Requirements Analysis of System for Research Data Management to Prevent Scientific Misconduct

Miho Funamori
Research Center for Open Science and Data Platform
National Institute of Informatics
Tokyo, Japan
e-mail: funamori@nii.ac.jp

Masaharu Hayashi
Research Center for Open Science and Data Platform
National Institute of Informatics
Tokyo, Japan
e-mail: mhaya@nii.ac.jp

Yusuke Komiyama
Research Center for Open Science and Data Platform
National Institute of Informatics
Tokyo, Japan
e-mail: komiyama@nii.ac.jp

Masatoshi Tsuchiya
Information and Media Center
Toyohashi University of Technology
Toyohashi, Japan
e-mail: tsuchiya@imc.tut.ac.jp

Kazutsuna Yamaji
Research Center for Open Science and Data Platform
National Institute of Informatics
Tokyo, Japan
e-mail: yamaji@nii.ac.jp

***Abstract*—With the pressing need for research data management to prevent scientific misconduct (RDM-PSM) in Japan, our goal in this study was to identify the basic functionalities and requirements of an RDM-PSM system. To achieve it, we first extracted the core elements of the guidelines on preserving research materials set forth by the Science Council of Japan and reorganized the elements to form the minimum requirements and two basic functionalities, "Content Management" and "Institutional Management," of an RDM-PSM system. Next, the minimum requirements for RDM-PSM were scrutinized in order to formulate the system requirements and associated functions needed to develop the system. Finally, the functions were mapped onto an assumed RDM-PSM workflow model at academic institutions to evaluate the consistency and usability of these functions. Even though this requirements analysis was conducted on the basis of Japanese guidelines, the identified requirements and analysis procedure are useful in international contexts to meet the increasing demand worldwide for RDM-PSM.**

Keywords-research data management; research materials preservation policy; data policy; scientific misconduct; Open Science; institutional repository

I. INTRODUCTION

There is growing need for academic institutions to systematically act on research data management (RDM). Managing research data has traditionally been the responsibility of individual researchers. However, after the National Institutes of Health in the US issued "Data Sharing Policy and Implementation Guidance" in 2003 and the Committee for Scientific and Technological Policy at the Organization for Economic Cooperation and Development (OECD) issued the "Declaration on Access to Research Data from Public Funding" the following year, RDM began to be pursued as a matter of compliance [1][2]. As an increasing number of funding agencies started to require data management plans (DMPs) to be submitted along with research grant proposals and a plan for monitoring research to ensure the DMPs have been met emerged, a need for systematically conducting and supporting RDM arose. According to Sherpa Juliet, which is a website summarizing the open access policies of funding agencies, 59 funders (42%) either require or recommend data archiving (require: 29%, recommend: 13%, unknown: 58%) [3]. The rising number of academic journals asking for evidence data to be submitted as supplements to research articles is also a push for RDM [4]. The growing number of retractions of research articles due to scientific misconduct has forced academic journals to introduce such policies [5][6].

In response to the above trends, academic institutions are gradually introducing RDM support and infrastructures. RDM support at academic institutions is often provided through the collaboration of the research office, research library, and IT center within the institution, with the library as the coordination point. RDM services at academic institutions mainly cover 1) training for the basic understanding of RDM and consultations for dealing with specific needs, 2) providing an overview of funding agencies' data policies and support for writing DMPs using DMP tools, and 3) providing a list of repositories where research data can be stored and made available [7][8][9]. Some academic institutions provide repositories on top of these services. However, when it comes to actual RDM, there are still no established methods on how to store and organize research data, what metadata and descriptions to be added to make the data findable and reusable, how to decide on the range of data to provide for various purposes, how long to keep the data, and the various methods to make data available. As of now, RDM infrastructure consists merely of storage places rather than places for effective RDM.

In Japan, RDM has been introduced with strong emphasis on addressing research integrity issues. After several incidents of serious scientific misconduct occurred in Japan, the Ministry of Education, Culture, Sports, Science

