Recommendations for Research Data Management at Academic Institutions

AXIES 大学ICT推進協議会

May 1, 2019

Introduction

Chairperson of the Academic eXchange for
Information Environment and Strategy
Masao Kitano

Open Science and research data management have been attracting attention as new academic movements in recent years. These movements are attempts to create new value out of vast amounts of data that are now generated, used, shared, and distributed in any academic disciplines owing to advances in digitization and widely available internet access. The Academic eXchange for Information Environment and Strategy (AXIES) believes that its member institutions should have a shared understanding of these new tides and become actively involved in fostering an environment for research data management and that in doing so, Japan's academic research will be at the forefront of the contemporary movements and eventually positioned to lead the world in research data management. This was why AXIES established the Research Data Management Subcommittee (RDM Subcommittee) in the summer of 2017.

The potential of the digital era no doubt lies in making full use of information sharing and data connection, which means that a certain level of standardization is advisable across research environments. Researchers have managed their research data thus far, including preparing data storage. Once academic institutions begin to provide the research environments that free their researchers from concerns about data security and backup, the researchers can fully focus on their research activities. Furthermore, standardized infrastructure provided by academic institutions will make information sharing and data connection remarkably easy. This will also help ensure the accountability of the institutions, which society demands today.

An academic institution needs to ensure greater coordination between their offices in order to establish an environment for research data management in such a way that will improve its research capacity, raise its profile, and work well with society. The institution's IT center should create and manage research data infrastructure, researchers and research support staff should use this infrastructure, the office that supports research projects should manage the administrative part of the research, the university library should engage in the distribution of academic information, and the units for public relations and industry-academia collaboration should work to establish external partnerships. A new

challenge that needed multiple stakeholders like this used to have the risk of being passed over in the process of establishing a research environment in an institution. This is why the RDM Subcommittee stated at the meeting in the spring of 2018 that it should require documents describing the need for research data management within an academic institution, along with a conceptual image of research data connection. Then these Recommendations were planned and put forward.

The Recommendations were produced for universities and other institutions of higher education that make up the majority of the AXIES members. The Cabinet Office has published the Guideline for Establishing Data Policy at National Research and Development Agencies, which demands that research institutions set their own guidelines on the retention, management, and use of their research data. On the other hand, the Cabinet Office has only issued a policy of the "ten-year research data retention rule" for institutions of higher learning in order to prevent research misconduct. No policy has been put forward to increase the use of research data as well as research capacity. These Recommendations present multiple points of view on research data management at institutions of higher education and then express the view that each academic institution needs a data policy that is based on its own strategies.

Effective research data management that uses information and communication technology (ICT) is the key to value creation in a new paradigm of research. I hope that these Recommendations help lead each academic institution in Japan to provide a research data management framework that allows its researchers to work comfortably on their research with a sense of security, thereby creating a new paradigm for Japan's academic research.

May 1, 2019

Recommendations for Research Data Management at Academic Institutions

Academic eXchange for Information Environment and Strategy

Table of Contents

Recommendations for Research Data Management at Academic Institutions

Advances in digitization have further necessitated proper research data management in each stage of research. Research data frequently form the basis of research arguments in any fields. This has created the need to properly retain and manage research data in each stage of a research process in order to ensure that research outputs are formed on top of former research achievements and become reproducible. The internet has also created more opportunities for collaboration than ever, which also necessitates a particularly high level of research data management. Researchers now use data generated by other researchers, while sharing their own data and making the data openly available for reuse by others. When research data are reused, information about the source and version of the data is vital. With research data growing in size and types, it is not an overstatement to say that researchers who have control over the use and management of research data will propel in research in the digital era.

