RESEARCH PAPER

Investigation and Development of the Workflow to Clarify Conditions of Use for Research Data Publishing in Japan

Yasuyuki Minamiyama[1], Ui Ikeuchi[2], Kunihiko Ueshima[3], Nobuya Okayama[4] and Hideaki Takeda[5]

[1] Research Center for Open Science and Data Platforms, National Institute of Informatics, Tokyo, JP

[2] Department of English Language and Literature, Faculty of Language and Literature, Bunkyo University, Saitama, JP

[3] Japan Data Exchange Inc, Tokyo, JP

[4] Smart Infrastructure Consulting Department, Hitachi Consulting Co., Ltd, Tokyo, JP

[5] Principles of Informatics Research Division, National Institute of Informatics, Tokyo, JP

Corresponding author: Yasuyuki Minamiyama (minamiyama@nii.ac.jp)

With the recent Open Science movement and the rise of data-intensive science, many efforts are in progress to publish research data on the web. To reuse published research data in different fields, they must be made more generalized, interoperable, and machine-readable. Among the various issues related to data publishing, the conditions of use are directly related to their reuse potential. We show herein the types of external constraints and conditions of use in research data publishing in a Japanese context through the analysis of the interview and questionnaire for practitioners. Although the conditions of research data use have been discussed only in terms of their legal constraints, we organize the inclusion of the non-legal constraints and data holders' actual requirements. Furthermore, we develop practical guideline for examining effective data publishing flow with licensing scenarios. This effort can be positioned to develop an infrastructure for data-intensive science, which will contribute to the realization of Open Science.

Keywords: Conditions of use; licensing; research data publishing; data curation; open science; data-intensive science

1. Introduction
1.1 Background
With the recent Open Science movement and the rise of data-intensive science (Kelling, S. et al. 2009), many efforts are in progress to make research results available on the web (Kowalczyk and Shanker, 2011). Numerous research papers have been digitized and published with Open Access options. A similar trend in research data underlying the papers has recently been observed. Moreover, various data within a specific domain have been used across domains, and a key element for driving both activities is research data publishing (Klump et al. 2006; Lawrence et al. 2011).

For efficient data publishing, data and metadata (Beagrie, 2006; Ruggles S., 2018) must be curated based on the defined standards and registered in a dedicated data repository (Assante, et al. 2016; Marcial and Hemminger, 2010). One of the most well-known community norm for sharing and publishing data is the FAIR Data Principles (Wilkinson, Dumontier, & Aalbersberg, 2016), specifying that data should be findable, accessible, interoperable, and reusable; they are widely supported across disciplines.

The terms and conditions of use must be classified when published data are shared in different fields (OECD, 2015). If the conditions of use for the data are ambiguous or not standardized, it causes an additional cost to interpret or reuse. In other words, in case of conditions of use described in an ambiguous/non-standard expression, there are few examples of how to interpret. The data may be interpreted differently from the data holder intended. Clear and unambiguous condition of use is better, but a bespoke condition

of use often impose additional costs on managing the data. As a result, it introduces significant delays to the process of publishing data. For the data generated by public research funds, we can reach a consensus through data publishing guidelines or policies in each country (ANDS, 2018; CODATA, 2019; RECODE Project Consortium, 2014). However, there are many research projects using non-public funded data that has more diverse nature and standardization efforts based on the actual situation for conditions of use for research data publishing has not been substantially explored.

1.2 Approach

We started clarifying the requirements based on the use-case of research data and researchers' needs to propose appropriate options for conditions of use. First, we conducted interviews with data repository practitioners to clarify the issues and actual conditions. Next, we conducted a web questionnaire survey for researchers and research supporters to obtain quantitative data for organizing and categorizing the possible conditions of use. Based on these surveys, we identified the constraints that come from types of research data, actual options granted, data holder's intentions, legal restriction, or nature of the research community.

Then we developed guideline based on the analysis of both surveys. This guideline provides a data publishing workflow for specifying and describing the categorized conditions of use. This guideline will enable researchers to select the standard conditions of use and treat cross-disciplinary data from different fields. This activity can be positioned to develop an infrastructure for data-intensive science, which will consequently contribute to Open Science.

This study was conducted as part of the Research Data License Subcommittee under the Research Data Utilization Forum (RDUF, https://japanlinkcenter.org/rduf/) activities. The RDUF is a voluntary inter-institutionary project for Open Science established to provide a place for stakeholders involved in utilizing research data to share information across individual organizations and fields. Each subcommittee sets the activity themes and aims to make policy proposals or guidelines.

2. Definition

In this Section, we define "Researcher," "Data holder," "Data user," "Conditions of use," and "Data publishing" in this paper.

"Researcher" is a person who conducts research using data. Both a data holder and a data user are included.

"Data holder" is either a person or an institution that have datasets to publish.

"Data user" is also either a person or an institution that use datasets.

"Condition of use" generally means use permission/prohibition, obligations, and constraints based on the types of research data. The license by the data holder is also included in "condition of use" in this paper (e.g., copyright sign, Creative Commons license, Open Data Commons license). Other contractual clauses (e.g., intellectual property rights in data, ownership, disclaimers, warranties, and liability for defects and damages) are also included, but not considered as the main issues. For convenience, the term "license" is used in the interview and questionnaire surveys that will be described below.

"Data publishing" means to make research data open to the public and is distinguished from "data sharing." "Data publishing" includes cases where it is published only to an unspecified large number of users who meet certain conditions. The assumptions of the conditions of use significantly differ between the case where data are shared to only a specific target and where data are provided to an unspecified large number of users.

3. Literature Review

Among the legal risks of research data, unclear copyrights and licenses are often a high barrier among the list of factors that hinder data reuse (Mayernik, M.S. et al. 2020). Van Panhuis mentioned lack of trust (that users will use the data properly), overly restrictive policies and unclear guidelines on data sharing, and confusion over the ownership of data (Van Panhuis, Williem G., et al. 2014). The Digital Curation Centre mentioned early on the importance of promoting licensing as a way of maximizing the economic and social impact of data publishing (Ball, A., 2014). A large international survey conducted by SpringerNature (Stuart et al. 2018) found that unsure about copyright and licensing (37%) are the second most crucial factor as the obstacles to publishing research data after organizing the data in a useful manner (46%). Kindling (Kindling, M. et al. 2017) also reported that the most widely used type of condition of use registered in re3data.org is "Other" at 57.2% and "Copyright" at 38.6%, with setting its own conditions of use or copyright notices being the most used situation. The use of standardized tools, such as Creative Commons (CC) licenses, is limited to 21.8% at most. In some cases, CC licenses are granted for "non-copyrighted" works. The conditions of use for research data required by data holders are very diverse; hence, the interpretive costs for reuse are significant.

Minamiyama et al: Investigation and Development of the Workflow to Clarify Conditions
of Use for Research Data Publishing in Japan

Art. 53, page 3 of 20

Initiatives for setting conditions of use have already been set up. Grabus and Greenberg reported 20 initiatives for dealing with ethical or legal issues for data sharing or publishing (Grabus, S. and Greenberg, J., 2019). Accordingly, the RDA/CODATA Legal Interoperability Interest Group (IG) published the article "Legal interoperability of research data: principles and implementation guidelines" (RDA-CODATA Legal Interoperability Interest Group, 2016). These guidelines are primarily for data produced in or funded by the public sector and focus on legal interoperability to address misunderstanding and lack of knowledge and guidance on the legal issues generally related to research data. Meanwhile, in the local context, the requirements for legal decisions on privacy and other matters slightly differ from country to country, and some solutions have not been introduced in Japan (Ministry of Economy, Trade and Industry, 2018).