978-1-5386-7447-5/18/$31.00 ©2018 IEEE
DOI 10.1109/IIAI-AAI.2018.00083

and Technology (MEXT) decided in 2014 to strengthen the "Guideline for Dealing with Scientific Misconduct," which was put into effect back in 2006 [10][11]. The 2006 guideline held individual researchers responsible for research transparency and preventing scientific misconduct. The 2014 guideline enlarged these responsibilities by also holding the academic institutions responsible for preventing scientific misconduct. The idea of preserving research data for a certain period to prevent scientific misconduct was also introduced. It was put forward by the Science Council of Japan (SCJ) upon request by MEXT. In 2015, SCJ issued the report "Reply: For the Enhancement of Soundness of Scientific Research" clarifying the details of the 2014 guideline [12]. The so-called "Ten-Years Preservation Rule for Research Data" was made clear in this report along with the "Guideline for the Preservation of Research Materials for Scientific Misconduct Prevention." Note that the SCJ guideline uses the term "research material" instead of "research data," which will be used hereinafter in this article. It should be added that there is also a push from funding agencies to comply with RDM. However, as DMPs are only required by some of the discipline-specific funders and not by the general funder, the Japan Society for the Promotion of Science (JSPS) to which all researchers across disciplines are eligible to apply, it can be said that RDM in Japan is mainly intended to prevent scientific misconduct [13][14][15].

To comply with the guidelines of MEXT and SCJ, many research-intensive academic institutions in Japan have introduced a research materials preservation policy to prevent scientific misconduct. However, the implementation of these policies is still on weak ground, as the guidelines and according institution's policies have been introduced just recently and there is no specific support system or infrastructure established to comply with the policies. In a policy, the institution holds the departments and the departments hold the researchers or research labs responsible for preserving the research materials for a given period. Most institutions do not provide any storage for long-term preservation of research materials, and storage is the responsibility of the researcher or research lab. Researchers are asked to report on their own publications and on what research materials have been collected for them. This reporting is done by filling out a table on an Excel spreadsheet, which the institution routinely sends out to researchers [16][17]. As the report of one research publication accounts for only a single row in the table, not much information can be conveyed on the whereabouts of research materials related to the research publication. As the Excel spreadsheet also does not provide a direct link to the research materials, one can assume that it would be extremely difficult to identify the research materials and test the reproducibility of the research findings for an incident of scientific misconduct. In fact, academics have voiced concern on the effectiveness of this approach to research materials preservation. The provision of appropriate tools and infrastructure for RDM is needed in order to prevent scientific misconduct.

This research aims to identify the basic functionalities and system requirements of a "System for RDM to Prevent Scientific Misconduct" (RDM-PSM system), which is urgently needed at Japanese academic institutions. The SCJ guideline, which provides detailed requirements for research materials preservation and associated RDM-PSM procedures at academic institutions, was analyzed to identify the core elements, and the minimum requirements and basic functionalities of an RDM-PSM system were derived from them. The minimum requirements for RDM-PSM in the guideline were scrutinized in order to formulate the system requirements and associated functions needed to develop an RDM-PSM system. Finally, these functions were mapped onto an assumed RDM-PSM workflow at academic institutions to evaluate the consistency and usability of the derived RDM-PSM functions.

II. Derivation of Basic Functionalities of RDM-PSM System

A. Identifying Reference Document for RDM-PSM

To derive the basic functionalities for developing an RDM-PSM system for Japan, the document that underlies the procedure for RDM-PSM in Japan was identified.

The MEXT and SCJ guidelines form the basis for the implementation of RDM-PSM at Japanese academic institutions. Even though these documents are guidelines and not mandates, most research-intensive academic institutions in Japan have been following them in response to calls for accountability after several incidents of serious scientific misconduct happened. MEXT conducted a survey in 2015 to evaluate the implementation of the guideline issued in 2014 dealing with scientific misconduct. Among all Japanese academic institutions, 790 (44%) reported having an effective research materials preservation policy in place. Among the national universities, which are mostly research-intensive, 76 (84.4%) reported having an effective policy in place [18].

Some institutions have stand-alone policies, while others have them built into their scientific-misconduct-prevention policies. The terms included in these policies, however, tend to follow the SCJ guideline. For example, the University of Tokyo, which is considered the most prestigious research university in Japan, has introduced its own guidelines specifying materials to be preserved, responsibilities of researchers and the institution, preservation periods, and preservation methods [19]. The Kyushu Institute of Technology, which was highlighted in the MEXT survey as a model for research materials preservation, includes similar guidelines in its "Policy for the Prevention and Measures for Scientific Misconduct" [16][17]. In fact, many universities use the MEXT survey as a reference document when drafting their own policies, so it is likely that they have similar policies as that of the Kyushu Institute of Technology.

As most academic institutions in Japan follow the MEXT and SCJ guideline and as the MEXT guideline sets out the principles, whereas the SCJ guideline sets out the detailed procedures, we decided to use the SCJ guideline as the reference document for our analysis. The "Guideline for the Preservation of Research Materials" mentioned in Section 2(2) of the SCJ guideline on "The basic duties and care of an

academic researcher, and the period and means to preserve research data" details the requirements for research materials preservation and was used for our requirement analysis of an RDM-PSM System.

