

# Open Data Infrastructure

Masaharu Hayashi, Yohei Hayashi  
National Institute of Informatics

# Number of Institute Repositories in Japan



507 / 789 IRs

IR built on JAIRO Cloud  
(include number of under built)  
(2017/6/1)

# What is JAIRO Cloud?

- SaaS-based system environment of institutional repository provided by NII and JPCOAR
  - Institutions pay some usage fee
- Built based on WEKO(software) developed by NII
  - WEKO is a repository module on NetCommons2(NC2) which is a sort of contents management system and groupware developed by NII

# Community site of JAIRO Cloud



### What's New 新着情報

に過去の情報が掲載されていますのでご確認ください。

- 件名に「はなまる」がついた質問は解決済みです。
  - 投稿いただいたスレッドについて解決した場合、投稿者は未解決記事のアイコンに「はなまる」マークをつけてください。【お手紙はこちら】
- 「みんなで作るマニュアル」は閉鎖いたしました。
  - この投稿はご存知でありました内容につきましては、過去の投稿に譲っていただいております。今後、内容を見直した上で、「マニュアル版」のページ等に転載することを予定しております。

### フォーラム (参加機関用)

ユーザーサポート(既記事一覧)

プラットフォームへ 1 2 3 4 5 次 ▶ 20件

● 一括登録にて、登録先が「指定」になる(2件)	20170609 (2件)
● トップページが変更になった(2件)	20170609 (2件)
● 博士論文のホームページのエラーについて(2件)	20170616 (2件)
● トップページ項目名の修正について(4件)	20170617 (4件)
● 新規アカウントの登録について(2件)	20170407 (2件)
● アドレスの一括登録について(2件)	20170418 (2件)
● ホームページの一括登録について(2件)	20170417 (2件)
● 更新履歴について(2件)	20170407 (2件)
● エディタバックアップの取崩しについて(1件)	20170117 (1件)
● アドレス変更履歴について(1件)	20170314 (1件)
● 更新履歴について(1件)	20170310 (1件)
● 対応のスマートフォンでの表示について(2件)	20170316 (2件)
● 一括登録用ソフト5000のバージョン確認方法について(2件)	20170306 (2件)
● 学内限定コンテンツの登録について(2件)	20170324 (2件)
● サイト更新について(2件)	20170324 (2件)
● エディタタイプが変更されない(2件)	20161130 (2件)
● 1件だけの検索がとれないアドレスについて(1件)	20170320 (1件)
● 新規登録もメニューから削除のアクセスできない(2件)	20170127 (2件)
● 更新履歴が非公開になってしまふ(2件)	20170301 (2件)
● エディタアクセス時のアドレスを一括登録サポートする(2件)	20170116 (2件)

プラットフォームへ 1 2 3 4 5 次 ▶ 20件

### サービス概要

JAIRO Cloudは、平成24年度より運用開始されたクラウド型の機関リポジトリ環境提供サービスです。従となる機関リポジトリソフトウェアに開発されたWEXOを採用しています。利用機関はシステム・サーバー管理から解放され、コンテンツ登録に集中できます。利用申請は随時受け付けます。

### JAIRO Cloud導入機関数

公開から閉鎖機関申請中の機関を合わせ

# 507機関

のみならずお使いいただけます。(2017/03/31現在)