In some research disciplines, data that are not funded by public funds are widely used in research. In the case of non-public-funded data, the data are held by an individual or a company, and the conditions of use are determined by the data holder, except in cases provided by law. A number of existing licensing tools can be applied to these data. **Table 1** shows a comparison of the CC and Open Data Commons (ODC) licenses and the "Government of Japan Standard Terms of Use" used for Japanese government websites.

These licenses are designed to provide flexible conditions of use, considering the copyright law. However, the presented laws that may apply to research data are more complex and do not fully reflect the actual conditions. In some cases, CC licenses are granted to data that are not subject to protection under the Japanese copyright law, such as numerical data, which may cause confusion about reuse. While there is no doubt that CC licenses are still a useful solution in many cases of data publishing, some challenges exist for handling non-copyrighted data. Some proposals have recently been pushed forward (e.g., UK Scholarly communication license (Baldwin, J. and Pinfield, S., 2018) and Microsoft Open Use of Data Agreement (https://github.com/microsoft/Open-Use-of-Data-Agreement)). However, the extent to which these proposals are effective in the Japanese context is still under consideration. In any case, various requests from data holders must be considered to realize a wider range of data publishing.

4. Survey and Analysis

This study aims to develop an infrastructure for handling research data from different fields across the borders to clarify and standardize the conditions of use. For this purpose, we set the following research questions (RQs) herein. In Section 5, we discuss the licensing based on the results of the RQs.

> RQ 1: What are the limitations that arise in using the research data?
> RQ 2: What do the data holders desire or request when they publish research data?
> RQ 3: What support can be effective in promoting reuse in aspects of the data holders' desire or data users' request?

We address these issues by conducting a survey to clarify the use-cases of the research data and the researchers' needs to identify the requirements.

4.1 Interview Survey

First, the preliminary interviews were conducted with data repository practitioners to organize the conditions of use that could be granted according to the types of data. This survey aims to primarily identify the external constraints that arise in the research data through interviews with the data repository operators. It also includes the purpose of obtaining clues for the subsequent questionnaire survey. The interview study was limited to Japan, since we think that external constraints must be judged in the local context.

We conducted a semi-structured interview with the following topic guide:

Main question:
· Sharing and publishing of research data
  ○ Outline and characteristics of the research data to be handled
  ○ Current status of data publishing in your own research, your institution, and your research community
  ○ Difficulties in sharing and publishing research data
  ○ What you want to ask data users to do/prohibit to your published research data
· Regarding granting licenses to research data
  ○ Type of license tools used, provisions of the licenses granted, and information/guidelines referenced at the time of granting the licenses
  ○ Difficulties in granting licenses

Table 1: License comparison chart.

License	Permissions			Requirements			Prohibitions
	Reproduction	Distribution	Derivative Works	Notice	Attribution	Share Alike	Noncommercial
Creative Commons Licenses (https://creativecommons.org/licenses/)							
CC0	X	X	X				
CC-BY	X	X	X	X	X		
CC-BY-SA	X	X	X	X	X	X	
CC-BY-NC	X	X	X	X	X		X
CC-BY-ND	X	X		X	X		
CC-BY-NC-SA	X	X	X	X	X	X	X
CC-BY-NC-ND	X	X		X	X		X
Open Data Commons (https://opendatacommons.org/)							
ODC-PDDL	X	X	X				
ODC-BY	X	X	X	X	X		
ODC-ODbL	X	X	X	X	X	X	
Government of Japan Standard Terms of Use (https://www.kantei.go.jp/jp/singi/it2/densi/kettei/gl2_betten_1.pdf)							
Government of Japan Standard Terms of Use	X	X	X	X	X		

Minamiyama et al: Investigation and Development of the Workflow to Clarify Conditions
of Use for Research Data Publishing in Japan

Art. 53, page 5 of 20

· Licensing of research data and promotion of legal interoperability
  ○ Need for licensing and legal interoperability of research data
  ○ Personal views on existing licenses and guidelines and existing discussions
  ○ The extent of protection of the research data and basis for the request for protection
  ○ Expectations of organizations that support licensing, rights management, and data publishing
· Other issues
  ○ How to deal with license violations

The survey included five experts, including data repository managers and researchers from space science, environmental sciences (interdisciplinary fields), social sciences, materials science, and humanities (digital archives). The criterion for selecting the fields is that multiple conditions of use must be attached to the research data. Also, the research request document stated that the responses would be incorporated into the questionnaire and that it would be made available as a report. This point was also confirmed verbally before the interviews. The personal information of the interviewees was not included in this paper. The survey period was from December 12th, 2017 to February 1st, 2018 and was conducted for approximately one hour each. **Table 2** shows the summary of the results.

With regard to data policies, the policies for research data publishing in research data repositories are generally clear, and automated processing is in progress in some research disciplines. The repository dealing with interdisciplinary data seems to have challenges in setting acceptance standards and data and metadata quality.

The types of data holders include the researchers who acquired the data, the institution they work for, the research funder, or a third party data supplier. In various cases, it is unclear who can claim rights because of the passage of time or the circumstances of the funding agency. The demand for rights protection varies by research discipline. In a research discipline that deals with both constrained and unconstrained data, there is an opinion that the less constrained datasets (e.g., no fixed term) tend to be more commonly used.

The interviewees provided several reasons why publishing data might fail. From an IP perspective, these include privacy (e.g., portrait rights), military/security perspective, and intent of the depositor.

4.2 Questionnaire Survey
We conducted a web questionnaire survey following the interview survey. This survey aims to clarify the actual situation and perceptions of using the conditions of use through a questionnaire survey for data holders. At the same time, we also aim to obtain a more specific knowledge of giving incentives to data holders, which have already been pointed out. Ten questions are provided, and no mandatory items are included. Some of the questions are expected to be difficult for some respondents to answer.

List of questions:

1. Which of the following terms best describes your research field?
2. Have you ever obtained or published any data, including the cases in which user registration and fees are required?
3. Are you familiar with the following license tools?
4. Have you ever used any of the following license tools to publish your data?
5. If you would like to publish your data, would you like to require the following to your users (those who use that data to publish results)?
6. If the license is complied with, would you be willing to publish the data?
7. If you are using public data for your part of the research, please choose the method of presentation that you think is appropriate.
8. Do you have any requests or concerns about using your published data for commercial activities, patents, press, literature, art, etc.?
9. Please select the initiatives that you believe are desirable to the use and publishing of data.
10. Free description (any problems or requests regarding the use or publishing of data).

The survey period was set from February 13, 2018 to March 20, 2018. A web questionnaire was distributed via some mailing lists and websites using Questant's questionnaire system. The survey not only mainly targeted researchers, but also data manager, and librarians. The final number of responses is 413, of which 409 are valid responses. It should be noted that two limitations of this survey are as follows: (1) This survey was not a random selection. (2) The respondents' research fields were biased from Social Science (17.4%) to Astronomy (0.2%). At the beginning of the questionnaire survey, we stated that (1) we plan to publish the aggregated results in oral presentation and published form, (2) no questions are personally identifiable,

Table 2: Summary of the interview survey results.