We decided to use the term "research materials" rather than "research data," since both the MEXT and SCJ guidelines use the term "research materials preservation." The RDM-PSM system is supposed to be an e-infrastructure where research data and research materials can be treated as the same electronic content. Additionally, we focused on only research materials linked to specific outputs such as research publications; materials related to other endeavors by researchers were not part of the study. Both guidelines call for research materials preservation to prevent scientific misconduct and for "effective self-defense in case of being suspected of scientific misconduct." Here, we should note that incidents of scientific misconduct are most often detected in research publications, and the SCJ guideline states that it makes sense to preserve only research materials directly or indirectly linked to tangible research outputs in consideration of the cost of preserving all research materials.

B. Element Analysis of SCJ Guideline—Derivation of Minimum Requirements of RDM-PSM System

The SCJ guideline was analyzed to identify the core elements that would form the basic functionalities of an RDM-PSM system.

The guideline consists of six instructions concerning: 1) preservation of lab notebooks in which the daily research activities are recorded, 2) preservation of research materials, 3) means and period of preservation of digital research materials, 4) means and period of preservation of physical research materials, 5) procedures for when a researcher or lab head is leaving to another institution or retiring, and 6) research materials that require special treatment. Each instruction consists of two to three sentences.

The element analysis of the guideline was conducted by breaking down each sentence of the guideline into distinctive elements. Table 1 shows the analysis procedure for instruction 2) above. The upper box denotes the sentences in the guideline. The numbers in that box denote the instruction number and sequential sentence number for each instruction. The lower box shows the elements extracted from each sentence. The elements are numbered sequentially. This procedure was carried out on all six instructions of the guideline.

After extracting all the elements from the six instructions, those overlapping in meaning were unified into a single element and the elements were reorganized as shown in the left column of Table 2. The column denotes the extracted elements, and the numbers in parentheses show from which element in Table 1 each was derived. For instance, "2.1.2 appropriate format for future reuse" and "2.2.1 future reuse is enabled" are merged as "reusable" in Table 2. The gray headings in Table 2 denotes the categories used when reorganizing the elements.

The extracted and reorganized core elements of the SCJ guideline in Table 2 form the minimum requirements for the development of an RDM-PSM system.

TABLE I. SAMPLE OF SCJ GUIDELINE ELEMENT ANALYSIS (TERM 2)

SCJ Guideline-instruction 2

2	**Preservation of Research Materials**
2.1	Research materials (documents, numerical data, images, etc.) used for research outputs, such as research publications and oral presentations, shall be preserved in appropriate format for the future reuse and testing of research outputs.
2.2	Upon preservation of research materials, proper measures shall be taken so that future reuse and referencing are enabled; measures include adding metadata and enabling of discoverability and trackability of research materials.

Extracted elements

2.1.1	research materials (documents, numerical data, images, etc.) used for research outputs such as research publications and oral presentations
2.1.2	appropriate format for future reuse
2.1.3	appropriate format for the future testing of research outputs
2.2.1	future reuse is enabled
2.2.2	future referencing is enabled
2.2.3	adding metadata
2.2.4	enabling discoverability
2.2.5	enabling trackability of research materials

C. Deriving Basic Functionalities of RDM-PSM System

The basic functionalities of an RDM-PSM system were derived from the minimum requirements in Table 2. First, we eliminated the elements that do not need to be considered to develop an e-infrastructure, and then, scrutinized the remaining elements to derive the basic functionalities of an RDM-PSM system, namely "Content Management" and "Institutional Management." As the aim was to derive only the basic functionalities of an RDM-PSM system, the fewer functionalities that are derived, the better it is. Thus, any added functionalities, such as features to enhance the user experience or the effectiveness of the system, are out of the scope of this study.

The elements that do not need to be considered for deriving the basic functionalities of an RDM-PSM system are eliminated. These are denoted as "N/A" or "(Prsvd Mat)" in Table 2. These elements were eliminated for the following reasons. The elements that are only concerned with physical research materials under "I. Materials to be Preserved" were eliminated because we considered only digital materials that can be handled by an e-infrastructure. Elements under the terms "II. Preservation Period" and "V. Considerations for Preservation" were eliminated, because these refer only to modification of the parameters at a system level or how people use the system. Likewise, the academic institutions drafting policies and researchers signing a memorandum upon hiring under "VI. Institutional Management Methods" were also eliminated. The elements denoted with "(Prsvd Mat)" under "I. Materials to be Preserved" were the digital content that an RDM-PSM system would deal with. Since the system will be able to handle all digital content equally,

