

# ZS Project

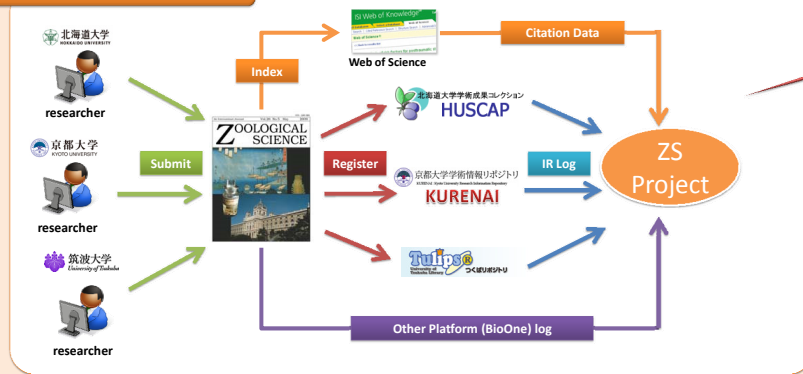
Sho SATO †, Yuji NONAKA ††, Masato ONISHI †††, Hiroshi ITSUMURA †  
 † Graduate School of Library, Information and Media Studies, University of Tsukuba  
 †† Library, Information systems division, Hokkaido University  
 ††† Electronic Resources Section, Technical Service Division, Kyoto University Library

## Usage log analysis of the contents of institutional repositories: user domains, types of referrals and content attributes

### Introduction

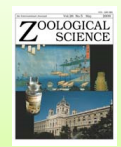
- I. What impact do institutional repositories have upon Open Access (OA)?
- II. Do articles made publicly available via institutional repositories inspire new citations?

#### Project Overview



#### ZOOLOGICAL SCIENCE

- ZOOLOGICAL SCIENCE was founded in 1984.
- It is published by the Zoological Society of Japan and devoted to the publication of original articles, reviews, essays and short communications in the broad field of zoology.
- The full text of all recent issues of the journal can be found on the Internet at BioOne2, UniBio Press, etc.



**JCR 2008 Data**  
 Total Cites: 2072  
 IF: 1.100  
 5-Year IF: 1.227  
 Articles: 152  
 Cited Half-life: 6.4

#### Analysis

- In cooperation with the Zoological Society of Japan, of the articles published in ZOOLOGICAL SCIENCE, we selected the papers written by researchers belonging to each university, and deposited them in the **author's own institutional repository**.
- Analysis of the usage logs from institutional repositories and BioOne in order to measure **how many times the articles have been downloaded**.
- Organization the changes in the "Times Cited Count", which indicates the number of times a published article has been cited by another papers in the Web of Science, and examination of the effects on citation status by cross-checking with the above usage logs.

As a preliminary step of the ZS project, we extracted and analyzed the 2008 usage logs from each repository in order to understand more about the current status of institutional repository usage.

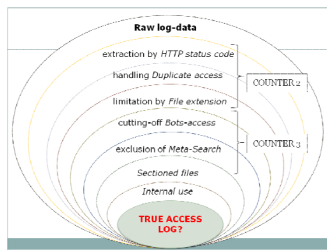
### Method

- Analysis of relationships between the use of content and **user domains, types of referrals, and content attributes** in four institutional repositories.
  - Hokkaido University: **HUSCAP (Contents: 25,542 items)**
  - Kyoto University: **KURENAI (28,536)**
  - Tsukuba University: **Tulips-R (7,899)**
  - IDE-JETRO: **ARRIDE (640)**
- Total downloaded items: **1,150,813**
- Analysis of the **2008 usage logs** from each repository

IP address    Access time    File name    HTTP status code

133.51.6.132 - [27/Jul/2008:07:57:01 +0900] GET /dspace/handle/2433/49548/1/ HTTP/1.1 200  
 424264 - Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; InfoPath.2; .NET CLR 2.0.50727)

#### Log Filtering Procedure



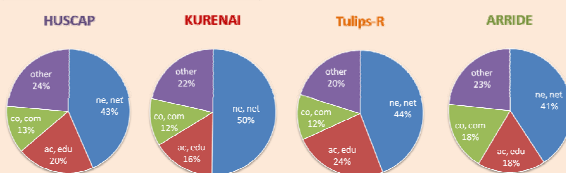
The original version is Japanese and English version was offered by Y. Sato.

