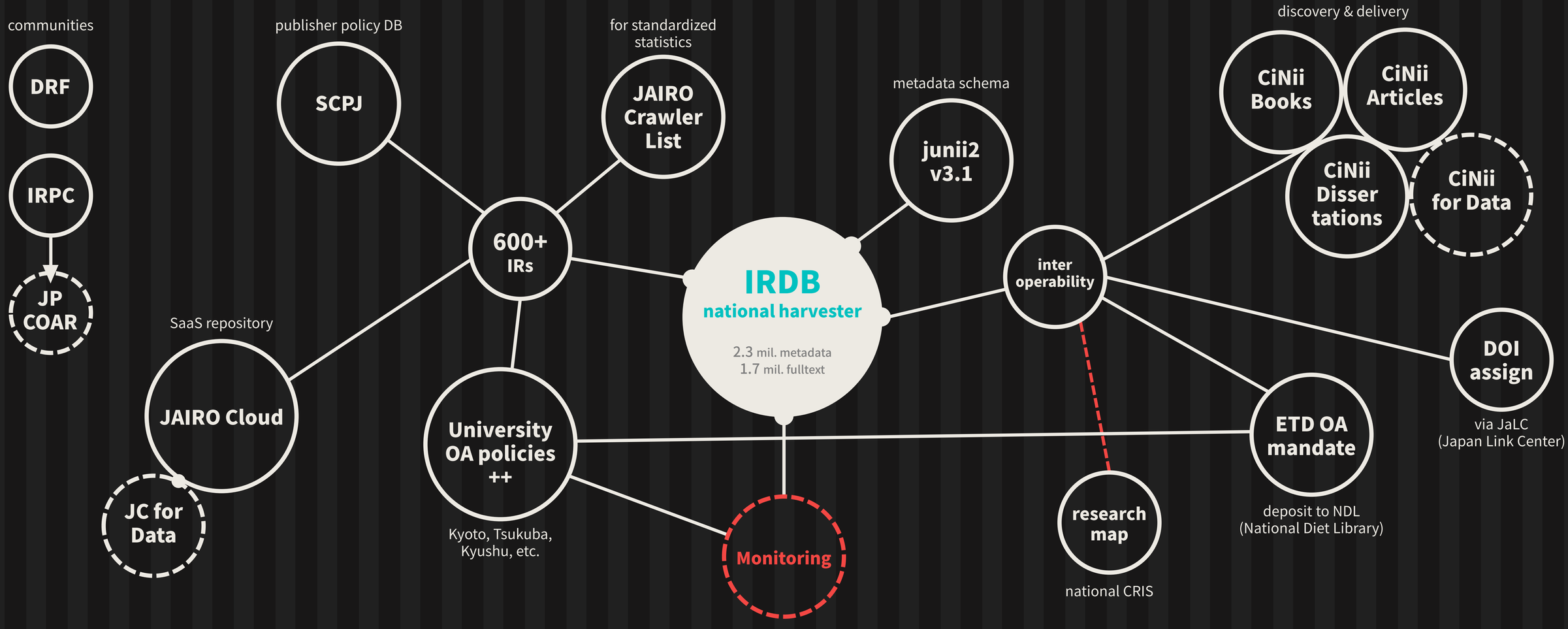
Emerging Infrastructure for Open Access Monitoring in Japan

Through our 10+ years experiences, we believe a key missing piece for driving OA is monitoring of OA activities.

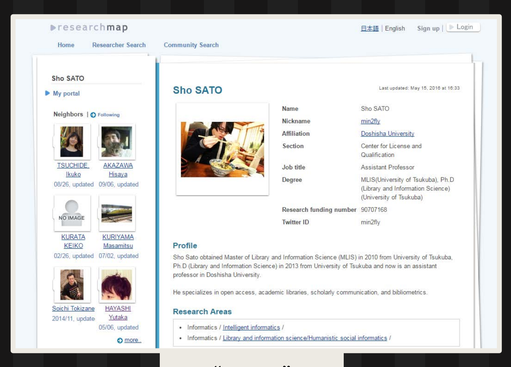


So we launched a project in 2015 to develop OA monitoring infrastructure, and tried two approaches first.