TABLE II. MINIMUM REQUIREMENTS AND BASIC FUNCTIONALITIES OF RDM-PSM SYSTEM DERIVED FROM SCJ GUIDELINE

Minimum Requirements for RDM-PSM		Basic Funct.
I. Materials to be Preserved		
	Research activities (1.2.1)	(Prsvd Mat)
	Condition of data acquisition (1.1.1, 1.1.2)	(Prsvd Mat)
	Research materials used for research outputs such as research publications and oral presentations (2.1.1, 3.1.1)	(Prsvd Mat)
	- Digital materials (3.2.1)	(Prsvd Mat)
	- Paper-based materials (3.3.1)	N/A
	- Physical materials such as specimens and devices (4.1.1)	N/A
	Research outputs (6.2.1)	(Prsvd Mat)
	Metadata (2.2.3)	(Prsvd Mat)
II. Preservation Period		
	Research materials need to be preserved for ten years after publishing of research outputs (3.1.2, 3.3.2)	N/A
	Physical materials, such as specimen and devices, need to be preserved for five years after publishing of research outputs (4.1.2)	N/A
III. Conditions for Preservation		
	Inalterable (1.1.3)	Cont Mgmt
	Findable (2.2.4)	Cont Mgmt
	Trackable (2.2.2, 2.2.5, 5.1.3)	Cont Mgmt
	Verifiable (2.1.3)	Cont Mgmt
	Submittable (5.1.4)	Cont Mgmt
	Reusable (2.1.2, 2.2.1, 3.2.4)	Cont Mgmt
IV. Preservation Methods		
	Creating back-ups (3.2.3, 5.1.2)	Cont Mgmt
	Adding and managing metadata (3.2.2)	Inst Mgmt
V. Considerations for Preservation		
	Space constraints for preservation (3.3.3)	N/A
	Fundamentally impossible to preserve (4.2.1)	N/A
	Immense cost for preservation (4.2.2)	N/A
	Existing legal regulations (6.1.1)	N/A
	Ethical consideration needs (6.1.2)	N/A
	Special agreement with funders (6.2.2)	N/A
VI. Institutional Management Methods		
	Academic institutions can test the research materials created by their researchers even upon the leaving or retirement of researchers and lab heads (5.1.1, 5.2.1)	Inst Mgmt
	Academic institutions can test the research materials created by their researchers, especially those directly related to research outputs (5.1.2, 5.2.2)	Inst Mgmt
	Academic institutions draft and put policy for research materials preservation into effect (5.3.1)	N/A
	Researchers sign a memorandum for the transfer of research materials to the institution upon hiring (5.3.2)	N/A

Note: 1) Basic Func.: Basic Functionalities of RDM-PSM System, 2) Prsvd Mat: Preserved Material, 3) Cont Mgmt: Contents Management, 4) Inst Mgmt: Institutional Management

the variation in research materials does not need to be considered.

Next, we identified the basic functionalities of an RDM-PSM system from the remaining elements. The elements under "III. Conditions for Preservation"—Inalterable, Findable, Trackable, Testable, Submittable, Reusable—are identified as functionalities for "Content Management." The two elements under "VI. Institutional Management Methods"—"Academic institutions can test the research materials created by their researchers even upon the leaving or retirement of researchers and lab heads" and "Academic institutions can test the research materials created by their researchers, especially those directly related to research outputs"—were identified as functionalities for "Institutional Management." Academic institutions also need to directly deal with digital content in case of scientific misconduct. However, they generally deal with metadata, such as who wrote what research publication, when, and on what media. Thus, "Adding and managing metadata" under "IV. Preservation Methods" was identified as elements for "Institutional Management," whereas "Creating back-ups" under "IV. Preservation Methods" was identified as elements for "Content Management." In the subsequent system requirement analysis, we considered a system to handle mainly digital content for "Content Management" and metadata for "Institutional Management."

III. SYSTEM REQUIREMENTS ANALYSIS OF RDM-PSM SYSTEM

A. Formulation of System Requirements and Functions of RDM-PSM System

To develop RDM-PSM as an e-infrastructure, the minimum requirements for an RDM-PSM system under the basic functionalities "Contents Management" and "Institutional Management" in Table 2 need to be re-formulated as system requirements. Table 3 lists the derived system requirements and associated functions according to each minimum requirement of RDM-PSM.

