Recent Trends of Social Science Data in Japan

2020-3-19

Yasuyuki Minamiyama, Chief Data Librarian
Research Center for Open Science and Data Platform (RCOS),
National Institute of Informatics (NII)
About me

• My Job
  • As a data librarian
    • Sept. 2019 - Present
      National Institute of Informatics, Research Center for Open Science and Data Platform
    • Oct. 2018 – Aug. 2019
      The Tokyo Foundation for Policy Research Policy Data Lab
  • As an academic librarian
    • Sept. 2005 - Sept. 2018
      National Institute of Polar Research Library
  • Extra
    • July 2007 - Mar. 2008
      Participated 49th Japanese Antarctic Research Expedition (JARE) Summer Party

• Expertise
  • Metadata Management
  • Research Data Management
  http://researchmap.jp/minamiyama_yasuyuki
Library-outside experience
Outline

1. Actual situation of research data sharing in Japan

2. Policy developments in Japan related to open science

3. How to obtain Japanese social science data

4. Japanese activities related to the ensuring of research data management human resources
### 1. Actual situation of research data sharing in Japan

#### Overview: Types of available data related to social sciences in Japan

<table>
<thead>
<tr>
<th>Types of available data</th>
<th>Data archive</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Public statistical data</strong></td>
<td>• e-Gov</td>
</tr>
<tr>
<td></td>
<td>• e-Stat</td>
</tr>
<tr>
<td></td>
<td>• data.go.jp</td>
</tr>
<tr>
<td></td>
<td>• Statistics Bureau of Japan</td>
</tr>
<tr>
<td>(1) Population census</td>
<td></td>
</tr>
<tr>
<td>(2) Trade statistics</td>
<td></td>
</tr>
<tr>
<td>(3) Budgets and financial data</td>
<td></td>
</tr>
<tr>
<td><strong>Non-public statistical data</strong></td>
<td>• SSJDA</td>
</tr>
<tr>
<td>**Examples:</td>
<td>• Keio University Panel Data Center</td>
</tr>
<tr>
<td>(1) Survey of Japanese People's Life Values</td>
<td>• Rikkyo University Data Archive (RUDA)</td>
</tr>
<tr>
<td>(2) Survey on State of New Business Start-ups</td>
<td>• Institute of Economic Research, Hitotsubashi</td>
</tr>
<tr>
<td></td>
<td>University</td>
</tr>
<tr>
<td>(3) Some public opinion polls</td>
<td>• JGSS etc.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
So, how have researchers “actually” been sharing the data?
Insights 1: actual situation of research data sharing (Sociology)

• “There is a lot of qualitative data in each researcher’s context, so the culture of data sharing is still unfamiliar”
  • Background information of data has not been enough verbalized
  • Importance of “human networks” has been emphasized

• “There is no mechanism for sharing data. It is the only way that researchers get research data information from conference presentations/published papers/books”

• To obtain other researcher’s data:
  ✓ Become their graduate student
  ✓ Propose the joint research after accumulating some achievements
Insights 2: actual situation of research data sharing (Political science) 1/2

• “There is a person who is responsible for storing master data in the research group. Our data will be stored in SSJDA after the end of the study, but some data has not been deposited yet for the following reasons:“
  • No documentation time (it is hard to gather past dissipated data information)
  • Do not understand right processing for newspaper-originated data

• “Data sharing is often taking place between researchers, but the response differs depending on the situation”
  • From a graduate student: contact their supervisor and confirm how the data will be used
  • From a researcher: request an acknowledgment
Insights 2: actual situation of research data sharing (Political science) 2/2

• “Secondary analysis is common for my discipline, but it may be faster to make appropriate data myself”
  • Currently, aggregate data is not be handled by SSJDA
  • There is another problem with whether students can make accurate data…

• To obtain other researcher’s data:
  • Contact the research group following their style
  • Look for appropriate data archive
Insights 3: actual situation of research data sharing (Economics)

• “I may use public data in my paper, but I want to use the latest data in conference materials”
  • not yet published on the website such as e-Stat
  • The latest tax statistics can be picked up from the National Tax Agency's Statistical Yearbook (only available in PDF format)