To ensure that research data are properly managed throughout the research process and that these data are shared and made open after the completion of the project, information infrastructure for stable and sustainable operation is essential. However, it would be too costly and inefficient for each researcher to prepare and, above all, maintain such infrastructure. Given that academic institutions perform academic work for society, it is these institutions' duty to create an environment for proper retention and management of research data and to make the data available to those who seek the opportunities to use them, while ensuring the efficiency of research by their researchers. Research data must be retained as invaluable academic resources for humanity and used to facilitate the advancement of humanity. Excellent research data, coupled with an environment for efficient use thereof, serve as the proof of an institution's academic excellence, which further attracts the attention of talented researchers and outstanding organizations. Furthermore, actively sharing research data and made them openly available eventually meets the demand for "open scholarship" in the digital era.

Academic institutions should use a common design to create an environment for efficient research data management, considering the need for data connection and the possibility of researchers moving to different institutions. Academic institutions are advised to work together to establish a standardized infrastructure in Japan as the foundation of their research data management, so that they will build and develop their own academic work upon it. The Academic eXchange for Information Environment and Strategy (AXIES) aims to "bring about advancement to education, research, and management at institutions of higher education and academic research by applying ICT, thereby contributing to education, academic research, culture, and industry in Japan." The AXIES advances the establishment of standardized infrastructure for research data management in Japan, while helping academic institutions in the country work together and develop their own environments.

The AXIES, a syndicate of academic institutions that promote the use of ICT, has drawn up these

Recommendations in hopes that institutions of higher education and academic research in Japan will adapt to the digital era, thereby connecting to the world beyond Japan to provide academic knowledge that is meant to enrich humanity. The Recommendations state the ideas and positions that should be noted as academic institutions manage their research data in the digital era. Research data management requires coordination and cooperation between various stakeholders within each institution. Hence, we hope that the Recommendations will be consulted by Chief Information Officers (CIOs), the administrations, and offices of the institutions of higher education and academic research in the AXIES community as they develop their research data management frameworks.

[Various Aspects on Research Data Management at Academic Institutions]

Actions/Objectives
I. Responsibilities of an academic institution in its role
a. Defining its own purpose of research data management as an academic institution
b. Exploring, developing, and operating the research data management system adopted by the university as an academic institution
c. Establishing and providing a digital platform for research data management
d. Announcing that the institution's research data management system is available and encouraging internal and external use of the system
II. Practical side of research data management introduced to an academic institution
a. "Open scholarship" delivered through dissemination of each institution's academic research outputs and cooperation with society
b. Increased visibility of the academic institution gained through publicity for its research projects and the promotion of external partnerships
c. Enhanced research capacity through the establishment of research infrastructure for researchers within the academic institution
d. Meeting requests from research funders or international academic journals regarding open research data sharing
e. Compliance with the ten-year research data retention rule by retaining and managing data as supporting evidence for academic research outputs
f. Greater protection provided to researchers and institutional compliance ensured through retention and management of research trails
g. Academic work passed to the next generation through long-term preservation and availability of research data for use
h. Students' data skills enhanced through the educational use of research data
III. Research data managed and provided by an academic institution and the use thereof
a. Data for management within the institution as supporting evidence for published research outputs in order to prevent research misconduct (institution)
b. Data for confirmation and use as supporting evidence for published research outputs (international academic journals, referees, researchers, etc.)
c. Research data from publicly funded projects for confirmation and use (researchers, the government, and the public)
d. Research data profile ("metadata") for search purposes and use (researchers, the government, and the public)
e. Databases or data archives for search purposes and use (researchers and the public)

f.	Research data, including long tails, made available for viewing and use (researchers and the public)
g.	Evaluation of research capacity based on research data generated within the institution (institution, research funders, the government, etc.)