Date	12,12,2017	12,18,2017	12,20,2017	01,30,2018	02,01,2018
Fields	Space Physics	Environmental Sciences	Social Sciences	Materials Science	Humanities
Characteristics	Mostly numerical data	Image data, numerical data, etc..	Survey data	Measurement data, calculation data, Material Informatics data, software	Image data, bibliographic data
Data holder (Representative)	The person(s) who is/are acquired the data	The funder(s)	The data provider	Institution of the data holder (with exception)	uncertain
Request for users	None	# Request for user name and purpose of use for searching metadata # Primary data are negotiated on an individual basis	# Only available by the researchers who belongs in the social sciences discipline # Submission of a research proposal is mandatory # Submit usage reports every year/Inclusion in the acknowledgments for using the data	Provide a provenance information of the data as well as literature citations	# The metadata should be CC0 # The conditions of use of the data should be clearly stated by data holder
Penal regulations	There are no publishing constraints from a scientific point of view	Under consideration	If the data is passed on to a third party without permission, the use of the data may be suspended	None	Considering the use of rightsstatements in case of publishing data
Rights protection	No rights protection is provided for the data to be published	There are two levels of access restrictions on data in the repository, depending on the contents	One-year and indefinite licenses are available according to the wishes of the data holder	Data marked as private will have restricted access	Non-public data will be considered for protection on an individual basis by contract. Among the public data, those with copyright properties will be subject to government of Japan standard terms of use and CC licenses
Issues	# There are few explanation for national security related data (especially the description of disclosure period)	# There are no institution to consult about research data rights in Japan	# In addition to an organization that supports data management, data management personnel and personnel who can handle the technical aspects of metadata are needed # The criteria for determining sensitive data change over time, so it cannot provide past data as it is	# A point of contact is needed to receive inquiries about published data	# Licensing standards for publishing non-copyrighted or obscure data # How to develop a culture of data protection, who will bear the cost, and how to spread it The future discussion points are whether or not to do so
Aspiration related to conditions of use	Development of data utilization laws	Enhancement of the university's intellectual property department function			

Minamiyama et al: Investigation and Development of the Workflow to Clarify Conditions
of Use for Research Data Publishing in Japan

Art. 53, page 7 of 20

and (3) if you do not want your free responses to be cited, you should state that. There are no personally identifiable aspects of the respondents in this paper. The aggregated results of the 409 valid responses are presented in the order of the questions presented before. The data from the survey are publicly available (Ikeuchi, U and Minamiyama, Y. 2020). The "n" in the chart indicates the number of respondents.

(1) Property of respondents

Table 3 shows the research fields of the respondents. Social sciences (17.4%), earth sciences (12.5%), and humanities (10.3%) were well represented among respondents, while mathematics and astronomy were not (both 0.2%). Other responses included 'library and information science', nursing, nutrition, and so on. Responses from library staff and related persons and private companies were also recorded. Fifty-eight (14.2%) respondents selected, "I am not currently engaged in any research activities."

Table 3: Research fields of respondents (n = 409).

Research Field	Number	Ratio
Social Sciences	71	17.4%
Earth Sciences	51	12.5%
Humanities	42	10.3%
Medicine	35	8.6%
Engineering	32	7.8%
Computer Science	20	4.9%
Biological Science	19	4.6%
Agricultural Science	18	4.4%
Psychology	16	3.9%
Physics	8	2.0%
Chemistry	2	0.5%
Mathematics	1	0.2%
Astronomy	1	0.2%
Other	35	8.6%
I am not currently engaged in any research activities	58	14.2%
Total	409	100.0%

(2) Experience in obtaining published data and publishing data by themselves

In this question, we asked for experience in obtaining published data and publishing data by themselves from the nine sources. The respondent's choices are "Obtain," "Publish," and "None" and set as follows: "Obtain" and "Publish" are multiple selections, and "None" cannot be selected when "Obtain" or "Publish" is selected. **Table 4** shows the aggregate results.

Highly selected answers as regards "where to obtain" are institutional repositories/data archives (62.3%), government repositories/data archives (48.4%), and personal/research lab websites and blogs (47.9%). Highly selected answers about "where to publish" are institutional repositories/data archives (25.7%), personal/research lab websites and blogs (23.5%), and academic SNS services (11.5%). Compared to the experience of obtaining data, the proportion of respondents with experience in publishing data is lower.

Table 5 presents the results of obtaining public data and having experience in releasing data. Respondents who selected "Yes" for one or more of the items in **Table 4** are tabulated as having "Yes" experience in obtaining and publishing. Consequently, 84.1% of the respondents had experience in obtaining data, and 46.5% of the respondents had experience in publishing data. One respondent did not respond at all.

(3) Awareness of existing licenses

We asked for the awareness of three licenses, which are well known in Japan to identify the extent to which existing licenses are recognized. To eliminate answers based on fuzzy memories, we also set a link to the license or a page explaining the license in this question form. **Figure 1** shows the aggregate results.

Table 4: Experience in obtaining published data and publishing data by themselves (n = 409).

Sources	Obtain	Publish	None	No Answer
Institutional repositories/data archives	62.3%	25.7%	29.1%	1.5%
Government repositories/data archives	48.4%	1.7%	46.0%	4.6%
Personal/research lab websites or blogs	47.9%	23.5%	41.8%	2.2%
Supplementary materials (in research paper)	36.7%	9.3%	54.3%	6.8%
Academic SNS services (e.g. Mendeley, ResearchGate)	32.0%	11.5%	58.2%	6.6%
Data repositories/archives in specific field	28.6%	8.3%	64.3%	5.4%
Code sharing services (e.g. GitHub)	24.4%	8.1%	69.2%	5.1%
Repositories/data archives by Commercial company	18.1%	1.5%	73.6%	7.1%
Other data publishing services (e.g. figshare, zenodo)	12.7%	3.7%	79.2%	6.8%

Table 5: Experience in obtaining published data and publishing data.

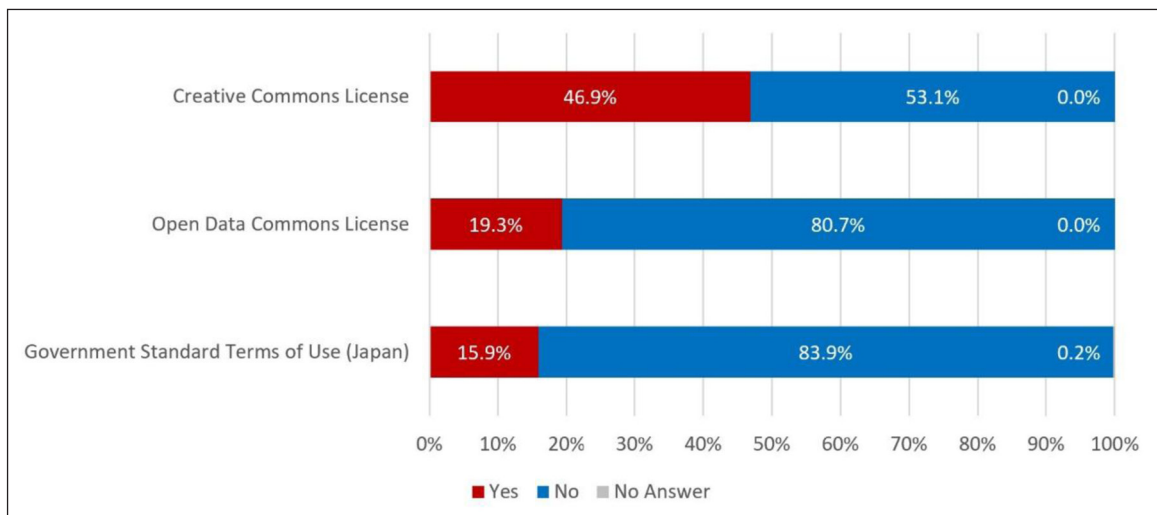|	Yes		No/No response		Total	
Obtain	344	84.1%	65	15.9%	409	100.0%
Publish	190	46.5%	219	53.5%	409	100.0%

Figure 1: Awareness of existing licenses (n = 409).

The highest recognition is for CC license, but less than half (46.9%) of the respondents are aware of it. ODC (19.3%) and Government Standard Terms of Use (15.9%) follow, and both are less than two in 10. The survey respondents would be expected to have some level of interest in licensing their research data, but awareness of existing licenses was not high.

(4) Usage of existing licenses

To ascertain the use of the licenses listed in the previous question (3), we asked for respondents who are aware of each license about their experience in using each one. **Figure 2** shows the aggregate results.

Fifty-nine respondents (30.7%) had used the CC license, which was the highest proportion, as was the case with (3). Only four (5.1%) and six (9.2%) respondents had experience using ODC and Government Standard Terms of Use.