• “In some cases, I will contact directly with the person in charge”
  • They often have accurate and easy-to-use data

• To obtain appropriate data:
  • Understand the purpose of use
  • Know how to contact the person in charge
Summary: Research data environment surrounding the social sciences in Japan

✔ Data sharing is mainly between research groups. Use of data archives is still limited

✔ Some data types are not handled in the data archive (such as aggregated data). There may not be just easy-to-use data

✔ You may need to create data by yourself in principle, but it may be effective to contact stakeholders through appropriate routes
2. Policy developments in Japan related to open science

• March 2015 Cabinet Office, "Promoting Open Science in Japan"

• January 2016 "The 5th Science and Technology Basic Plan"

• February 2016 Council for Science and Technology, "Promoting Open Access to Academic Information"

• July 2016 Science Council of Japan (SCJ), "Recommendations Concerning an Approach to Open Science that Will Contributes to Open Innovation"

• June 2018 Headquarters for Japan's Economic Revitalization, "Growth Strategy 2018 - Reform towards Society 5.0 and Data-driven Society"

• June 2018 Cabinet Office, "Guideline for Establishing Data Policy at National Research and Development Agencies"
National guidelines regarding research misconduct

‘Guidelines for responding to misconduct in research’ (2014.8.26)
- “Preserving and disclosing research data in research institutions for a set period”

- JST, AMED - requires their Data Management Plan in 2017
- JSPS - under consideration

- Most of university / research institution adopted / corresponded this policy
  “Except for under special circumstances, the preservation period to be stipulated in the preservation plan shall be at least 10 years after the presentation of the said research papers”

Figure:
http://www.mext.go.jp/component/english/__icsFiles/afieldfile/2016/02/23/1367533_005.pdf
Development of research data infrastructure and repository by NII and universities/institutions

https://www8.cao.go.jp/cstp/stmain.html
Program to Promote the Construction of Data Infrastructure for Humanities and Social Sciences

The image of the federate data catalog

platform for discovering academic resources

Data users

Data centers and universities

Researchers may try to register their data directly

Data repositories / Institutional repositories
Data held by each data centers / universities

• SSJDA:
  • Provides about 1500 individual survey data

• Hitotsubashi University:
  • Provides mainly processed data of government statistics created in each research project

• Keio University Panel Data Center:
  • Provides some household panel surveys

• JGSS:
  • Provides “Japan General Social Survey” data (regional data)
Interface image (under construction)

WEKO3

Top  WorkFlow  Communities

Type and press enter to search

Index Tree

► Sample Data Repository
  ► New Index
  ► Institute for Quantitative Biosciences
  ► Panel Data Research Center at Keio University
  ► Institute of Economic Research, Hitotsubashi University
  ► JGSS Research Center at Osaka University of Commerce
  ► SSJ Data Archive

► Contents Type
  ► Faculty of Humanities and Social Sciences
  ► Faculty of Business Sciences
  ► Faculty of Pure and Applied Sciences
  ► Faculty of Engineering, Information and Systems
  ► Faculty of Life and Environmental Sciences
  ► Faculty of Human Sciences
  ► Faculty of Health and Sport Sciences
  ► Faculty of Art and Design
  ► Faculty of Medicine
  ► Faculty of Library, Information and Media Science

Index List

Sample Data Repository <156950566479>

usaha <1569505650817>

JGSS Research Center at Osaka University of Commerce <1569979259543>

Institute of Economic Research, Hitotsubashi University <1569979260320>

Panel Data Research Center at Keio University <1569979675486>

Institute for Quantitative Biosciences <157016199893>

New Index <1570232891360>

Item Lists

Sample Data Repository

Export
3. How to obtain Japanese social science data:

Pattern of research data publishing

Independent
Data/dataset and metadata are published at data repository

Supplemental
Data/dataset are published accompanying the paper

Data paper
Data paper is published accompanying the data/dataset

# data presented within the paper
# data presented at data repository

Ref: https://www.nii.ac.jp/openforum/upload/20190530PM_Contents2_03_kato.pdf
Before introduction… Caution!!