Methods/Features/Maintenance

IV. What is needed to establish research data management at an academic institution

    a.   Clearly defined purposes of research data management at the academic institution and rules thereof

    b.   A complete digital platform for research data management

    c.   Clearly named stakeholders within the institution and a system for coordination between them

    d.   A well-defined position on how the internal cost is covered to maintain the research data management system, and stable financial resources

    e.   Established rules about research data retention and destruction, and trained experts for decision-making

    f.   Established rules and procedures regarding the institution's positions on rights in research data

    g.   Established procedures for handling research data that need to be kept confidential

V. System for research data management at an academic institution

    a.   Specifying the institution's offices involved in research data management and establishing a system for the management (planning and research offices)

    b.   Providing a digital platform for use to manage research data (IT center)

    c.   Announcing that the institution's research data management system is available and encouraging the use of the system (planning and research offices)

    d.   Engaging in research activities, and retaining and managing research data (researchers and research support staff)

    e.   Registering research data for institutional reporting or data sharing purposes, and adding explanatory information (researchers and research support staff)

    f.   Managing data as supporting evidence for published research outputs and handling any case of research misconduct (a research integrity office)

    g.   Checking research data to be made available open and adding metadata as necessary (university libraries, URA, and research officeresearch office)

    h.   Publicizing and offering research data for use, and promoting external partnerships (public relations, industry-academia collaboration with local communities, university libraries, and research officeresearch office)

    i.   Retrieving and using research data (students, researchers outside the institution,

companies, and society)

j. Assessing the institution's research capacity using research data (IR, URA, research office, research funders, and the government, etc.)

VI. Functional requirements of a digital platform for research data management

a. Functions to register and manage data as supporting evidence that are connected to published research outputs

b. Functions to retain research data and relevant documents and to manage research trails

c. Functions to authenticate institutional users and to control access

d. Functions to introduce a system of persistent identifiers for users and research data, among others

e. Functions to connect with a research data analysis tool and other external tools

f. Functions to save an analysis tool and an execution environment

g. Functions to make research data and relevant documents openly available and to add metadata

h. Search function that connects research data and relevant information

i. Functions to customize a digital platform for the institution and to manage the use of the platform

VII. Research data management personnel training

a. Basics of research data management (graduate students, researchers, research support staff, institution's office in charge, etc.)

b. Hands-on training for advanced research data management in each specialized field (graduate students, researchers, research support staff)

c. Technical knowledge and skills of research data management (research support staff)

d. Basics of data science for research data analysis and processing (undergraduate students, graduate students, researchers, and research support staff)

e. Hands-on training in data science for research data analysis and processing (graduate students, researchers, and research support staff)

[Recommendations for Research Data Management at Academic Institutions (appendix)]

(Actions/Objectives)

    I. Responsibilities of an academic institution in its role

    II. Practical side of research data management introduced to an academic institution

    III. Research data managed and provided by an academic institution and the use thereof

(Methods/Features/Maintenance)

    IV. What is needed to establish research data management at an academic institution

    V. System for research data management at an academic institution

    VI. Functional requirements of a digital platform for research data management

    VII. Research data management personnel training

Actions/Objectives

I. Responsibilities of an academic institution in its role

a. Defining its own purpose of research data management as an academic institution
b. Exploring, developing, and operating the research data management system adopted by the university as an academic institution
c. Establishing and providing a digital platform for research data management
d. Announcing that the institution's research data management system is available and encouraging the use of the system inside and outside the institution

Research data have been managed by researchers and always will be. Researchers manage the research data that they handle during their own research activities in ways that are most appropriate for their research.

Researchers have conventionally prepared their own environments for managing, retaining, sharing, and publishing research data. This means that each environment created by a researcher has been different than the others and so has the level thereof. When researchers from different institutions work together, each of them is also required to prepare his/her own means of communication and use of data for collaboration. This creates a significant workload and makes the process of the collaborative efforts extremely inefficient. A digital platform as a basic tool provided by academic institutions for their researchers eliminates this workload and boosts research productivity. A standardized digital platform installed across institutions also facilitates communication and coordinated joint efforts between researchers [1][2].

