

# System for Cross-Organization Big Data Analysis of Japanese Institutional Repositories

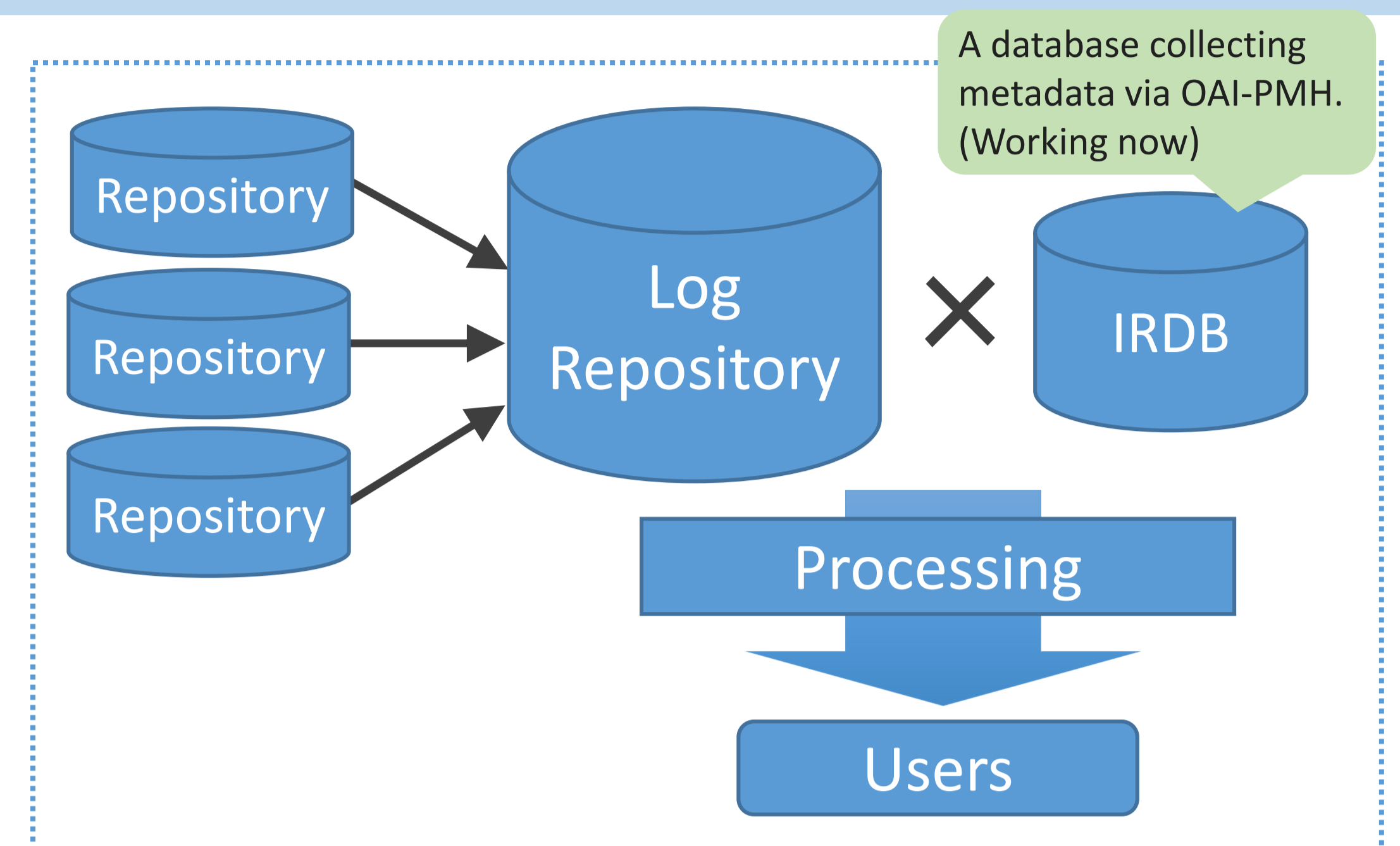
Takuro Kawamura<sup>\*1</sup>, Kenichi Igarashi<sup>\*2</sup>, Hiroshi Kato<sup>\*3</sup>, Akira Maeda<sup>\*3</sup>, Toshihiro Aoyama<sup>\*4</sup>, Kazutsuna Yamaji<sup>\*3</sup>, Sho Sato<sup>\*5</sup>

<sup>\*1</sup>Hiroshima University, tkawa@hiroshima-u.ac.jp <sup>\*2</sup>Keio University <sup>\*3</sup>National Institute of Informatics

<sup>\*4</sup>National Institute of Technology, Suzuka College <sup>\*5</sup>Doshisha University

## Introduction

We are developing a system that use access log of institutional repositories(IRs) and metadata of contents. Major work of the system is data analysis. This function enables cross-organization comparison and standardization of statistics. The system is also attached other functions to promote utilization of IRs.