The system requirements describe under what conditions the minimum requirements need to be met. To prevent leakage of system requirements, we first listed all possible system requirements that would fulfill the minimum requirements without referring to the SCJ guideline. We then chose the ones that would be needed for RDM-PSM. Finally, we cross-checked whether the chosen system requirements do not deviate from and are in support of the SCJ guideline. By first deriving the system requirements without referring to the SCJ guideline, several requirements which were out of scope of SCJ guideline could be detected. For instance, the system requirement "Able to declare the user license of digital content" under the minimum requirement "Reusable" is needed for ensuring the reusability of digital content, even though this requirement is not considered in the guideline.

In the "Functions" column in Table 3, functional components which meet each system requirement are laid out. In some cases, there are multiple functions needed for one system requirement, as can be seen in the functions derived for the minimum requirement "Trackable." By lay-

TABLE III. SYSTEM REQUIREMENTS AND ASSOCIATED FUNCTIONS TO MEET MINIMUM REQUIREMENTS FOR RDM-PSM SYSTEM

Minimum Requirements		System Requirements	Functions
Contents Management	Findable (Content Search)	・Able to find digital content concerning scientific misconduct ・Able to add metadata of digital content needed for research data management, such as creator of content, associated research publications, grant number, and citations.	1 Metadata retrieval (including keyword search, navigation, filtering, sorting function) 2 Metadata management
	Trackable (Content Tracking)	・Able to track digital content and related researchers and institutions upon incident of scientific misconduct	1 Persistent identifier assignment and management for digital content 2 Provenance management for digital content 3 Linking digital content with related researchers and institutions
	Verifiable (Content Verification)	・Able to verify digital content upon incident of scientific misconduct ・Able to preserve the relations of files generated at different research stages such as the processing of raw data to primary and secondary data, and the final research output ・Able to preserve and associate the research workflows and software needed for data validation	1 Research-trail recording (including version control, editing and update history, and file relationship management) 2 Preserving research workflows 3 Preserving electronic experiment notes
	Submittable (Content Submission)	・Principal investigator is able to submit verifiable digital content anytime within certain period in response to an institutional investigation upon incident of scientific misconduct	1 Digital-content acquisition and archiving
	Reusable (Content Reuse)	・Able to edit and process digital content using standard format (or open format anyone can edit and modify) ・Able to declare the user license of digital content ・Able to provide or reproduce processing environment of digital data to ensure research reproducibility	1 Conversion and migration to standard format 2 User-license declaration 3 Virtual machine and container virtualization technology
	Inalterable (Content Inalterable)	・Able to prevent the tampering with digital content during research and upon scientific misconduct ・Able to preserve and confirm research trail of digital content	1 Preventing the tampering with digital content (time stamp, electronic signature, digital watermark, etc.) 2 Recording and viewing research trail
	Creating backups (Content Backup)	・Able to preserve and backup digital data for long term in a sustainable manner to ensure verifiability	1 Data backup 2 Disaster recovery 3 Automatic switching of hot and cold storage
Institutional Management	Academic institutions can test the research materials, especially those directly related to research outputs, created by their researchers (Institutional Verification)	・Able to have an overview of research outputs of the institution with associated digital content at metadata-level	1 Associating research outputs with related digital content 2 Retrieving research outputs and related digital content
	Academic institutions can test the research materials created by their researchers even upon the leaving or retirement of the researcher or lab head (Mobility Management)	・Able to devolve the digital data to another researcher in the same institution upon retirement of the principal investigator ・Able to devolve the digital data to another researcher in another institution upon retirement of the principal investigator ・Able to store provenance information of digital content to be able to track the content ・Able to delegate access rights to digital content to others upon principal investigator's demise	1 Devolving digital content to others in the same institution 2 Devolving digital content to others in another institution 3 Provenance management for digital content devolved to other institutions 4 Delegating access rights to digital content to others

Note) Research-trail: The records of creation and modification of files during research activities

-ing out the functions for system requirements in detail, the functionalities for application development are clarified. The functions are described in technical terms to be understandable for system developers without RDM knowledge. However, as the three system requirements for the minimum requirement "Verifiable" are closely linked to RDM, the related functions needed to include RDM terms such as "Research-trail recording," "Preserving research workflows," and "Preserving electronic experiment notes." It should be noted that the derived functions are not unique and that there could be some variations in the functions; especially with advancements in technology, other sets of

functions could be considered to meet the system requirements. We have laid out the most basic functions that are commonly available with current technology and comprise the bare minimum to develop an RDM-PSM system.