#### Filtering Results

	HUSCAP	KURENAI	Tulips-R	ARRIDE
Raw log-data	13,885,718	15,756,128	5,207,989	2,198,116
Downloaded PDF log (100%)	790,653	1,872,902	595,811	95,824
Extraction by HTTP status code (200, 304)	95.6%	98.5%	99.2%	86.8%
Handling Duplicate access	86.6%	88.7%	85.6%	81.2%
Cutting-off Bots-access	50.8%	32.6%	24.3%	19.3%
Cutting-off Internal use (except ARRIDE)	50.3%	32.0%	23.9%	19.3%
Results after filtering	50.3%	32.0%	23.9%	19.3%
Translation between IP and host name failed.	13.1%	7.9%	5.5%	7.4%
Translation between IP and host name succeeded.	37.2%	24.1%	18.4%	11.9%

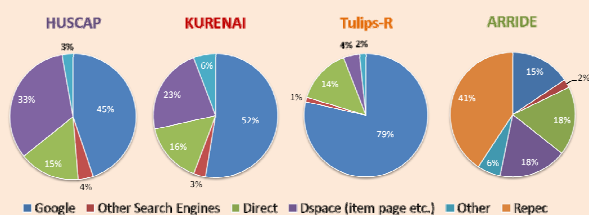
### Results

#### User Domains



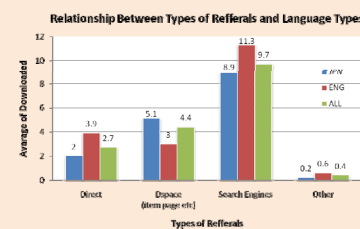
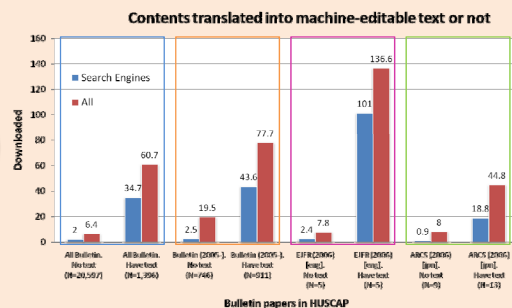
Accessed from not only researchers in education-related agencies, but also various members of people **the public and industries**.

#### Types of Referrals



Most users found repository contents **via search engines**, especially **Google**.

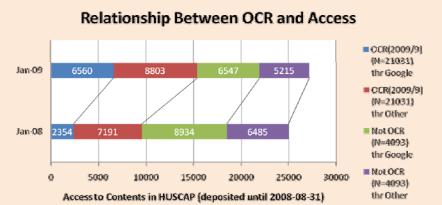
#### Content Attributes



**English papers**, in particular, are more likely to be accessed via search engines than are Japanese papers.

#### Make your papers accessible via machineries

Contents translated into **machine-editable text** were downloaded many times more frequently than those that have not been translated.



#### Relationship Between OCR and Born Digital

	Downloaded	ALL	Search Engines
JPN/OCR (N = 5,486)	Avg. 31.6	21.2	9.0
JPN/BornDigital (N = 434)	Avg. 65.2	44.8	14.0
ENG/OCR (N = 2,123)	Avg. 51.6	35.6	20.0
ENG/BornDigital (N = 236)	Avg. 6.7	4.6	0.0

\* Bulletin papers in KURENAI (Except RIMS Kokyuroku)

Contents that have a hidden text layer inserted by the **OCR process** behind the image also have increased access via search engines.