It should also be noted that only academic institutions are entitled to retain, manage, share, and publish research data generated by their researchers even after the researchers are transferred to different institutions or retire from research activities. Given that many research data are generated in publicly funded projects, it is the duty of academia to facilitate open access to the data for use in order to maximize the benefits of the data for humanity [3][4][5]. Moreover, academic institutions are required to retain their research data for a certain period in order to prevent research misconduct [6][7]. That is, as a rule, when a researcher generates data, it is the institution where the researcher works that must manage the data and handle any case of research misconduct even after the researcher moves to a different workplace or retires.

Hence, academic institutions are expected to provide a digital platform as research infrastructure on which they will operate their research data management systems. Note that each of these research data management systems established by academic institutions differs from the others. This is because each

platform, along with the operating policy thereon, is designed according to why research data are managed. That is, an institution may have its data management system to provide its researchers with an environment that facilitates efficient research activities, thereby strengthening the institution's research capacity; to boost the dissemination of research data for stronger partnerships with external entities; or to become capable of handling cases of research misconduct that require the institutions to take actions, no matter how minor they might be. For this reason, when an academic institution creates a research data management system within it, it should specify the purpose and design the system and platform accordingly before introducing the infrastructure.

Each academic institution must disseminate how its research data management system works and also promote the use of the system. Researchers, who are requested to manage and share their research data by research funders and/or international academic journals, need to be aware of the services offered within their institutions in order to meet those requests [8][9][10]. If an institution provides research infrastructure for the use by its researchers during their research, it is advisable that the institution can emphasize, for example, how securely their research data can be stored on the common platform, how easily the data can be shared with their co-researchers, how the data can be connected to an external analysis tool, and other features that help improve research efficiency. The institution should also encourage its researchers to share their research data and make these data open, so that the researchers will feel positive about making their data externally available. Actively sharing excellent research data and making the data open will lead to boost the academic institution's research capacity and raise the institution's profile, attracting the attention of more talented researchers and outstanding organizations, and eventually to advance the overall research at the institution [11][12]. To achieve this, each institution should inform external entities about the research data it has made available and facilitate the use of the data as well as coordination between the entities and its researchers.

II. Practical side of research data management introduced to an academic institution

a. "Open scholarship" delivered through dissemination of each institution's academic research outputs and cooperation with society

b. Increased visibility of the academic institution gained through publicity for its research projects and the promotion of external partnerships

c. Enhanced research capacity through the establishment of research infrastructure for researchers within the academic institution

d. Meeting requests from research funders or international academic journals regarding open research data sharing

e. Compliance with the ten-year research data retention rule by retaining and managing data as supporting evidence for academic research outputs

f. Greater protection provided to researchers and institutional compliance ensured through retention and management of research trails

g. Academic work passed to the next generation through long-term preservation and availability of research data for use

h. Students' data skills enhanced through the educational use of research data

The movement toward Open Science around the world has clarified the idea that research data should be managed and made open for use [4][13][14][15][16]. This is the position that there should be "open scholarship" by making research data as well as academic papers available to anyone outside the institution that manages the data [17][18]. The position has gained momentum as the digital era has made academic research outputs, including research data, more retainable and shareable than ever, and as the world today increasingly expects academia to take up challenges that will create a social impact. Academic institutions fulfill their societal duties when they disseminate their academic research outputs, including data, to work and cooperate with society.

Publicly announcing what research data they internally retain also helps the institutions raise their profiles and gain a competitive edge. Their profiles will be further enhanced if the institutions actively support the reuse of their research data, along with partnerships between their researchers and eternal organizations through the reuse. These efforts also attract the attention of many talented researchers to the institutions.

Furthermore, the digital platforms the institutions provide to their researchers for efficient research activities will eventually strengthen their research capacity [11][12]. Researchers need an environment in which their research data are securely retained and shared with co-researchers and in which the versions of the files are automatically managed to ensure that research trails are left. They also need to be connected to external analysis tools and lab facilities. An excellent research environment attracts

the attention of talented researchers and outstanding organizations, serving as a base of research excellence.