The system requirements could be scrutinized in detail by laying out the associated functions. For example, the system requirement for the minimum requirement "Trackable" is described as "Able to track digital content and related researchers and institutions upon incident of scientific misconduct." The associated functions were laid out as: "1 Persistent identifier assignment and management for digital content," "2 Provenance management for digital content," and "3 Linking digital content with related researchers and institutions." Whereas functions 2 and 3 can be easily derived from the system requirement, function 1 seems to be less related to the minimum requirement "Trackable." However, to ensure that digital content is "Trackable," there needs to be a persistent identifier assigned to the content. Also, considering that researchers tend to move across institutions, there needs to be a management function for the persistent identifiers when digital content is transferred to another institution.

B. Requirements Analysis of Functions for RDM-PSM System Using RDM-PSM Workflow Model

The functions for an RDM-PSM system laid out in Table 3 were scrutinized using a workflow model for RDM-PSM at academic institutions.

The assumed workflow for RDM-PSM at academic institutions is depicted in Figure 1. This workflow illustrates the RDM-PSM stages step by step. The researcher conducts research, publishes the research output, archives the research materials, and at some point moves on to another institution or retires. The upper workflow is for the normal research activities. The lower workflow is for in case of an incident of scientific misconduct. The left part of the workflow shows the actions by researchers, whereas the right part shows the actions by academic institutions.

The functions derived from the system requirements in Table 3 are mapped onto the workflow. Since the time line is not taken into consideration in Table 3, Figure 1 enables us to identify when each function is called. The functions mapped onto the workflow in Figure 1 are described in mnemonic expressions by combining the minimum requirements in parentheses in Table 3 and the function number. For instance, the function on the top of Table 3 "Metadata retrieval" is described in Figure 1 as "Content Search 1." Hereafter, the mnemonic expressions of the functions are called function units.

All the function units in Table 3 could be mapped to Figure 1. This means that the blocks of the workflow cover all the minimum requirements for academic institutions stated in the SCJ guideline. Also, there was no overlapping in the function units across normal research activities and in case of scientific misconduct, except for "Institutional Verification 2." This means that in designing the details and user interface of each function unit, we only need to assume single settings, namely, normal research activities or in case of misconduct. Only "Institutional verification 2" is used both in normal research activities and in case of scientific misconduct. However, since this function unit is originally the function for "Retrieving research outputs and related digital content," which is quite basic, it can be assumed that the usage of the function is almost the same for normal research activities and for scientific misconduct.

The function units in Table 3 were derived and categorized according to the minimum requirement derived from the SCJ guideline element analysis. However, scrutinizing each block of the workflow in Figure 1, each block contains function units from different categories. This means that, for instance, the functions of "Content Search" needs to be called at different stages. Depending on the combination of function units, the appropriate user interface can be derived as a set of functions. The mapping of function units onto the workflow enables us to identify how each function unit is going to be used at different work stages, which ensures the firm implementation of the SCJ guideline.

It should be noted that all the function units attributed to "Institutional Management" in Table 3 are mapped as function units of academic institutions, whereas function units attributed to "Content Management" in Table 3 are mapped as function units of both researchers and academic institutions. The function units attributed to "Content Management," which are mapped to academic institutions, are "Persistent identifier assignment and management for digital content," "User-license declaration," "Metadata retrieval (including keyword search, navigation, filtering, and sorting function)," and "Recording and viewing research trail." These functions could not be derived just from the terms of the SCJ guideline. The procedures to analyze the SCJ guideline in terms of its elements and mapping the system requirements onto the workflow enabled us to clarify the functions needed for either the researcher or academic institution.

IV. Conclusion and Discussion

The basic functionalities, system requirements, and associated functions of an RDM-PSM system were derived and scrutinized using the SCJ guideline. Since the guideline only prescribes the requirements for research materials preservation and not the procedures for RDM-PSM, our requirement analysis, namely, the derivation of minimum requirements and basic functionalities of an RDM-PSM system and the formulation of the system requirements of associated functions, was a necessary step towards specifying the details of an RDM-PSM system. The system requirements and associated functions of an RDM-PSM system in Table 3 and the workflow of RDM-PSM, where the functions are mapped to different stages in Figure 1, form the basis for developing an RDM-PSM system. However, to implement an RDM-PSM system, further usability issues need to be worked out in cooperation with universities, as the system requirements and associated functions derived describe only the basic functions of an RDM-PSM system. It will also be worthwhile to work out metrics to measure the usability and effectiveness of such a system in advance. The metrics should include usability for institutions as well as