English pages on the Japanese website often have less information and update slowly.
General catalog (for data librarians)

# Contains Japanese public and private statistics in various fields (such as Japan's land, population, economy, society, and culture)

# Also available for looking up statistics by related keywords

# Unfortunately, you can download only PDF file
-> You can use e-Stat

English translated by google:
http://www.stat.go.jp/data/nenkan/index1.html
Search by category (for data librarians)

**Economics, Society, Education**

This page introduces tools for finding information on economics, society, and education.

**How to check**

You can search by keyword for search guides that introduce related information sources.

<table>
<thead>
<tr>
<th>Economic and social fundamentals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examine statistics</td>
</tr>
<tr>
<td>Examine the roster</td>
</tr>
<tr>
<td>Investigate Land</td>
</tr>
<tr>
<td>Phone book</td>
</tr>
<tr>
<td>time table</td>
</tr>
<tr>
<td>poll</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Economy and industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investigate economy, finance and finance</td>
</tr>
<tr>
<td>Find a company</td>
</tr>
<tr>
<td>Industry information Guide</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Society, labor, education, sports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investigate society</td>
</tr>
<tr>
<td>Examine labor</td>
</tr>
<tr>
<td>Check education</td>
</tr>
<tr>
<td>Investigate sports</td>
</tr>
</tbody>
</table>

**Visit and check**

1. Science and Technology Information Room
2. In-house contacts database for economic, social, and educational fields

**Related organizations/sites**

1. Japan External Trade Organization (JETRO)
2. National Institute for Educational Policy Research, Education and Research Information Center, Education Library
3. Comprehensive window for government statistics for Japan
4. EDINET

**Examination method**

**By field**

1. Information related to reconstruction support for the Great East Japan Earthquake
2. Science and technology
3. Economics, Society, Education
4. Politics, law, administration
5. Humanitarian / General
6. Asian countries

English translated by google: https://rnavi.ndl.go.jp/business/index.php
Academic data archive: example 1 (SSJDA)

# You can find both data and secondary analysis paper

https://ssjda.iss.u-tokyo.ac.jp/Direct/?lang=eng
Academic data archive: example 2 (Dataverse)

https://dataverse.harvard.edu/dataverse/harvard
How to look for the paper’s evidence data

Disentangling the relationship between sociotropic and egotropic trade attitudes: A survey experiment in Japan

Abstract
Public support for protection is typically attributed to economic self-interest. Beyond pocketbook anxieties, a competing approach, however, contends that sociotropic attitudes dictate foreign policy preferences. Researchers, however, have faced difficulty in disentangling sociotropic attitudes from pocketbook concerns in observational studies. This article addresses this problem by utilizing a priming experiment to examine the relationship between socio and egotropic attitudes. In line with the predictions of the

Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/S1468109919000124 and https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/KZK409

https://doi.org/10.1017/S1468109919000124
Data paper/Data journal

An FAQ dataset for E-learning system used on a Japanese University

Yasunobu Sumikawa and Masaki Fujisaki

Data in Brief
Volume 25, August 2019, 104001

Data Article

In this data article, we present an FAQ dataset written in Japanese and its translation to English in order to train chatbot models for e-learning systems. We first collected raw Q&A data reported as the difficulties from April 2015 to July 2018 by users of the e-learning system introduced at Tokyo Metropolitan University. We then divided them into 11 categories according to features provided by the e-learning system. Finally, we integrated questions with the same answers in order to create the FAQ form. The dataset contains 427 questions and 79 answers that were examined by experts with experience in using the e-learning system for more than three years. Using this dataset, we performed statistical analyses to evaluate the qualities of the FAQ dataset. The proposed applications of the dataset include not only academic research but also activities; for example, translating from Japanese to

https://doi.org/10.1016/j.dib.2019.10400

1

https://doi.org/10.5281/zenodo.2783642

Japanese FAQ dataset for e-learning system

Yasunobu Sumikawa, Masaaki Fujisaki, Masashi Hatakeyama, and Makoto Nakai

This dataset includes FAQ data and their categories to train a chatbot specialized for an e-learning system used in Tokyo Metropolitan University. We report accuracies of the chatbot in the following paper:


This dataset is based on real Q&A data about how to use the e-learning system asked by students and teachers who use it in practical classes. The duration we collected the Q&A data is from April 2015 to July 2018. We attach an English version dataset translated from the Japanese dataset to ease understanding what contents our dataset has. Note here that we did not perform any evaluations on the English version dataset, there are no results how accurate chatbot responds to questions.