(5) Desired condition of use when respondents publish their research data

We asked respondents to select their desired conditions of use from a list to quantify the extent of the requests they would make. The list was assembled from the results of the interview survey and the CC license

Minamiyama et al: Investigation and Development of the Workflow to Clarify Conditions of Use for Research Data Publishing in Japan

Art. 53, page 9 of 20

elements. **Figure 3** depicts the aggregate results, with the following order: the sum of "Yes" and "It depends on cases" is the highest.

The highest percentage of "Yes" and "It depends on cases" is for "Credit on the results" (93.4%). The "Yes" percentage was higher than the credit indication (80.0%) for the "Prohibition of use when improper use of data." The total of "It depends on cases" was 90.5%. The items for which the total of "Yes" and "It depends on cases" exceeded 80% were "Request to use the latest version" (84.1%), "Impose the same conditions when publishing results" (83.1%), and "Noncommercial" (83.1%).

On the contrary, 43.5% of the respondents selected "No" when they asked for "Nothing (freely available)." In other words, just over 40% of respondents wanted to set some kinds of conditions to release their data. In addition, 40.3% of the respondents answered "No" to the question of "Fee for use," indicating that a certain number of respondents do not want to be compensated for data publishing. Note that there is an error of 0.1% between the **Figure 3** and the main text due to rounding off numbers after the decimal point.

(6) License compliance and willingness to publish data

We asked whether they would be willing to publish their own data if the conditions listed in (5) were complied with. This question was asked to all respondents, including those whose data had already been exposed. **Figure 4** depicts the aggregate results.

Consequently, 64.1% of the respondents said that they were "Agree," and 24.4% said they were "Somewhat agree" (total: 88.5%), exceeding "Somewhat disagree" (4.6%) and "Disagree" (2.4%).
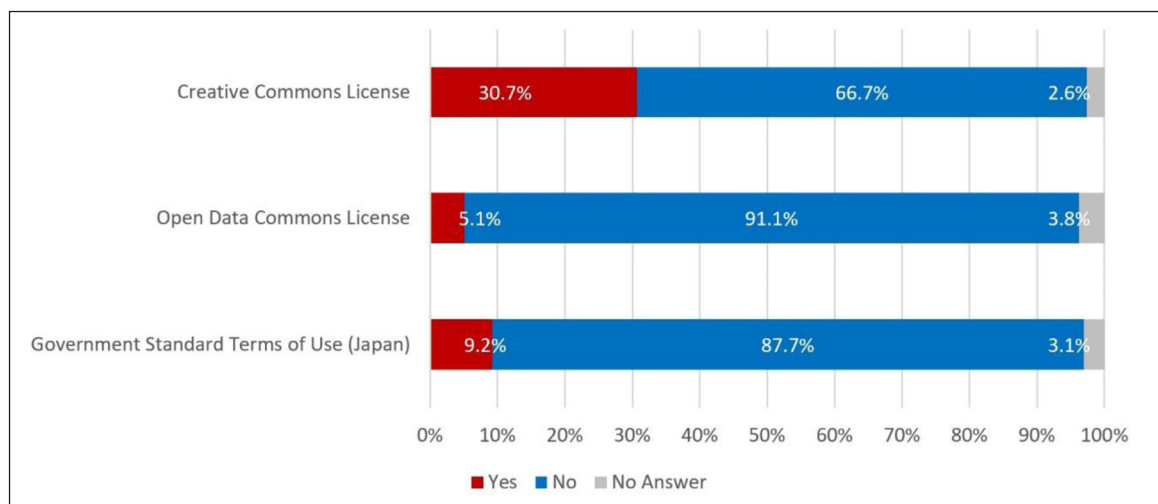

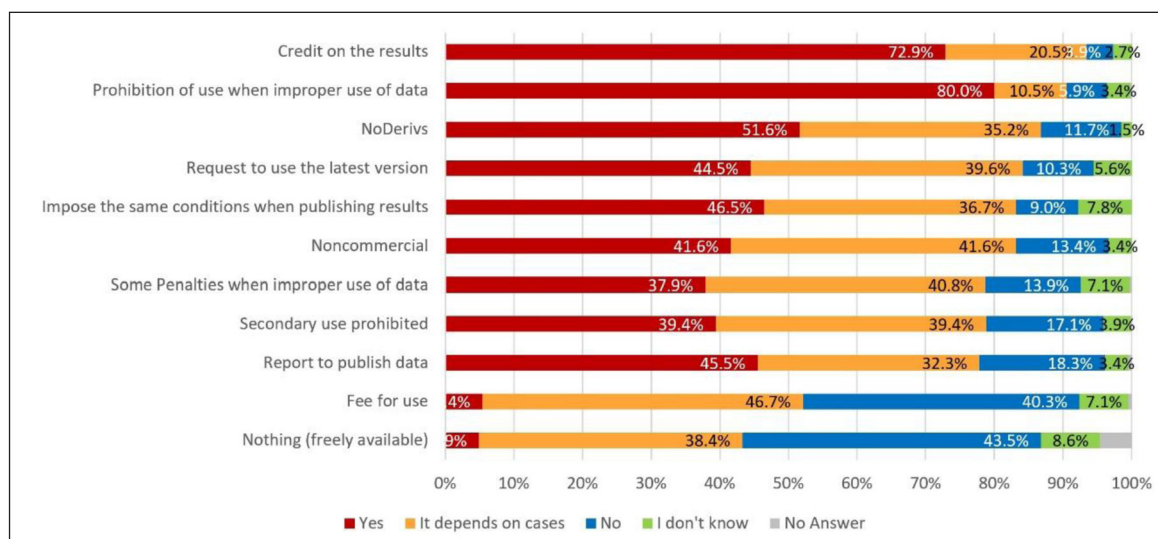**Figure 2:** Usage of existing licenses (n = 409).

Figure 3: The desired conditions of use when respondent's publish their research data (n = 409).

Figure 4: License compliance and willingness to publish data (n = 409).

(7) An appropriate method of displaying the use of published data

The respondents were asked about the display method they thought was appropriate for using published data by a third party. We allowed the respondents to select multiple choices. **Table 6** presents the number and percentage of respondents who chose each option. Note that three respondents, who did not select any of the options, were excluded from the tabulation.

The highest selection rate was "Cite the source of the data in a paper (include it in the bibliography)" (90.4%). Most of the respondents judged that it would be appropriate to cite the data and the paper if the data were used. 55.2% of the respondents selected "Include source of the data information in the main text" as their next choice. None of the respondents selected, "It is not necessary to describe the data in a paper."

Table 6: Appropriate methods of displaying the use of published data (n = 406).

Choices	Numbers	Rates
Cite the source of the data in a paper (include it in the bibliography)	367	90.4%
Include source of the data information in the main text	224	55.2%
Include source of the data information in the acknowledgment	99	24.4%
Add the data holder name as a co-author	47	11.6%
It is not necessary to describe the data in a paper	0	0.0%

(8) Requests and concerns about data reuse

The respondents were asked, "Do you have any requests or concerns if the data you've published will be used for commercial activities, patents, press, literature, art, etc.?" in an additional comment space. This question aims to identify any other requests or concerns not raised in the literature or interview survey. As a result, 197 respondents responded. The major concerns were as follows: citation or indication of authorship (99 respondents), concern about misuse or inappropriate use (35 respondents), and concern about commercial use (14 respondents).

(9) Desired approach to data use and publishing

The respondents were asked in a multiple-choice format about their preferred approach of data use and publishing. The choices were made with reference to the interview survey results. **Table 7** presents the number and rate of respondents who chose each option. Note that five respondents who did not select any of the options were excluded.

Table 7: Desired approach for data use and publishing (n = 404).