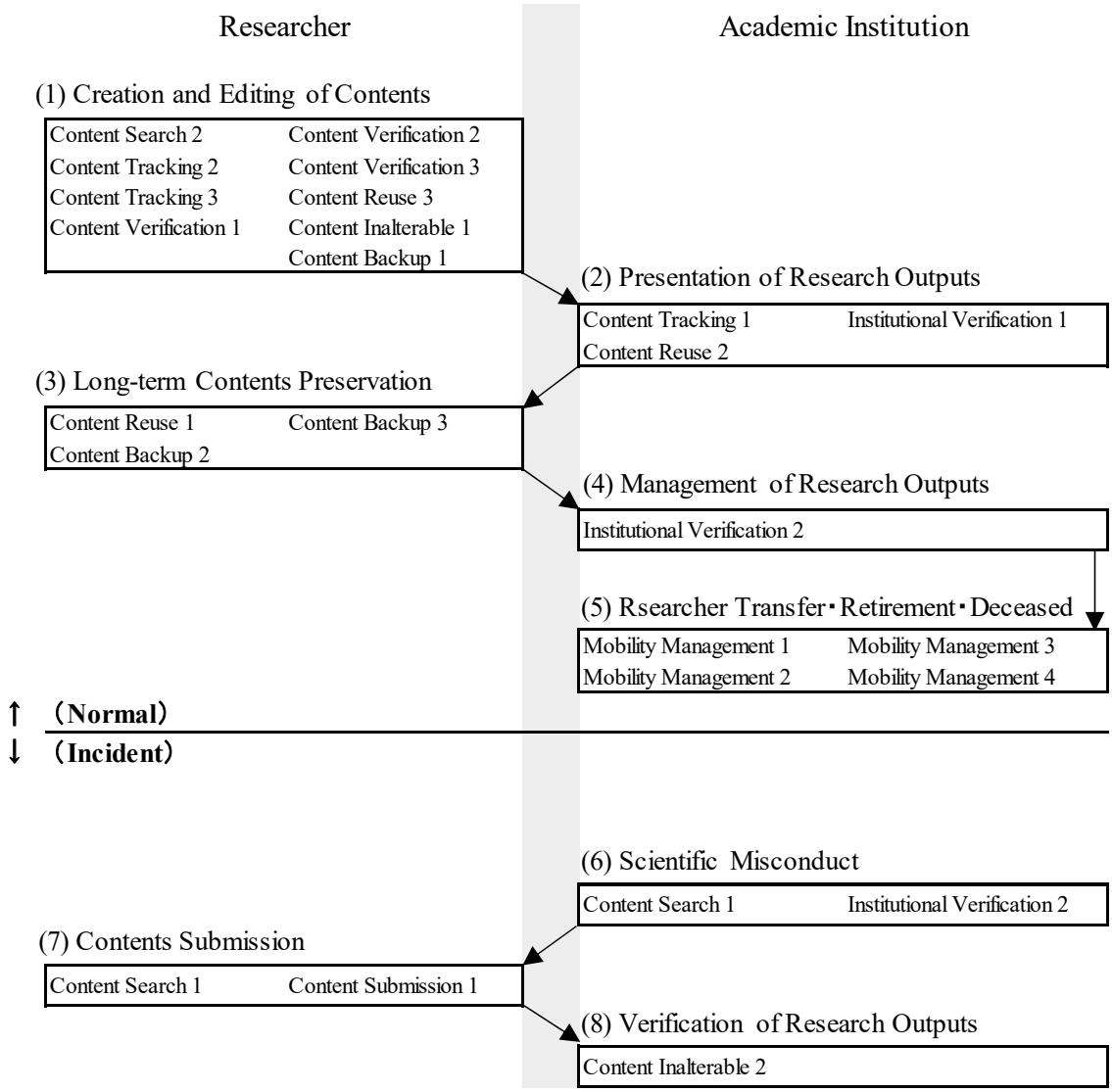

Figure 1. RDM-PSM workflow and mapped functions of RDM-PSM system

researchers and effectiveness for handling research misconduct as well as provision of a convenient research environment for normal research activities.

The procedure for formulating the system requirements and associated functions can also be used for RDM in Findable, Accessible, Interoperable, and Reusable (FAIR) Data Principles [20]. FAIR Data Principles are regarded as the underlining principles in the Open Science context, which is nowadays sought worldwide. Whereas RDM-PSM requires the research materials to be findable and accessible in case of scientific misconduct, FAIR RDM requires them in general and additionally requires the research data to be interoperable and reusable. The system requirements for FAIR RDM can be regarded as a derivation of an RDM-PSM system; thus, the procedure used to formulate and scrutinize the system requirements and functions for an RDM-PSM system can be applied to them.