There are also some factors that make research data management mandatory for academic institutions. One is that research funders around the world demand that publicly funded research data be shared and open. Making publicly funded research outputs openly available is expected to help avoid redundant investments and accelerate the research [3][4]. For this reason, researchers applying for a research grant are required to submit a research data management plan (DMP). A DMP should state a policy on sharing of research data, along with where the data will be retained when they are shared [8][9]. It is academic institutions' duty to provide their researchers with an environment for retaining and sharing research data in an essentially reliable and permanent repository, so that the researchers can apply for any grants they need. In recent years, international academic journals also demand that academic institutions take this action in order to ensure the reproducibility of academic papers [10]. Academic institutions should make sure that, while encouraging their researchers to contribute papers to these journals, research data acquired in Japan will not be leaked outside the country for national security reasons.

In Japan, it is almost mandatory for academic institutions to manage their research data in order to prevent research misconduct, rather than to join the movement toward Open Science [6][7]. These institutions are required to retain data that serve as supporting evidence for research outcomes to be published for about ten years. This requirement must be met even after the researcher who has generated the data move to a different institution or retire. In this context, the institutions do not routinely view or use the research data. Nevertheless, since the institutions must promptly check research data and any relevant information if research misconduct is committed, they should manage their research data coupled with documents related to published research outputs acquired using the data, research project members, the facilities and analysis tool used, and other relevant information.

In the context of the prevention of research misconduct, only data that serve as supporting evidence for published research outputs are required. However, to verify the reproducibility of research, a dataset that contains a wider range of data (e.g., raw data; processed data in each stage of research) may be needed. Hence, if research misconduct occurs in a routine research environment provided by a researcher, it is reasonable to use the research trail saved in this environment. Researchers would also find it easier to have their research trails automatically recorded in their routine research environment than to retrieve only the data that serve as supporting evidence for published research outputs and report them to their institutions. When academic institutions provide research infrastructure their researchers routinely use in order to save and manage their research trails, they also provide protection for their researchers and further ensure their institutional compliance.

Academic institutions are also increasingly expected to treat their research data as important academic assets for humanity and preserve the data for posterity. People have long retained academic research

outputs for generations only in the form of literature (e.g., academic papers and books) or physical materials archived in museums. The digital era has enabled numerous academic research outputs to be saved as digital data, dramatically expanding the range of research outputs that can be preserved for posterity. These digital data include not only academic papers and research data, but the digital environment and the analysis tool that have created research data, along with physical documents converted into digital form. If the digital data are not stored in a field-specific repository, the academic institution that manages the data should sift through the digital content that is mass-generated each day for long-term preservation for posterity.

Given that the digital era has tremendously increased the opportunities to use digital data, university graduates are fully expected to have skills needed to use those data. Academic institutions will be able to meet this need of society by using the research data, which they retain and manage, in their education. The use of research data this way also serves as practical training in academic research for students.

III. Research data managed and provided by an academic institution and the use thereof

> a. Data for management within the institution as supporting evidence for published research outputs in order to prevent research misconduct (institution)
>
> b. Data for confirmation and use as supporting evidence for published research outputs (international academic journals, referees, researchers, etc.)
>
> c. Research data from publicly funded projects for confirmation and use (researchers, the government, and the public)
>
> d. Research data profile ("metadata") for search purposes and use (researchers, the government, and the public)
>
> e. Databases or data archives for search purposes and use (researchers and the public)
>
> f. Research data, including long tails, made available for viewing and use (researchers and the public)
>
> g. Evaluation of research capacity based on research data generated within the institution (institution, research funders, the government, etc.)

Research data are used on various occasions according to the purpose of research data management specified by each academic institution.

If an institution adopts the minimum research data management, it only manages data that serve as supporting evidence for published research outputs, along with information about the management, in order to prevent research misconduct [6][7]. In this case, research data are not required to be open, are managed only by a office in charge within the institution, and are used only when any case of research misconduct occurs. The institution may also publish data that are the same as the research data collected inside the institute in response to a request from an international academic journal [10]. Once such data are published, they are available for reference and use by the international academic journal and its referees of papers, as well as internal and external researchers and the public.