Choices	Numbers	Rates
Establishment of standard data licenses (conditions of use)	312	77.2%
Development of appropriate guidelines for data licensing	285	70.5%
Establishment of a data licensing consultation, support, and management department (organization)	168	41.6%
Enabling a license to be specified in the data retrieval system	155	38.4%
Development of data rights legislation	148	36.6%
Establishment of a governing body for data licensing (external organization)	95	23.5%
Nothing in particular	21	5.2%

Minamiyama et al: Investigation and Development of the Workflow to Clarify Conditions of Use for Research Data Publishing in Japan

Art. 53, page 11 of 20

The highest rate was "Establishment of standard data licenses (conditions of use)" (77.2%), followed by "Establishment of guidelines for data licenses" (70.5%). Moreover, "establishment of a data licensing consultation, support, and management department (organization)" (41.6%) was selected higher than "establishment of a data licensing management organization (external organization)" (23.5%) as a contact point of data licensing issues.

(10) Free comments

A total of 84 respondents described the situation in the free comments. Regarding data publishing in general, various issues were pointed out, including inadequate systems, infrastructure, and technical difficulties and concerns about data publishing.

5. Design of conditions of use for publishing research data

We discussed the design of conditions of use that shall apply to research data publishing based on the results of the two survey analyses.

The two possible reasons for not publishing research data are as follows: 1) external constraints, such as legal or customary restrictions, and 2) data holder's intention. Responses to violations would be different; hence, we clearly separated the two and discussed them. We also refer to the data user's perspective.

5.1 External constraints on research data publishing

In this section, we organize the external constraints regarding when to publish research data based on the input from the "Legal Interoperability of Research Data" guidelines (RDA-CODATA Legal Interoperability IG, 2016) and survey results. **Table 8** shows its category, definition, constraint subject matter, and some examples. Note that the examples described are not exhaustive.

1) Discipline agreement and international treaties

In some cases, research data publishing is restricted by the discipline agreement of the field or research community, such as cases in which the research data publishing causes harm to the research subject or cases in which the subsequent research activities themselves are severely affected. Although protection policies are established as international treaties in many cases, let us keep in mind that these policies are not always clearly defined, known, nor applied.

2) Personal Information

In some cases, research data publishing is restricted to protect personal information. Its cases also include a restriction for disclosure, transfer, and anonymization or so by relevant local Japanese laws, cross-border rules (e.g., GDPR), and specific globalized guidelines (e.g., medical information).

3) Diplomatic/national security

If the data is related to national security or international relations (please see examples above), research data publishing is restricted. These data are strictly operated in the global context, including the conditions of use.

4) Agreements, contracts, and intellectual property rights

In some cases, data disclosure is restricted by contracts. For example, when a company and a researcher collaborate on a research project because it is not a direct matter of concern, the conditions of use and the embargo period for publishing are not uniformly determined in many cases. Then many agreements, contracts, and intellectual property rights are concluded in a local context.

5) Data Policy

In various cases, the data policy is defined by the data holder's organization (department) or research funding agency. There are many possible reasons, for example, a research funder has a policy on limiting data publishing for the research to be funded; when a patent has been applied for; or the commercialization of research results is expected. These cases are restricted as an individual strategic decision and similar to previous data disclosure through contracts. The difference is that it is a management decision based on an "open-closed strategy", which is a strategy to handle data by separating what should be released (open) and what should be protected (closed) based on the characteristics of the data.

Table 8: List of external constraints on research data publishing.

Category	Definition	Subject	Example
Discipline agreement and international treaties	Practices and standards in a specific discipline or research community that limit the data publishing. In some cases this is stated as an international treaty, but in others it is not always explicitly stated.	Disciplines & Norms	Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES)
			Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property 1970
			Convention Concerning the Protection of the World Cultural and Natural Heritage
			The Convention on the Protection and Promotion of the Diversity of Cultural Expressions
			The Nagoya Protocol on Access and Benefit-sharing
			Recommendation on the Safeguarding of Traditional Culture and Folklore
			Bereaved family's request
Personal Information	It stipulates the handling of data that can identify individuals. It includes guidelines that define individual policies on anonymization and information disclosure.	Societies	The Personal Information Protection Commission, Government of Japan. "Laws and guidelines" (only in Japanese)
			Japan External Trade Organization(JETRO). "About General Data Protection Regulation (GDPR)" (only in Japanese)
			Ministry of Health, Labor and Welfare (Japan). "About research guidelines" (only in Japanese)
Diplomatic/ National security	Research data pertaining to national security. Data related to the development of weapons of mass destruction, etc. (as defined in the Foreign Exchange and Foreign Trade Act) and defense secrets (the Self-Defense Forces). law), important data that may affect national life (e.g., domestic energy (e.g., location of resources, blueprints for critical equipment, etc.).	State	Japan Society for Intellectual Production. "Security Trade Control Guidelines for Researchers in universities and other institutions of higher education. Revised 2nd ed"
Agreements, contracts, Intellectual Property rights	An agreement with a research partner, contractor, etc. that restricts the data publishing in joint research or contract research.	Companies, etc.	Ministry of Economy, Trade and Industry (Japan). "Operation guidelines for data management in contract research and development" (only in Japanese)
			Ministry of Economy, Trade and Industry (Japan). "Contract Guidelines on Utilization of AI and Data. Data Section"
Data Policy	Where the research funder has a policy on limited data sharing for the research to be funded, or where a strategic business decision restricts the data publishing relating to pending industrial property rights or research data where the commercialization of the research results is envisaged.	Institutions	National Institute for Environmental Studies (Japan). "NIES Data Policy" (only in Japanese)
			Teikyo University (Japan). "Intellectual Property policy in Teikyo University" (only in Japanese)
			Japan Agency for Medical Research and Development (Japan). "Data sharing policy for realization of genomic medicine" (only in Japanese)

Minamiyama et al: Investigation and Development of the Workflow to Clarify Conditions of Use for Research Data Publishing in Japan

Art. 53, page 13 of 20

As a result, external constraints consist of five categories, and the specific constraint requirements are determined by localizing in each subject matter. However, some external constraints have an ambiguity that arises from a global perspective, such as international treaties or GDPR. When standardized conditions of use are to be designed, the requirements of each external constraints must be localized for a category.

5.2 Setting of conditions of use by the data holder

This section discusses the setting of conditions of use at the request of the data holder. In the previous questionnaire survey, requests and concerns about reusing data included citations, responses, disclaimers for misuse and inappropriate use, commercial use, alteration, and reporting of use. **Table 9** shows each condition of use analyzed from the perspective of expected users, duties, and constraints. We also categorized the conditions of use indicated in the questionnaire as "Preferable," "Available," and "Not Preferable" from the perspective of data publishing.

Other items should also be categorized as "Requests" rather than included as "Conditions of Use." "Requests" for the public are not legal contracts, but mainly moral matters; hence, no-obligation, prohibition, nor permission easily occur. The violation does not immediately imply termination of use, but data providers ask data users to comply as much as possible with the data holder's request for the appropriate use of their data. The survey results did not allow us to judge the survey's validity; therefore, we did not include it in the table. A further study is needed.

Table 9: List of each condition of use analyzed from the perspective of expected users, duties, and constraints.