1. CRIS-based approach: matching "researchmap" and IRDB

http://researchmap.jp/

- a national CRIS service (SaaS)
- since 1998
- operated by JST & developed by NII
- 240,000 researcher profiles
- millions of metadata: papers, presentations, misc.
- few fulltext files



"paper" de-duplicated **1,080,556 records**

Objective
Calculating the number of metadata records in researchmap which exist also in IRDB (Green OA).

Method
Matching the two set by an n-gram based original algorithm with some threshold values.

a. the subset of de-duplicated metadata in researchmap of type "paper" (# = 1,080,556)
b. the subset of de-duplicated metadata in IRDB of type "journal article" and with fulltext (# = 1,525,592)

Result
65,589 records in (a) are matched to records in (b). In this case, the open access ratio is 6.1%.

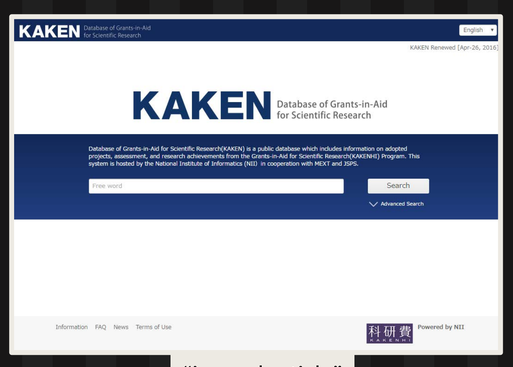
6.1% (65,589 records) matched

Discussion
- Relatively low metadata coverage of researchmap is problematic.
- Journal articles may be included in other content types than "paper".
- Gold OA status cannot be understood by the IRDB-based approach.

2. Funder-based approach: matching "KAKEN" and IRDB

http://kaken.nii.ac.jp/en

- a database of research projects funded by KAKENHI (MEXT/JSPS), the largest research fund in Japan (FY2015: 19.1 billion €) [1€ = 120 yen]
- since 1996
- operated & developed by NII
- 820,000 projects
- millions of metadata



"journal article" with DOI **467,082 records**

Objective
Calculating the number of metadata records in KAKEN database which exist also in IRDB (Green OA).

Method
Matching the two set by a DOI-based algorithm.

a. the subset of metadata in researchmap of type "journal article" and with DOIs (# = 467,082)
b. the subset of metadata in IRDB of type "journal article" and with DOIs and fulltext (# = 1,525,592)

Result
11,118 records in (a) are matched to records in (b). In this case, the open access ratio is 2.4%.

2.4% (11,118 records) matched

Discussion
- The result is limited to KAKEN-funded research outputs.
- Expansion to journal articles without DOIs or other content types is needed.
- Gold OA status cannot be understood by the IRDB-based approach.

IRDB

1,525,592 records with fulltexts de-duplicated

46,676 records "journal article" with fulltexts, DOI

Lessons learned and our future plans

1. Update metadata schema

We plan to complete updating our standard metadata schema "junii2" by the end of FY2016, to precisely describe
- OA status (Green, Gold, Embargoed, Closed, ...)
- Funding information (Funder id, Project id)
- Research Data metadata

in a way harmonized with OpenAIRE or RIOXX.

2. Capture Gold OA

Our government takes Green route, but the number of Gold OA articles is increasing. Only by the IRDB-based approach we cannot understand our present Gold OA situation. Thomson's InCites gives us some figures, but research outputs written in Japanese (language) are not indexed in Web of Science. Some different approaches are needed.

3. Develop "Research Product Database"

So now we are developing a "Research Product Database" to increase coverage and refine the calculation of OA ratio. It could collect metadata of research output created by Japanese researchers from multiple sources, integrate and enrich them, and manage with OA status and funding information. Interoperability with OpenAIRE is considered.

HAYASHI Yutaka* (Kyushu University Library, Japan), ORCID: 0000-0001-7761-3444
MISUMI Kenichi (Obihiro University of Agriculture and Veterinary Medicine Library, Japan)
KATAOKA Shin (National Institute of Informatics, Japan), ORCID: 0000-0002-6435-8445
*Corresponding author: hayashiyutaka@gmail.com
Our project is funded by Institutional Repositories Promotion Committee (IRPC), Japan.

Open Repositories 2016 (OR2016)
Trinity College Dublin, Ireland, June 13-16, 2016