Making research data and any relevant information open raises the profile of an academic institution as a research entity, which may lead to partnerships with external organizations or individuals and to the institution's greater research capacity [11][12]. This also helps preserve research data generated in publicly funded projects for posterity as humanity's academic assets [3][4]. That said, it should be the institution's researchers who hold the right to publish their research outputs. Hence, instead of making the whole research data open, the institution may opt to have a certain period before making the data freely available (an embargo) or to publish only metadata about the research data. Then the institution may publicize the research activities and pursue opportunities to partner with external entities. Some researchers and research groups compile a well-organized database or data archive for publication and public use. The institution may also publicize such a database or archive if it has any. In addition to

all these options, long-tail datasets that researchers obtain for individual subjects may be made open through the institution's repository for public use.

Once an academic institution establishes an internal process for systematic and comprehensive organization of research data generated within the institution, it may use the process to evaluate its research capacity and its researchers. The institution may also consider forging research partnerships with other institutions that have similar processes or employing researchers from those institutions. When research data are used to assess research capacity, it is advised that the research data be linked to such information as published research outputs that use the data, related research projects funded by grants, and relevant researchers and research support staff. Any research data that are generated in publicly funded projects should be organized and managed by each institution and be available for evaluation by the research funders who directly sponsor those projects as well as by the government.

IV. What is needed to establish research data management at an academic institution

a. Clearly defined purposes of research data management at the academic institution and rules thereof

b. A complete digital platform for research data management

c. Clearly named stakeholders within the institution and a system for coordination between them

d. A well-defined position on how the internal cost is covered to maintain the research data management system, and stable financial resources

e. Established rules about research data retention and destruction, and trained experts for decision-making

f. Established rules and procedures regarding the institution's positions on rights in research data

g. Established procedures for handling research data that need to be kept confidential

To adopt a research data management system, an academic institution should define its own purpose of research data management in line with the realities surrounding the institution. Providing a digital platform alone would not compel researchers to use it, nor does it elicit much-needed support for smooth operation of the data management system from relevant units within the institution. For this reason, each academic institution should specify the purpose of research data management as a set of written rules (e.g., "Research Data Management Policy") [19].

Since a research data management system essentially involves multiple offices and people within an institution, opinions offered by prospective users of the system as well as relevant offices should be well incorporated as the institution works on its research data management policy. At the same time, discussions with stakeholders and efforts to achieve a consensus should continue not only during the policy development process, but throughout the stages of operation and review of the system (multi-stakeholder approach [20]). Involving multiple offices may obscure where responsibility lies. Hence, it is vital to clarify who is responsible for the establishment of the research data management system and who is in charge of promoting the system and coordinating with the relevant offices.

Academic institutions should also be aware that financial backing is needed to create a digital platform for research data management and to develop and maintain a framework to operate the platform. How to cover the cost of a repository for research data particularly requires detailed discussions. Volumes of research data considerably vary by discipline, research subject, and research method. Hence, academic institutions may consider introducing a variable fee system in which researchers pay for any

extra storage space they need in addition to a certain capacity during their projects, instead of providing the space that accommodates all of their data. It should also be noted, however, that the institutions will essentially be unable to have their researchers continue to directly cover the cost of long-term retention and publication of research data after the projects end or the researchers move to different organizations or retire. In other words, each institution should establish its own sustainable cost model to ensure that research data are retained and managed during and after research and that the data are shared, made open, and reused.

It is also vital to set standards for research data to be discarded or preserved for a long term after a certain period of retention, so that research data are not accumulated without limit. In addition, it is necessary to train experts so that they can make responsible decisions according to those standards. Archivists who conventionally manage academic resources in libraries or museums may be suitable for this role of experts. There are also supporting data associated with published academic research outputs [7] to consider, since institutions must retain and manage such data for ten years in order to prevent research misconduct. Each institution needs to determine whether to manage those data as part of research data for long-term preservation as academic assets or to manage them separately, according to its position on research data management.