Condition of use	Expected user	Type of duty	Target of constraint	Compatible with CC licenses	Suggested categories
1) Waiver	Public	–	–	CC0	Preferable
2) Credit on the results (CC license term: Attribution)	Public	Obligation	Redistribution	BY	
3) Impose the same conditions when publishing results (CC license term: Share-Alike)	Public	Obligation	Redistribution and combination	SA (only for redistribution)	Available
4) Noncommercial	Public	Prohibition	Redistribution and data processing	NC (only for redistribution)	
5) NoDelivs	Public	Prohibition	Redistribution and data processing	ND	
6) Improper use of data	Public Specific	Prohibition	Data processing	–	
7) Reporting	Public Specific	Obligation	Continue to use	–	
8) Secondary use prohibited	Specific	Prohibition	Redistribution	–	Not Preferable
9) Request to use the latest version	Public (latest version only)	Obligation	Redistribution	–	
10) Fee for use	Specific	Obligation	Redistribution	–	

Basis of categorization

This section presents a discussion of the abovementioned three categories. From data publishing perspective, for data to be considered published, an unspecified number of users (the public) must be given access, even if on some limited conditions. We categorized the targets assumed by each condition of use as "public," "specific," or both. If the conditions of use are only "specific," we categorized these conditions of use as "Not Preferable" for data publication.

We then analyze what obligations are imposed in the conditions of use and summarize the types and targets of these obligations. Consequently, we found two cases in which restrictions were placed on the

"redistribution of data" and on the "data itself." Restrictions on the use of the data are not desirable from the data-intensive science/open data perspective. The conditions of use not classified as "not Preferable" were classified as "Preferable" when restrictions were placed on the redistribution of data. We classified the rest as "Available."

Preferable
1) Waiver

This declaration waives copyright and all other related rights; hence, it can be evaluated to be definitely intended for the public. The cost to the data user is minimized because it is consistent with the legal requirements. The author's name need not be displayed; therefore, responsibility for misunderstanding (social risk) is less likely to occur. However, data providers are not credited, and no incentive is given to publish. Moreover, changing the conditions of use in a manner that makes them more stringent is extremely difficult, even when there is a desire to prevent unwanted use due to changes in circumstances, such as increased property values. Although this condition of use is ideal from the viewpoint of use or redistribution, note that the number of data actually published may be limited.

2) Credit on the results (CC license term: Attribution)

This condition of use requires displaying the creator's name and data URL information. The cost to data users can be assessed as a negligible level because it remains a "minimal constraint" found in the openness debate. Therefore, no problems are encountered when evaluating this condition of use aiming for the public.

The creator's name is displayed; thus, a certain incentive can be given to the data provider. In addition, the data citation expectation is particularly high (from Q7) as a method of presentation considered to be appropriate when published data are reused by a third party. Although a certain amount of recognition is given for data inclusion in the text and acknowledgments, the direction of the data is that they should be treated as independent artifacts rather than as complements to a specific article.

On the contrary, as data-specific concerns, it may be too costly and impractical for the data user to deal with a lot of different data sources as the source data for data-intensive science (e.g., machine learning). Describing all credits in the presence of multiple data sets takes time and effort. One of the commenters stated that this should be resolved as a problem with citation and notation methods. Although it is out of the scope of this paper, much discussion on this topic has taken place within the Data Citation Synthesis Group in FORCE11 (Data Citation Synthesis Group, 2014) and elsewhere.

3) Impose the same conditions when publishing results (CC license term: Share Alike)

This condition of use requires the same condition of use under redistribution and combination of multiple pieces of data. Unlike "Attribution," it may prevent them from combining the data with other sets that have an incompatible license. Although it remains the "minimum constraint" found in the openness debate for redistribution, from the viewpoint of data utilization, it should be used with caution.

4) Noncommercial

This condition of use requires the "non-commercial" use of data. The habit of prohibiting commercial use is deeply rooted in the academic community, and it seems unavoidable given the significance of academia's freedom from the society. On the contrary, although we may consider it to be out of the philosophy of open data, the criteria for judgment fluctuate depending on people because "commercial use" is not clearly defined. Also, the limitation is on redistribution and data processing (e.g., selling visualizations derived from the data). his result implies that the data of academia may be more public by adopting this condition of use, but it should be used with a more clarifying scope of commercial use. The ambiguity of commercial use was also pointed out in discussions on copyrighted materials (Creative Commons, 2009). A more careful survey by each field will be necessary for the future.

Available
5) NoDerivs (No Derivatives)

This condition of use prohibits data publishing after any modification. Although opening to the public is not restricted, based on these conditions of use, the cost to the data user is high because data use requires permission. The data will generally be published for new knowledge through processing or combination. From the viewpoint of data utilization, it should be used in a limited manner.

From the viewpoint of the data holders, the most frequent concerns are data alteration, falsification, fabrication, and misuse.

Minamiyama et al: Investigation and Development of the Workflow to Clarify Conditions of Use for Research Data Publishing in Japan

Art. 53, page 15 of 20

Furthermore, the case in which the prohibition of modification is effective is presumed to largely depend on the type of data and the manner of use (e.g., image data that are practically a work of art). Another survey for each type of data should be conducted, and more specific conditions of use for each must be set.

6) Improper use of data

This condition of use prohibits "Improper use" of data in the data processing phase. The data can be reused by both public or specific situations under this condition of use. However, the inappropriate use of data in the legal context would be covered by the Unfair Competition Prevention Act after its publication. Therefore, it is just a clear statement to users that legal and customary inappropriate treatment is prohibited. The definition of inappropriate use will probably depend on the conventions of the field. However, unclear conditions of use lead to contraction of usage. Therefore, the "improper" use in terms and conditions of use must be enumerated and specified.

7) Reporting

This condition of use requires post-reuse reporting, suggesting that the objective is to know detailed usage practices rather than mechanical access statistics. Although the data can be reused by both public and specific situations, the condition of use is stricter because of the "proactive" obligation. On the other hand, effectiveness may be realized if the data user is identified by linking to the relevant data. However, traceability cannot be guaranteed in the case of published data. In reality, it may be only at the level of a "request."

Not Preferable
8) Secondary use prohibited

This condition of use prohibits the secondary use of data. A mix of concerns about misuse/responsibility for quality and a desire to accurately understand the users have been observed. This license clearly prohibits data redistribution, translation, or adaption and is intended for one-on-one use of its original form. Although the data are not restricted, they cannot be re-distributed at all and have to be excluded from the definition of data publishing.

9) Request to use the latest version

This condition of use is used to limit the use of data to the latest one. Data in the past cannot be reused; thus, a large amount of data will be replaced when the latest data are published, resulting in marked costs of data usage. In addition, it is impossible to know in advance when the condition will be violated, and the manner to notify the version update is very limited.

10) Fee for use

This condition of use requires some fee for data use. The requirement of a user fee before data use is considered to be out of the scope of a condition of use that assumes that the data will be open to the public. The survey results also suggest that approximately half of the respondents are still uncomfortable with the act of monetizing data. However, given the sustainability of the data repository, monetization may be a major challenge in the future. The fee could be obtained in various ways, including shareware on a request basis, charging through a freemium model, download speed limits, and whether or not ads are displayed. In cases where the data holder itself requires a user fee, under what conditions the fee will be incurred must be clarified.

5.3 Data user's perspective

According to the questionnaire survey, we can see that there is a lot of concern in the topic of citation, misuse/inappropriate use, commercial use. Also, as the "Desired approach to data use and publishing," 70% or more mentioned that the standard data conditions of use and licensing guidelines had been established. From the data holder's perspective, compliance with the granted conditions of use leads to safe data publishing. The data user is required to understand the external constraints behind the granted conditions of use. However, as observed in Section 5.1, specialized knowledge is needed to determine whether external constraints will occur. In other words, it may not be possible to solve the problem by setting clear documents for conditions of use, and it is likely to be necessary to adopt the use of the system to suit easy-to-understand usage. And it may be possible to position the establishment of standard conditions of use and guidelines as a method for providing easy-to-understand usage. This survey was conducted from the data holder's viewpoint. Although not directly derived from this survey, the standard condition of use selection tools for research data is required (e.g., CC licenses for copyrighted works). Further research and data analysis are needed to establish the standard conditions of use and guidelines from the data user's viewpoint.