We identified the basic functionalities and system requirements of an RDM-PSM system, which is urgently needed at Japanese academic institutions. Even though RDM for preventing scientific misconduct may be a unique concept in Japan, considering there is an increasing number of incidents of scientific misconduct and retraction of research articles worldwide, an RDM-PSM system could be used other countries. Since a FAIR RDM system will be a derivation of an RDM-PSM system, the procedures for formulating the system requirements and functions can be used universally.

REFERENCES

[1] National Institutes of Health, “Data Sharing Policy and Implementation Guidance,” 2003.

[2] Committee for Scientific and Technological Policy, OECD, ”Declaration on Access to Research Data from Public Funding,” 2004.

[3] Jisc, “Juliet Statistics,” Sherpa Juliet Website <http://v2.sherpa.ac.uk/view/funder_visualisations/1.html>. last accessed 2018.3.11.

[4] Alice Meadows, “To Share or not to Share? That is the (Research Data) Question…,” Scholarly Kitchen, 2014. <https://scholarlykitchen.sspnet.org/2014/11/11/to-share-or-not-to-share-that-is-the-research-data-question/>. last accessed 2018.3.11.

[5] Fang FC, Steen RG, Casadevall A. “Misconduct accounts for the majority of retracted scientific publications,” Proc Natl Acad Sci USA, vol. 109(42), p.17028–33, 2012. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3479492/>

[6] Jocelyn Kaiser, “Former U.S. Research Fraud Chief Speaks Out on Resignation, 'Frustrations',” 2014. <http://www.sciencemag.org/news/2014/04/former-us-research-fraud-chief-speaks-out-resignation-frustrations>

[7] California Digital Library, “University of California Curation Center” <https://www.cdlib.org/services/uc3/index.html>, last accessed 2018.3.11.

[8] Purdue University Research Repository (PURR), “Advice and answers from the PURR team,” <https://purr.purdue.edu/help>. last accessed 2018.3.11.

[9] The University of Edinburgh, “Information Services: Research Data Service,” <https://www.ed.ac.uk/information-services/research-support/research-data-service>. last accessed 2018.3.11.

[10] Council for Science and Technology, “Guideline for Dealing with Scientific Misconduct,” 2006. < http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu12/houkoku/__icsFiles/afieldfile/2013/05/07/1213547_001.pdf>. last accessed 2018.3.11.

[11] Minister of Education, Culture, Sports, Science and Technology’s Decision, “Guideline for Dealing with Scientific Misconduct,” 2014. < http://www.mext.go.jp/b_menu/houdou/26/08/__icsFiles/afieldfile/2014/08/26/1351568_02_1.pdf>, last accessed 2018.3.11.

[12] Science Council of Japan, “Reply: For the Enhancement of Soundness of Scientific Research,” 2015. < http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-k150306.pdf>, last accessed 2018.3.11.

[13] Japan Science and Technology Agency, “Basic Principles at JST for the Promotion of Open Science of Research Outputs,” 2017. <http://www.jst.go.jp/pr/intro/openscience/>. last accessed 2018.3.11.

[14] Japan Agency for Medical Research and Development, “Japan Genomic Medicine Project: Data sharing Policy for the Realization of Genomic Medicine,” 2016. <https://www.amed.go.jp/content/000017356.pdf >

[15] Japan Society for the Promotion of Science, “Open Access Implementation Policy for the Funding Programs of the Japan Society for the Promotion of Science,” 2017. <https://www.jsps.go.jp/data/Open_access.pdf>. last accessed 2018.3.11.

[16] Ministry of Education, Culture, Sports, Science and Technology, “Model Case Study for the Promotion of Research Integrity at Research Institutions,” 2017, <http://www.mext.go.jp/a_menu/jinzai/fusei/1386957.htm>, last accessed 2018.3.11.

[17] Kyushu Institute of Technology, “Policy for the Prevention of Research Misconduct” 2014. < https://db.jimu.kyutech.ac.jp/cgi-bin/cbdb/db.cgi?Page=DBRecord&did=206&rid=246 >, last accessed 2018.3.11.

[18] Ministry of Education, Culture, Sports, Science and Technology, “F.Y. 2016 Survey of the Implementation of Guideline for Dealing with Research Misconduct,” 2015. < http://www.mext.go.jp/a_menu/jinzai/fusei/1368869.htm>, last accessed 2018.3.11.

[19] The University of Tokyo, “Policy for the Preservation of Research Materials,” 2016. < https://www.u-tokyo.ac.jp/gen01/reiki_int/reiki_honbun/au07410541.html >, last accessed 2018.3.11.

[20] FORCE 11, “FAIR Data Principles,” < https://www.force11.org/group/fairgroup/fairprinciples>, last accessed 2018.3.11.