# IRs based on JAIRO Cloud

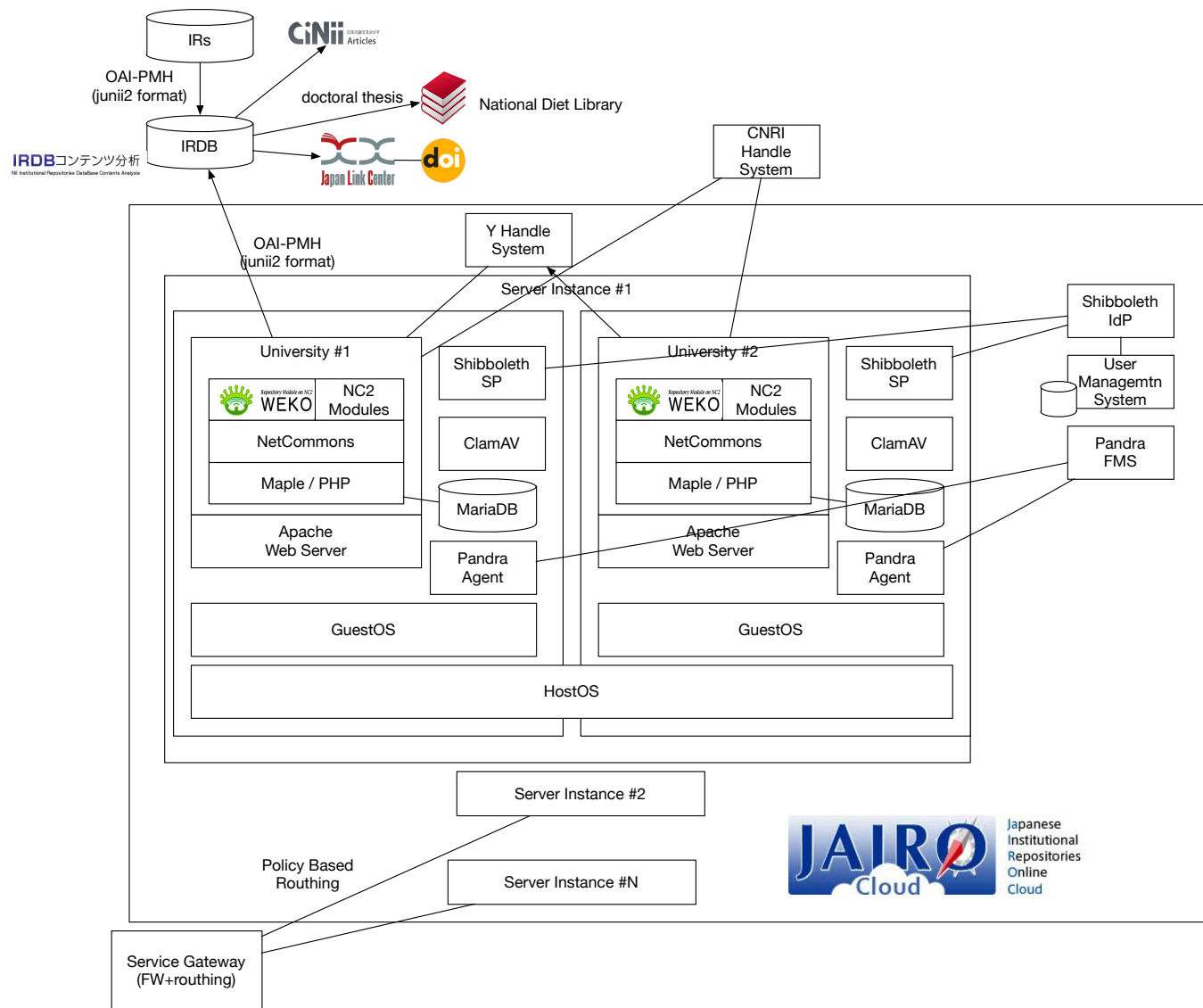
The screenshot shows the homepage of the Tsukuba Repository. At the top left is the logo for 'つくばリポジトリ University of Tsukuba Repository'. Below the logo is a navigation bar with links: 'リポジトリトップ', 'リポジトリについて', 'File Upload', 'Q&A', and 'お問い合わせ'. The main content area features a search bar with '検索' and '戻る' buttons, and a sidebar with 'Language' (set to Japanese) and 'インデックスツリー' (Index Tree) listing various faculties and their document counts.