6. Implementation

Based on the survey and understanding of the conditions of use for research data publishing discussed in the previous Sections, we developed the "Guideline for specifying conditions of use in research data publishing" (Research Data Utilization Forum, 2019) as a tool to help researchers and stakeholders in common understanding and make appropriate publication decisions. This Section introduces the guideline framework and describes what can be achieved by using them and their limitations.

The survey results show that respondents are willing to make their data public if they can demand conditions of use. Therefore, data publishing may proceed with the appropriate conditions of use (and guaranteed for feasibility). This guideline provides necessary information and examples that should be considered when publishing research data. This guideline can also be used as a tool to easily understand the outline of the conditions of use required by data holders. This guideline is intended for researchers (universities, companies, etc.), engineers who publish or use data, and persons in charge who support data publishing in their institutions (academic institutions, libraries, academic societies, academic publishers, etc.).

6.1 Overview of the guideline

The scope of the guideline is limited to publishing data for the public. The guideline suggests standard conditions, called Covenants, as a framework. The option is designed to set appropriate conditions of use by following a workflow. As discussed in the previous Section, some conditions of use arise from external constraints, while others could be set by the data holder. If you handle the data that are in borderline with copyrighted works, the conditions of use for these research data must be selected in a manner that is compatible with the existing licensing tools. This framework is designed to provide data protection equivalent to that of a copyrighted work and set appropriate conditions of use by following a process.

Figure 5 shows the data publishing flow with licensing scenarios. The flow consists of five steps. By taking these steps, the user (i.e., mainly, the data holder or data user) can check the data publishing procedure step by step. First, the user identifies the data to be published in Step 1. Next, the user confirms the external constraints that may occur in data publishing in Step 2. In Step 3, for the constraints identified in Step 2, the user confirms the necessary processes for enabling data publication (e.g., setting the embargo period). Steps 2 and 3 clearly state that expert consultation will be held because expert knowledge may be required for judgment. In Step 4, the user can select the appropriate data repository for the data judged to be open to the public. Finally, the user chooses appropriate conditions of use with detailed guidance in Step 5. The details for each step are shown below.
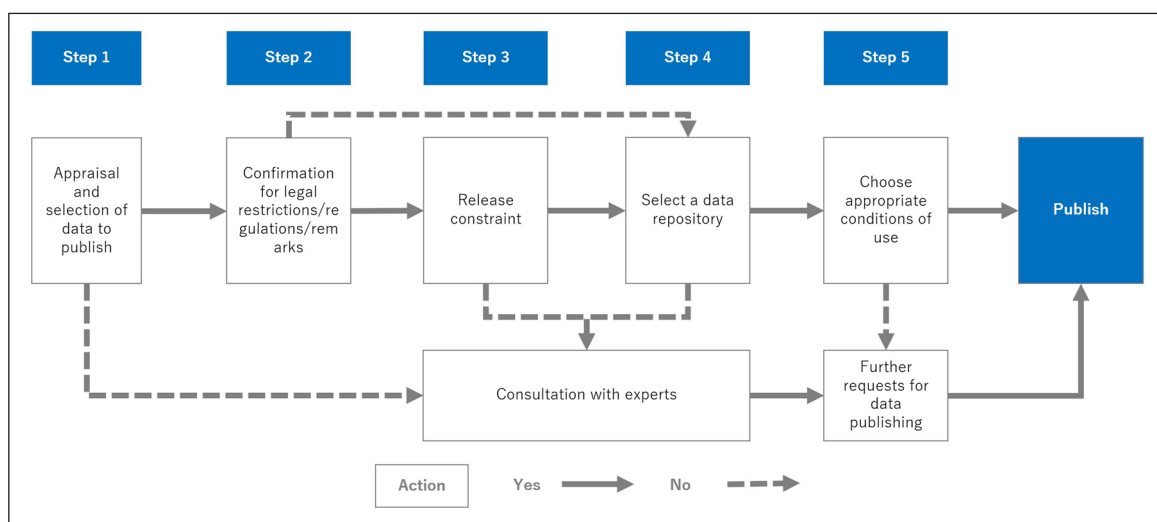

Figure 5: Data Publishing flow with licensing scenarios.

Step 1: Appraisal and selection of data to publish

In this step, the data holder identifies various data used in the research, which can be curated and made available to the public. There are various types of data publishing motivation; mandated by publishers, funders, or institutional policies, and by researchers' requests. Although the scope of "research data" differs

Minamiyama et al: Investigation and Development of the Workflow to Clarify Conditions of Use for Research Data Publishing in Japan

Art. 53, page 17 of 20

depending on the field of expertise, this guideline defines "research data" as data that can be managed by digital means and released as research results and do not include physical objects such as samples, specimens, and recording media (paper, disks, etc.). In addition, although research articles and software can be treated as research data, the guideline does not change or override the established methods for publishing in each content area (e.g., CC licenses for paper publishing and GPL or other software licenses for software publishing). If a researcher has received research funding, she/he should follow the rule of the treatment of research data defined by the funding agency. The guideline does not apply to such data; hence, their rules should be applied.

Step 2: Confirmation for legal restrictions/regulations/remarks
In this step, the data holder considers whether or not the data identified in Step 1 falls under the following categories of external constraints shown in Section 5:

- Disciplinary customs, including international treaty
- Personal information
- Diplomatic/national security
- Agreements, contracts, and intellectual property rights
- Data policy

The abovementioned factors correspond to the constraints set out in 5.1, which can be confirmed with some examples.

Step 3: Release constraint
In this step, the data holder identifies and sets the conditions or time period required before the constraints found in Step 2 can be lifted by category. The terms or periods set out here will be written into the conditions as special conditions.

Even in cases where legal/customary restrictions are imposed, restrictions may be lifted with appropriate data processing (e.g., anonymization) or data release restrictions for a certain period time. At this time, there is no legal provision for the termination of the protection period for data, as there is for copyrighted works. For example, even if the term of the collaboration agreement has expired, the data are apparently not open to the public after any length of time, unless the term is clearly defined. To prevent these unnecessary restrictions, the guideline provides explicit steps for lifting the restrictions and encourage users to keep them to a minimum. If the research data publishing cannot be decided at the time of review, the "in case research data cannot be published" option recommends creating metadata and data storage to enable a later decision.

Step 4: Select a data repository
In this step, the data holder selects the repository, where he/she wants to publish the data. Well-known repositories/archives are likely to be the first candidate. However, the confirmation of external restrictions in Step 2 is the step built on the premise of Japanese law or regulations. Therefore, repositories constrained by other foreign rules may not necessarily cover all points for consideration. There are some famous registry sites such as re3data.org and FAIRsharing, however, only a limited number of registrations are available in Japan. In light of this background situation, we provide the list of recommended domestic data repositories in cooperation with the Japan Data Repository Network subcommittee under the RDUF. We also prepared a list of legal measures that can be applied under the Japanese law and clearly indicated them to eliminate them and respond to concerns about inappropriate use, which may accompany the lifting of restrictions.

Step 5: Choose appropriate conditions of use
In this step, the data holder selects the conditions of use that fulfils the requirements of data users and completes the standard conditions of use (covenants) that set out conditions consistent with those protected by copyright law. The data requirements are more diverse than those for copyrighted works, and the situation has not yet been systematically organized. Furthermore, there are high demands for standardization and simple explanations. As discussed, a concern has been raised: the simple recommendation of open licenses avoiding the copyright problem will not lead to the promotion of use.

We already analyzed the questionnaire survey results in Section 5 to identify the "Preferable" conditions of use. However, as a practical consideration, we added "Impose the same conditions when publishing results," "Noncommercial," and "Noderivs" to the preferable requirements in the guideline because some contents are difficult to distinguish from copyrighted works when giving conditions for data usage. To the previous CC license discussions and to ensure interoperability, we provided "Impose the same conditions when publishing results," "Noncommercial," and "Noderivs" with the explanation of its validity only under limited conditions. We take care to minimize the effort involved in setting the conditions by explaining how to describe specific conditions and usage information (agreements). These standard conditions of use (covenants) are designed to function as a part of the data usage policy of each data repository.