## Development

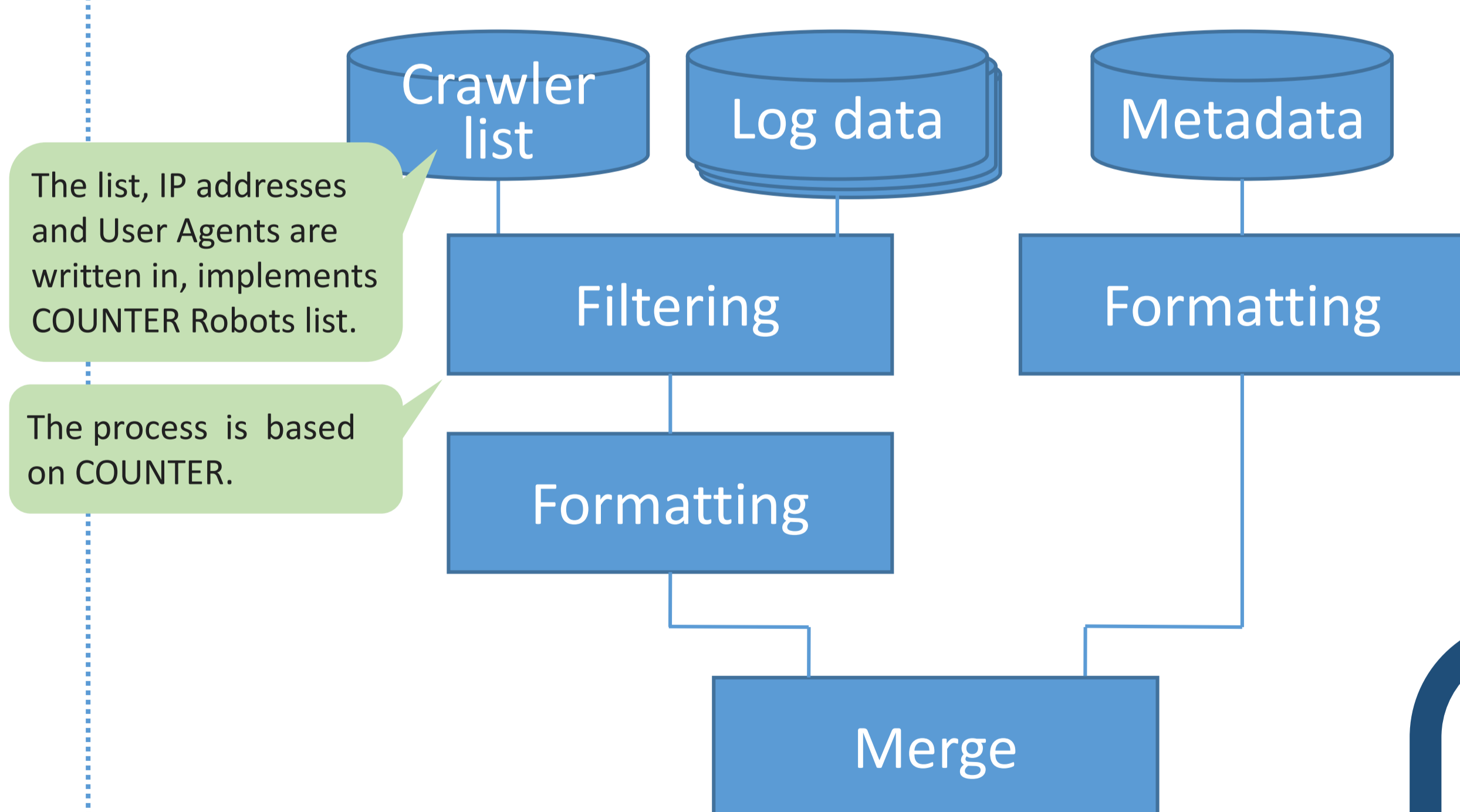
The project has just started. We are discussing the detail of the service, creating a prototype with the data of JAIRO Cloud on Microsoft Azure. Our current plan is as follows.

### What's JAIRO Cloud?

It is a SaaS-type cloud service for IRs. The service have a repository software module, WEKO. More than 250 institutions in japan use it.