<https://tsukuba.repo.nii.ac.jp/>

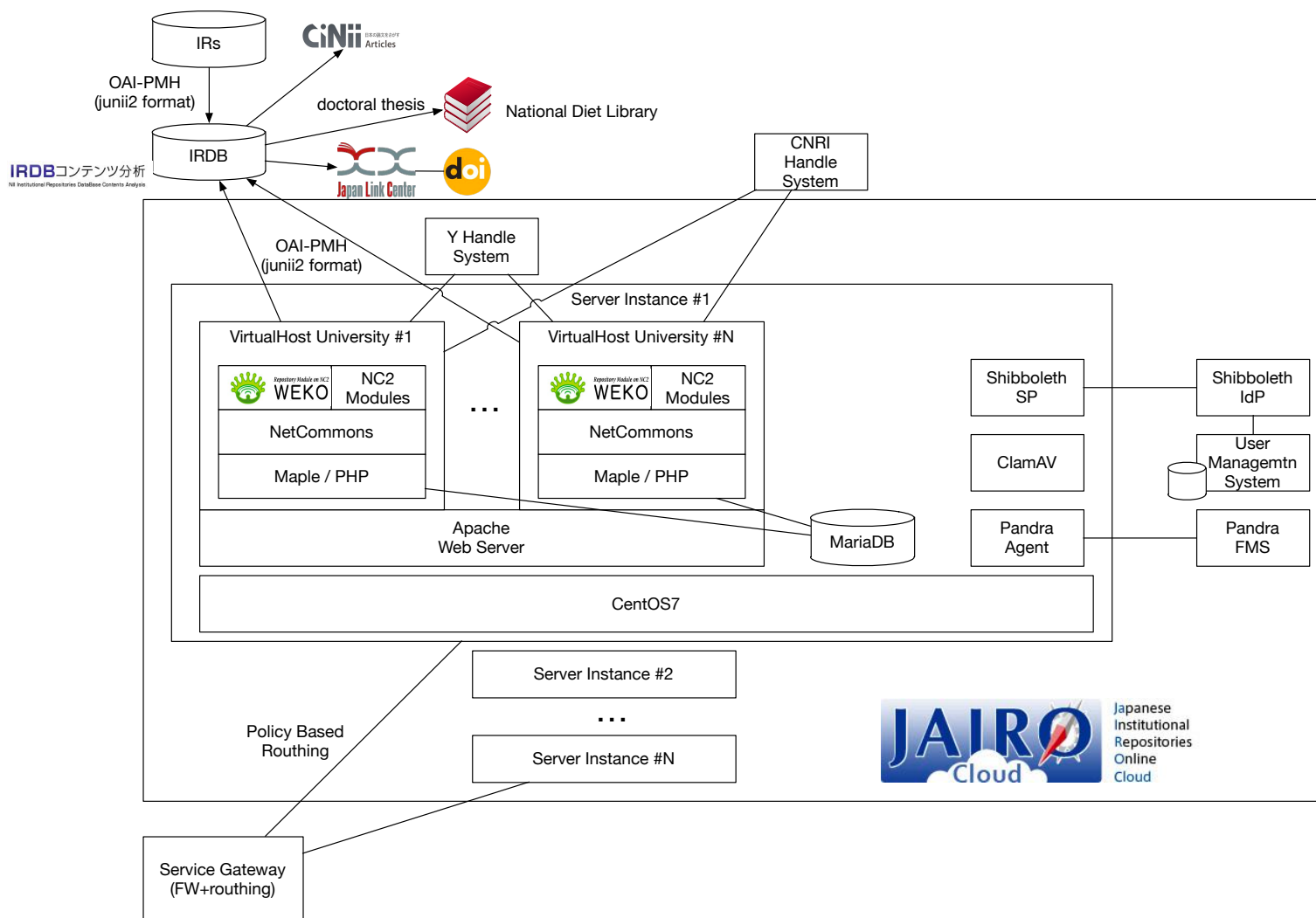
The screenshot shows the search results page of the Joensuu University Repository. The header reads '上越教育大学リポジトリ'. The search bar contains '092' and shows '092 高田博昭'. Below the search bar is a list of items, including '高田博昭: 第140号: 昭和元年十二月期用書通帳, 昭和元年12月' and '高田博昭: 第139号: 昭和元年十一月期用書通帳, 昭和元年11月'. Each item has a 'pdf' button. The right sidebar contains information about the repository and navigation links.

<https://juen.repo.nii.ac.jp>

# System Overview of JAIRO Cloud (OLD)



# System Overview of JAIRO Cloud



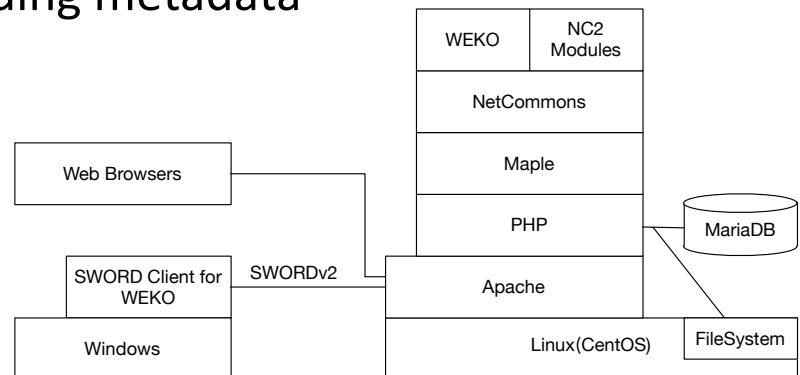
# What is WEKO?