6.2 Remaining issues of the guideline

The guideline prioritizes practical use and presents the preferable requirements in a manner consistent with existing licensed tools. Therefore, the additional categories of conditions of use revealed in the questionnaire survey analysis are not included. They should be included in the future version systemically, i.e., some controlled vocabulary or ontology is needed. The crucial point is that the use of research data is expected to be different from the original use of data set creation, such as using the data set as training data for machine learning.

7. Conclusion

In this study, we investigated and developed the workflow to determine conditions of use for research data publishing in Japan. There are two reasons to prevent from publishing research data. One reason comes from the external constraints. The external constraints consist of five categories, and the specific constraint requirements are determined by localizing in each subject matter. The other comes from the condition of use. We found that the conditions of use by the data holder is more varied than copyrighted works, and that many are not standardized.

Based on the above the observation and the discussion, we then developed the categorization of the condition of use from the perspective of the data publishing and the publishing workflow with licensing scenarios. By using this category, it can be expected to clarify the actual meaning of conditions of use and their interpretation in different local contexts and different requests by the data holder. Furthermore, this flow helps to organize the diverse conditions of use that are disciplined in local contexts while maintaining interoperability at the conceptual level in global contexts.

We believe that the work contributes not only to reduce daily efforts in research data publishing but also to develop an infrastructure for data-intensive science, which will consequently lead to the realization of Open Science. In the future, the data holder requirements will be clarified through a higher resolution by collecting data on the granting of conditions of use based on this guideline.

Competing Interests

The authors have no competing interests to declare.

References

ANDS. 2018. Publishing and sharing sensitive data: ANDS Guides: 23p. https://www.ands.org.au/__data/assets/pdf_file/0010/489187/Sensitive-Data-Guide-2018.pdf.

Assante, M, Candela, L, Castelli, D and **Tani, A.** 2016. Are Scientific Data Repositories Coping with Research Data Publishing? *Data Science Journal*, 15: 1–24. DOI: https://doi.org/10.5334/dsj-2016-006

Baldwin, J and **Pinfield, S.** 2018. The UK Scholarly Communication Licence: Attempting to Cut through the Gordian Knot of the Complexities of Funder Mandates, Publisher Embargoes and Researcher Caution in Achieving Open Access. *Publications*, 6(3): 31. DOI: https://doi.org/10.3390/publications6030031

Minamiyama et al: Investigation and Development of the Workflow to Clarify Conditions
of Use for Research Data Publishing in Japan

Art. 53, page 19 of 20

Ball, A. 2014. 'How to License Research Data'. *DCC How-to Guides.* Edinburgh: Digital Curation Centre. http://www.dcc.ac.uk/resources/how-guides.

Beagrie, N. 2006. Digital Curation for Science, Digital Libraries, and Individuals. *International Journal of Digital Curation*, 1(1): 1–6. DOI: https://doi.org/10.2218/ijdc.v1i1.2

CODATA. 2019. 'The Beijing Declaration on Research Data'. 6p. DOI: https://doi.org/10.5281/zenodo.3552330

Creative Commons. 2009. Defining "Noncommercial": A Study of How the Online Population Understands "Noncommercial Use". https://mirrors.creativecommons.org/defining-noncommercial/Defining_Non-commercial_fullreport.pdf

Data Citation Synthesis Group. 2014. 'Joint Declaration of Data Citation Principles'. *FORCE11.* DOI: https://doi.org/10.25490/a97f-egyk

Grabus, S and **Greenberg, J.** 2019. The Landscape of Rights and Licensing Initiatives for Data Sharing. *Data Science Journal*, 18(1): 29. DOI: https://doi.org/10.5334/dsj-2019-029

Ikeuchi, U and **Minamiyama, Y.** 2020. Data supporting "Investigation and Development of the workflow to clarify conditions of use for research data publishing in Japan" (Version 1.0). *NII institutional repository.* DOI: https://doi.org/10.20736/00001468

Kelling, S, Hochachka, WM, Fink, D, Riedewald, M, Caruana, R, Ballard, G and **Hooker, G.** 2009. Data-intensive science: A new paradigm for biodiversity studies. *Bioscience*, 59(7): 613–620. DOI: https://doi.org/10.1525/bio.2009.59.7.12

Kindling, M, et al., 2017. The Landscape of Research Data Repositories in 2015: A re3data Analysis. *D-Lib Magazine*, 23(3/4). DOI: https://doi.org/10.1045/march2017-kindling

Klump, J, Bertelmann, R, Brase, J, Diepenbroek, M, Grobe, H, Höck, H, Lautenschlager, M, Schindler, U, Sens, I and **Wächter, J.** 2006. Data publication in the open access initiative. *Data Science Journal*, 5: 79–83. DOI: https://doi.org/10.2481/dsj.5.79

Kowalczyk, S and **Shankar, K.** 2011. Data sharing in the sciences. *Annual Review of Information Science and Technology*, 45(1): 247–294. DOI: https://doi.org/10.1002/aris.2011.1440450113

Lawrence, B, Jones, C, Matthews, B, Pepler, S and **Callaghan, S.** 2011. Citation and peer review of data: Moving towards formal data publication. *International Journal of Digital Curation*, 6(2): 4–37. DOI: https://doi.org/10.2218/ijdc.v6i2.205

Marcial, LH and **Hemminger, BM.** 2010. Scientific data repositories on the web: An initial survey. *Journal of the American Society for Information Science and Technology*, 61: 2029–2048. DOI: https://doi.org/10.1002/asi.21339

Mayernik, MS, et al., 2020. Risk Assessment for Scientific Data. *Data Science Journal*, 19(1):10. DOI: https://doi.org/10.5334/dsj-2020-010

Ministry of Economy, T and **Industry.** 2018. FY 2017 Industry and Economy Research Commissioned Project (Research on Data Protection Systems Abroad) Research Report (translated from Japanese). https://www.data.go.jp/data/en/dataset/meti_20180312_0106.

OECD. 2015. Enquiries Into Intellectual Property's Economic Impact, Chapter 7: Legal Aspects of Open Access to Publicly Funded Research in Enquiries into Intellectual Property's Economic Impact. https://www.oecd.org/sti/ieconomy/KBC2-IP.Final.pdf.

RDA-CODATA Legal Interoperability Interest Group. 2016. Legal Interoperability of Research Data: Principles and Implementation Guidelines. DOI: http://doi.org/10.5281/zenodo.162241

RECODE project consortium. 2014. Policy Recommendations for Open Access to Research Data. *RECODE*, 44. DOI: https://doi.org/10.5281/zenodo.50863

Research Data Utilization Forum. 2019. Guideline for specifying conditions of use in research data publishing, ver. 1.0. 32p. DOI: https://doi.org/10.11502/rduf_license_guideline

Ruggles, S. 2018. The Importance of Data Curation. In: Vannette, D and Krosnick, J (eds.), *The Palgrave Handbook of Survey Research.* Cham: Palgrave Macmillan. DOI: https://doi.org/10.1007/978-3-319-54395-6_39

Stuart, D, et al. 2018. Whitepaper: Practical challenges for researchers in data sharing. DOI: https://doi.org/10.6084/m9.figshare.5975011.v1

Van Panhuis, WG, Paul, P and **Emerson, C,** et al., 2014. A systematic review of barriers to data sharing in public health. *BMC Public Health*, 14: 1144. DOI: https://doi.org/10.1186/1471-2458-14-1144

Wilkinson, MD, Dumontier, M and **Aalbersberg, IJ.** 2016. The FAIR Guiding Principles for scientific data management and stewardship. DOI: https://doi.org/10.1038/sdata.2016.18