To make research data generated within an academic institution available for use while maintaining certain control of the data for long-term preservation, an institution needs to reach an agreement with the creator of the data in advance on who owns the research data and on rights in those data and make arrangements accordingly. Given that data are pivotal to researchers' research activities, the researchers or research group that generate data should be able to use the data whenever they need. The data should remain available to these researchers even after they move to different organizations. On the other hand, it should be noted that many research data are generated in publicly funded projects, and thus other researchers, companies, and the public also have the right to benefit from the data [3][4][21]. For example, some data can be more useful when they are organized into a database or combined with other data, and these data should especially be shared and made open. If researchers decide to withdraw from academic research activities because they move to a different workplace or retire, their appropriate action would be to leave the research data generated through their projects in the hands of their academic institutions, so that the data will be preserved for use by posterity. As for research data that are generated according to a research agreement between an academic institution and a research funder, the government, or an enterprise, among others, the rights specified in the written agreement prevail. Means to reserve researchers' rights, such as having an embargo period, should also be considered in order to ensure that researchers will fully benefit from the data they have generated.

When an academic institution provides a system for sharing and publishing research data, the institution should also establish a procedure for handling research data that must remain confidential.

These research data include those that contain sensitive information such as personal data, those that are partly owned and/or licensed by an external entity, and those that must be carefully handled for a national security purpose [22]. Researchers often have difficulty determining whether the research data they have created qualify as confidential data that fall under any of these categories. Therefore, academic institutions should set accurate standards for confidential data and establish their own process to review and manage potentially sensitive data, while essentially having their researchers determine the confidentiality of their data according to the standards. Some confidential data, on the other hand, may be offered for external use if they are anonymized, for example. Hence, academic institutions are advised to set their own rules about how to process data in order to establish a system to provide technically processed data.

V. System for research data management at an academic institution

a. Specifying the institution's offices involved in research data management and establishing a system for the management (planning and research officeresearch offices)

b. Providing a digital platform for use to manage research data (IT center)

c. Announcing that the institution's research data management system is available and encouraging the use of the system (planning and research offices)

d. Engaging in research activities, and retaining and managing research data (researchers and research support staff)

e. Registering research data for institutional reporting or data sharing purposes, and adding explanatory information (researchers and research support staff)

f. Managing data as supporting evidence for published research outputs and handling any case of research misconduct (a research integrity office)

g. Checking research data to be made available open and adding metadata as necessary (university libraries, URA, and research office)

h. Publicizing and offering research data for use, and promoting external partnerships (public relations, industry-academia collaboration with local communities, university libraries, and research office)

i. Retrieving and using research data (students, researchers outside the institution, companies, and society)

j. Assessing the institution's research capacity using research data (IR, URA, research office, research funders, and the government, etc.)

Once generated by researchers at an institution, research data are handled by multiple offices before they are managed by the institution. Hence, it is vital to specify in advance the offices that will be involved in the workflow of these procedures for research data management in order to facilitate research data management at the institution. This is done by the institution's planning and research offices. The following paragraphs list the offices that will likely be involved in the operation of an institution's research data management system and their roles.

First, an institution must install an internal digital platform for research data management. The institution's IT center or any unit that is equivalent to it is in charge of the installation. If the institution uses a standard inter-institutional platform, the unit liaises with the National Institute of Informatics and other relevant organizations. Settings are configured for each organization, and the institution customizes the functions it needs. During this process, opinions about internal needs are collected within the institution.