File contents:
- FAQ data (tex):
  - data2.csv: Categories of answers.
  - data3.csv: Titles of answers.
  - answers.csv: Lists for answers and texts of answers.
  - categories.csv: Names of categories for answers.
  - questions.csv: Titles of questions and their corresponding answer IDs.
  - answers_english.csv: Lists for answers and texts of answers written in English.
  - categories_english.csv: Names of categories for answers and their corresponding English names.
  - questions_english.csv: Titles of questions and their corresponding answer IDs written in English.

https://doi.org/10.5281/zenodo.2783642
Other possibilities

Clarivate Analytics. Data Citation Index
http://wokinfo.com/products_tools/multidisciplinary/dci/

Data Cite. DataCite
https://search.datacite.org/

Elsevier. Discover Mendeley Data
https://data.mendeley.com/

IRDB.
https://irdb.nii.ac.jp/en
Ref: useful databases/links

http://worldjpn.grips.ac.jp/indexPC-ENG.html

English translated by Google:
https://www.nistep.go.jp/research/scisip/data-and-information-infrastructure/datalink_country_detail
If still not found...

- **Ask a librarian!**
  - (Probably) Japanese university library staff will respond flexibly
  - If you can request via the Japanese university library, you may be processed quickly

- When you contact other than library, **please use Japanese**
  - Some governmental departments may not even be read your message…
4. Japanese activities related to the ensuring of research data management human resources

**Issue: Lack of Research Data Management Staff in Japan**

<table>
<thead>
<tr>
<th>Status of Research Data Management Planning in Academic Institutions in Japan (N = 151)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>We mandate researchers to submit data management plan</td>
<td>0.0%</td>
</tr>
<tr>
<td>We encourage researchers to write data management plan</td>
<td>11.9%</td>
</tr>
<tr>
<td>We don’t know DMP</td>
<td>41.1%</td>
</tr>
<tr>
<td>Other</td>
<td>2.6%</td>
</tr>
<tr>
<td>No answer</td>
<td>4.0%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
</tr>
</tbody>
</table>

- Librarians are good at handling academic papers
  - (e.g. such as managing institutional repositories)
- However, few librarians have research data management skills.

**Development of human resources to support RDM is necessary to promote research data activity**

https://doi.org/10.1241/johokanri.60.119
Development of RDM training courses

Our efforts to develop RDM training courses with JPCOAR

- NII and JPCOAR have developed RDM training material named “RDM training tool” since 2015.
- JPCOAR’s Main activities about RDM
  - Development of training materials
  - Rescue of the databases
  - Aggregate and analyze RDM case study

First, We started developing RDM training materials.
Japan Consortium for Open Access Repository (JPCOAR)

JPCOAR is the consortium of academic institutions operating institutional repositories.

- Japan Associations of Universities
- Japanese University Libraries
- Board Committee

JPCOARE Partners

- Chair
- Auditors
- Steering Committees
- Training working group
- PR working group
- Operation working group
- Task forces
- Administration Office

Universities

- Launched in July 2016
- 629 member institutions as of January 2020
- https://jpcoar.repo.nii.ac.jp/

JPCOAR works with NII in the operation JAIRO Cloud and activities wanted to open science.
Short History of Our Activities

Institutional Repositories Promotion Committee (pre-JPCOAR)

2015

Developing “RDM Training Tool”

2016

2017.6 Release ”RDM Training Tool”

6,436 DL

2017

2017.11-2018.1 Launch Online Courses “Research Data Management in Open Science Era” in JMOOC/gacco

The number of participants: 2,305
(Completion Rate: 25%)