- “out-of-the-box” repository module on NC2
- Functionality
  - It has almost all functionalities you need as a repository system (such as directory/keyword search, metadata management, ingest workflow support, OAI-PMH provider and harvester, multimedia data viewer(image, audio, video and PDF), contents versioning )
- Easy to use
  - All functions can be customized and operated by web browser
  - SWORD Client for WEKO(SCfW) can deposit multiple contents to WEKO with MS Excel file including metadata



*Repository Module on NC2*

# WEKO





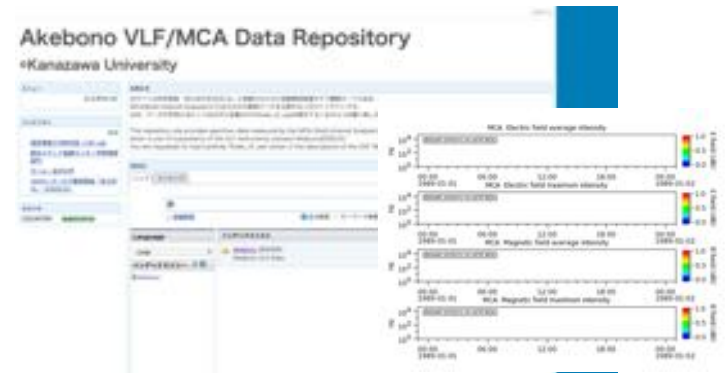
# Other Application using WEKO

Online Publication repository  
in academic society



<https://ipsj.ixsq.nii.ac.jp/>

Aurora Observation Satellite Data repository



<https://akebono-vlf.db.kanazawa-u.ac.jp/>



<https://jsai.ixsq.nii.ac.jp/>

Museum resources repository



<https://sci-instrument.repon.org/>  
<https://wallchart.repon.org/>

# Issues of Current JAIRO Cloud

- It is not yet ready for Research data publication
  - WEKO is mainly designed as Journal Repository, it has a limit to supporting to the diversity of research data
- It is difficult for system maintenance
  - We have not enough technical documentation and depend on knowledge of limited developers
- It has limitations on scalability and extensibility
  - WEKO is based traditional web application architecture, it is difficult to scale out
- It is necessary to replace in 1 or 2 years
  - Upcoming NC2 End of Life, first version of NC3 was released at October 31, 2016

We need to develop next WEKO2  
(the following referred to as WEKO3)



# WEKO3 Design Assumptions

- We must support current JAIRO Cloud functions
  - Over 500 IRs are built/under built on current JAIRO Cloud
- We must support supplemental data of journal articles and various research data
  - We must support next junii2 schema, JPCOAR schema is designed for supporting research data
- We should provide a system that can be used for 5 or 10 years or more
  - We need to satisfy the requirements of COAR Next Generation Repository
- We should provide a system that can be scalable, extensible and easy-to-manage and deploy
  - We are going to apply contentious integration (Devops) of software development method

# Current Schedule

- April - May ,2017 (delay)
  - Organizes functions of JAIRO Cloud
- June - July ,2017 (small delay)
  - Compares functions and architecture of existing repository software
  - Organizes functions of WEKO3
  - Defines specifications of WEKO3
- August – September, 2017
  - Starts to development of WEKO3
- March, 2018
  - Releases first version of WEKO3
- April, 2019
  - Starts an operation of next JAIRO Cloud(WEKO3)

# Comparison of existing repository software for development of WEKO3

- DSpace
- Eprints
- HUBzero
- Zenodo (invenio)
- Hyrax or Hyku (Hydra/Fedora Commons)
- Islandora (Fedora Commons)
- Dataverse

# Comparison matrix(draft)

name	Lang.	Framework	CMS	DB	Metadata	Fulltext search	OR mapping	Message queue system	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 + Flask	-	PostgreSQL	PostgreSQL (JSON)	elasticsearch	-	Rabbitmq, celery	-
Hyrax	Ruby + Java	Samvera+ Fedora	-	-	Fedora(JDBC)	Solr	ActiveFedora	Sidekiq	-
Hyku	Ruby + Java	Hyrax	-	PostgreSQL	Fedora(JDBC)	Solr	ActiveFedora	Sidekiq	IIIF
Islandra	PHP+ Java	Drupal+ Fedora	Drupal	-	Fedora(JDBC)	Solr	ActiveFedora	-	-
dataverse	Java	-	-	PostgreSQL	PostgreSQL	Solr	-	-	R package
WEKO2	PHP	Maple	NC2	MariaDB	MariaDB	MariaDB (Mroonga)	-	-	-

# Conclusion

- We are designing WEKO3 Now
- We are comparing existing repository softwares for development of WEKO3
  - Currently we are considering “invenio” and “hydra” as candidates of base system of WEKO3
  - Both repository softwares are “microservice architecture” and suitable for our requirements
  - “microservice architecture” is an approach to developing a single application as a suite of small services
- We will continuously design and develop for WEKO3 and aim for the realization of Devops





Thank you