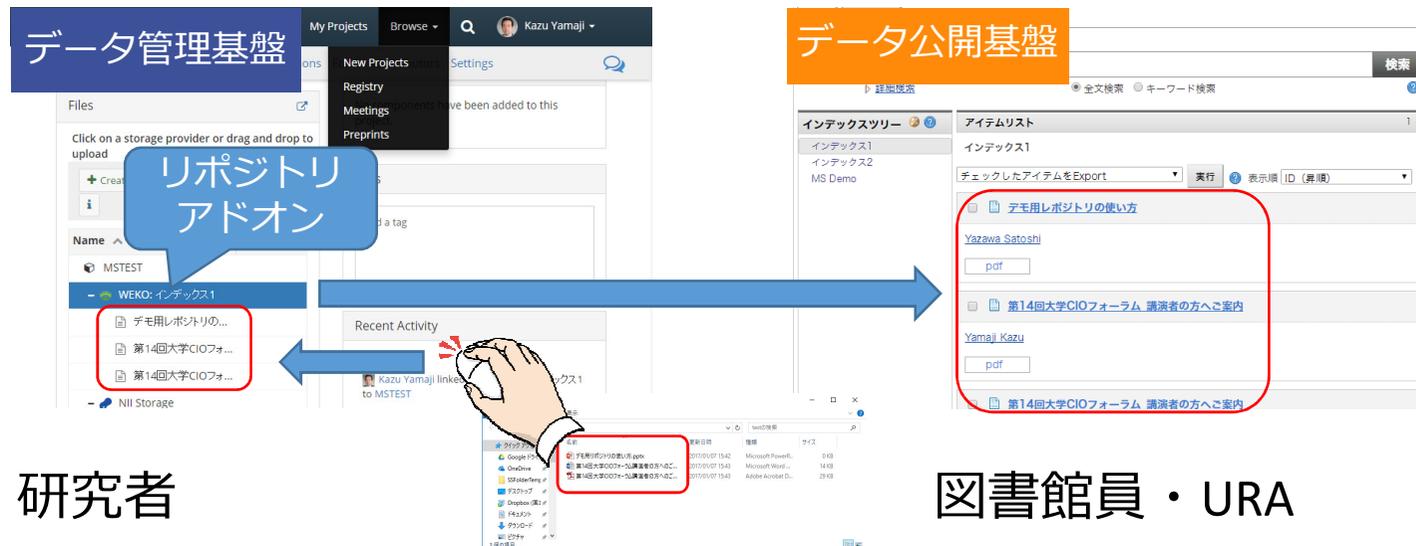
学認と連携しVO(仮想組織)
メンバーでファイルを共有



豊富なアドオンで必要な
クラウドストレージを接続



データ管理基盤 v.s. データ公開基盤



研究者

- データ管理機能
- メタデータ管理機能
- 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
- ...

どちらがどのような機能を提供するのが実ワークフローに最適か？

JPCOAR研究データTFの活動

- 2016年度

機関リポジトリ推進委員会 研究データTF

- 研究データ管理（RDM）トレーニングツールの開発
- ケーススタディによる研究データ管理ノウハウの蓄積

- 2017年度

JPCOAR 研究データTF

- RDMトレーニングツールの開発
- データベースレスキュー
- データリポジトリ

RDMトレーニングツール教材概要

- 目的
 - 学習者が研究データ管理に関する基礎的な知識を習得する
 - 各機関における研究データ管理サービス構築の足掛かりを得る
- 内容
 - 全7章構成：第1章 導入／第2章 データ管理計画（DMP）／第3章 保存と共有／第4章 組織化、文書化、メタデータ作成／第5章 法・倫理的問題／第6章 ポリシー／第7章 研究データ管理サービスの検討
 - 講義スライド、スクリプト
- 対象者
 - 図書館、IT部門、研究支援部門の職員（URA）等
 - オープンサイエンスや研究データ管理に関心のある教職員

無料オンライン講座「オープンサイエンス時代の研究データ管理」

2017年11月15日開講にむけ、現在受講者募集中

https://lms.gacco.org/courses/course-v1:gacco+ga088+2017_11/about

みなさまのご参加をお待ちしています！