2018

2018.8 Release of Training Materials “Plan and Practice of Research Data Management”

3,495 DL

2018.8-10

Experimental project on online courses

2019

Start Trial for Online Courses on NII’s LMS

※Source: https://www.slideshare.net/YuiNishizono/20190528joss2019

DL=Download Number (as of Nov, 2019)
Slide-based Training Materials on RDM basics

- Training materials that provide a comprehensive understanding of research data (Total of 7 chapters)
- Licensed Creative Commons 4.0 Attribution (CC-BY)
- http://id.nii.ac.jp/1458/00000023/

We provide training materials for people with no experience in RDM
Lifecycle of Research Data and RDM Training Tool

1. Chapter 1: Introduction
2. Chapter 2: Data Management Plan
3. Chapter 3: Preservation and Sharing
4. Chapter 4: File Management, Documentation and Metadata Generation
5. Chapter 5: Laws and Ethical Issues
6. Chapter 6: Policies
7. Chapter 7: Designing Research Data Management Service

The UK Data Archive (http://www.data-archive.ac.uk/create-manage/life-cycle)
RDM Support - basic training course for information specialists (http://dx.doi.org/10.6084/m9.figshare.1285313)
Launch MOOCs with RDM Training Tools

- from November 15, 2017 to January 15, 2018
- Available on MOOCs platform "gacco"
- RDM training tool reorganized to 4 weeks of learning
- Video learning, quiz, and Q & A through bulletin boards

https://www.nii.ac.jp/service/jmooc/rdm

2,305 enrollments

Our online training course enables people are interested in RDM
2nd Training Materials: “Plan and Practice of RDM Services” (2017)

RDM Supporter Specific Training Materials

- Practical training materials for librarians supporting RDM services
- Include contents on how to design and operate RDM services
- The training materials are arranged according to general research process.

http://id.nii.ac.jp/1458/00000107/

We helped RDM supporter acquire practical knowledge and skills for RDM services
The UK Data Service, Research data lifecycle
http://www.data-archive.ac.uk/create-manage/life-cycle

Mariette van Selm, RDM Support - basic training course for information specialists
http://dx.doi.org/10.6084/m9.figshare.1285313
New Online Training Course for RDM Supporters

RDM online courses on NII education Platform

- NII launched Moodle as own Education Platform
- Mounted existing 2 training materials (RDM basics & Designing RDM services)
- The platform enable institutions to monitor RDM training of their staff

We intend to leverage training materials for institutional RDM training
3rd RDM Training Tool: for Early Career Researcher and Graduate Students

We are restructuring existing materials for early career researchers and graduate students

Looking at RDM training for researchers from the standpoint of providing a foundation for open science

1. Context-based training materials
   - TF members survey detailed their researcher’s actual environment and needs

2. Modularize existing training tools
   - Provide existing training tools in module (an easily customizable form) at each institution
   - Examine how training materials should be easily embedded in the context of each institution
Summary: our RDM training tool activities

<table>
<thead>
<tr>
<th>#</th>
<th>Target</th>
<th>Skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1(^{st}) RDM Training Tool</td>
<td>People interested in RDM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RDM Basics</td>
</tr>
<tr>
<td>2</td>
<td>Online Course: Research Data Management in the Open Science Era</td>
<td>People interested in RDM, RDM Basics</td>
</tr>
<tr>
<td>3</td>
<td>2(^{nd}) Training Materials “Plan and Practice of Research Data Management”</td>
<td>RDM supporters in academic institutions</td>
</tr>
<tr>
<td>4</td>
<td>Trial for Online Courses on NII’s LMS (on-going)</td>
<td>RDM supporters in academic institutions</td>
</tr>
<tr>
<td>5</td>
<td>3(^{rd}) RDM Training Tool for Early Career Researcher and Graduates (on-going)</td>
<td>Researcher and Graduate Students</td>
</tr>
</tbody>
</table>
Thank you!

Next, I will introduce the online questionnaire survey during Feb. 12 to Mar. 4, 2020…

Special Thanks:
* Yukio Maeda (University of Tokyo)
* Ui Ikeuchi (Bunkyo University)
* Makoto Asaoka (National Institute of Informatics)