## Dataflow



An example of aggregated data

repo	Freq
rikkyo.repo.nii.ac.jp	5505
sapporo-u.repo.nii.ac.jp	3026
repo.lib.tut.ac.jp	2257
ncu.repo.nii.ac.jp	2046
redcross.repo.nii.ac.jp	1646
shonan-it.repo.nii.ac.jp	1554
fukuyama-u.repo.nii.ac.jp	1452

An example of API's response

```

{
  "Results": {
    "output": [
      {
        "Type": "Table",
        "Table": {
          "ColumnNames": [
            "repo",
            "Freq"
          ],
          "ColumnTypes": [
            "String",
            "Int32"
          ],
          "Values": [
            [
              "rikkyo.repo.nii.ac.jp",
              "5505"
            ],
            [
              "sapporo-u.repo.nii.ac.jp",
              "3026"
            ],
            [
              "repo.lib.tut.ac.jp",
              "2257"
            ],
            [
              "ncu.repo.nii.ac.jp",
              "2046"
            ],
            [
              "redcross.repo.nii.ac.jp",
              "1646"
            ],
            [
              "shonan-it.repo.nii.ac.jp",
              "1554"
            ],
            [
              "fukuyama-u.repo.nii.ac.jp",
              "1452"
            ],
            [
              "shonan-it.repo.nii.ac.jp",
              "1422"
            ],
            [
              "fukuyama-u.repo.nii.ac.jp",
              "1283"
            ],
            [
              "waseda.repo.nii.ac.jp",
              "1241"
            ],
            [
              "riko.repo.nii.ac.jp",
              "1191"
            ]
          ]
        }
      }
    ]
  }
}
    
```

## Functions

### Usage statistics for repository manager

Report of each IR  
Report of Japanese IRs

There are now no tools that enable comparison between IRs or analysis of all the entire IRs in Japan. The module will allow those.

### Ranking for institution, author

Institution ranking  
Content ranking

The number calculated by different method of each institution is often used now. This feature will provide meaningful information by common processing.

### Recommender for visitor

Similar Items

The function will promote use of content.

### Research for the system, IRs' future

Improvement this/IR system  
Mining knowledge

The system will provide research environment that enables us to use raw-close data.

Combined data

IP	weko	action.type	action.item	repo	actionDate	dayOfWeek	hourOfDay	repoIRID_ID	item_機関名	タイトル	NIItpe	pub_year
xxx.xxx.xxx.xxx	weko	repository_action_common_download	00001111	xxx.repo.nii.ac.jp	2015-01-01T08:44:22Z	yday	5	xxx.repo.nii.ac.jp:00001111	NII大学	鳥居式エグゼクティブと赤塚式スキームの発展	Departmental Bulletin Paper	2030/12/31
xxx.xxx.xxx.xxx	weko	repository_action_common_download	00001201	xxx.repo.nii.ac.jp	2015-01-01T08:44:22Z	yday	5	xxx.repo.nii.ac.jp:00001201	NII大学	MLA選定におけるそれぞれの役割	Departmental Bulletin Paper	2022/12/31
xxx.xxx.xxx.xxx	weko	repository_action_common_download	00002571	xxx.repo.nii.ac.jp	2015-01-01T08:44:22Z	yday	5	xxx.repo.nii.ac.jp:00002571	NII大学	船橋市の私立図書館と市立図書館	Departmental Bulletin Paper	2012/12/31
xxx.xxx.xxx.xxx	weko	repository_action_common_download	00001731	xxx.repo.nii.ac.jp	2015-01-01T08:44:22Z	yday	5	xxx.repo.nii.ac.jp:00001731	NII大学	船橋市の私立図書館の存続について	Departmental Bulletin Paper	2045/12/31
xxx.xxx.xxx.xxx	weko	repository_action_common_download	00001732	xxx.repo.nii.ac.jp	2015-01-01T08:44:22Z	yday	5	xxx.repo.nii.ac.jp:00001732	NII大学	曇天文庫と著作権	Departmental Bulletin Paper	2023/12/31
xxx.xxx.xxx.xxx	weko	repository_action_common_download	00001121	xxx.repo.nii.ac.jp	2015-01-01T08:44:22Z	yday	5	xxx.repo.nii.ac.jp:00001120	NII大学	エンディングモーションと資本理論	Departmental Bulletin Paper	2022/12/31
xxx.xxx.xxx.xxx	weko	repository_action_common_download	00001320	xxx.repo.nii.ac.jp	2015-01-01T08:44:22Z	yday	5	xxx.repo.nii.ac.jp:00001320	NII大学	知的管理推進機関はいかに成果を挙げたか	Departmental Bulletin Paper	2042
xxx.xxx.xxx.xxx	weko	repository_action_common_download	00001322	xxx.repo.nii.ac.jp	2015-01-01T08:44:22Z	yday	5	xxx.repo.nii.ac.jp:00001322	NII大学	和紙と紙魚の関係	Departmental Bulletin Paper	2042/12/31
xxx.xxx.xxx.xxx	weko	repository_action_common_download	00001124	xxx.repo.nii.ac.jp	2015-01-01T08:44:22Z	yday	5	xxx.repo.nii.ac.jp:00001124	NII大学	機関リポジトリにおけるログデータの活用	Departmental Bulletin Paper	2018/12/31