Each institution's planning and research offices establish the purpose of research data management at

the institution, along with a research data management system designed to meet the purpose, and then announce the launch of research data management across the institution, urging everyone to use the system. Researchers use the research data management platform, following the guidance. If the platform is provided as research infrastructure, researchers routinely use it to save research data and other related files, share files with their co-researchers, and analyze and process data. Files on the platform are regularly backed up, and their versions are managed, thereby enabling researchers to leave research trails without conscious effort. This is confirmed by the institution's research integrity office should any case of research misconduct arise.

When they publish research outputs or request research data to be open, researchers register the outputs or data in a designated location on the digital platform. The metadata and explanatory information about the research data should also be added, so that the data can be found and used by third parties. The addition may be done by research support staff. When these research data are already scheduled to be open, the institution's library or any relevant unit checks the data, and prepares and publishes metadata about the research data as necessary [23]. In regard to data as supporting evidence that accompany the publication of research outputs, the research integrity office manages the registration of the data and checks back on it should any research misconduct be committed.

Open research data are found by a search engine for the use of researchers and companies outside the institution, and anyone who needs them. Teachers and students may use the data as part of education. If the institution plans to demonstrate particular data to the public, it may actively publicize and use them to partner with external entities. In this case, the institution's offices for public relations, industry-academia collaboration with local communities, and research promotion work on the process. Since some research data may need to be accompanied by an extra explanation or be anonymized, the institution's library or any relevant unit may act as a coordinator to fine-tune the data and/or offer guidance on the use of the data.

Research data that are registered for each publication of research outputs eventually comprise an exhaustive dataset, which the institutions may use to assess its research capacity. In this case, IR[1], URA[2], and the research office work on the process. Research funders, the government, or other entities outside the institution may also use the dataset to assess the institution's capacity.

---

[1] IR: An institutional research staff/office at a university or an equivalent institution. IR collects and analyzes data and information in order to improve the institution's activities to help with the institution's decision-making. IR is classified by function into three types, namely IR for university management, education and learning, and research. The IR in these Recommendations refers exclusively to IR for research.
[2] URA: University research administrators, a position separate from that of full-time researchers. URAs plan and manage research activities and advance the use of research outputs, thereby helping invigorate researchers' activities and strengthen the management of research and development.

VI. Functional requirements of a digital platform for research data management

a. Functions to register and manage data as supporting evidence that are connected to published research outputs
b. Functions to retain research data and relevant documents and to manage research trails
c. Functions to authenticate institutional users and to control access
d. Functions to introduce a system of persistent identifiers for users and research data, among others
e. Functions to connect with a research data analysis tool and other external tools
f. Functions to save an analysis tool and an execution environment
g. Functions to make research data and relevant documents openly available and to add metadata
h. Search function that connects research data and relevant information
i. Functions to customize a digital platform for the institution and to manage the use of the platform

A digital platform for research data management is for the use of researchers, academic institutions, and external entities. It also serves as a medium for mutual collaboration between these users [1][2]. Functional requirements of a digital platform vary according to each institution's purpose of research data management. These requirements should also be regularly reviewed in order to keep pace with advances in digital technologies and to adapt to changing societal demands. Moreover, they must continue to be stably operated to ensure that academic information is protected and preserved for posterity.

If an institution manages research data for the purpose of preventing research misconduct, its platform needs to have a function to register the fact that certain research outputs have been published, coupled with any relevant documents, and a function that allows the institution to manage the details of the registered information [6][7]. These relevant documents to be registered may be academic papers, along with data as supporting evidence for these papers. Or users may only register where these data supporting the outputs can be found, instead of the data themselves. In this case, however, the institution should make sure that, to be prepared for any incident of research misconduct, it always has access to the research data and relevant documents.

An investigation into research misconduct, if any is committed, requires not only the data supporting the published research outputs in question, but the history of all that has been done with the research data, from generation to instances of processing. For this reason, it is advised that, going forward, each institution provide a digital platform to its researchers as a place for their routine research activities, so that all actions taken to generate, edit, save, or delete files on the platform will automatically be

recorded for the purpose of research trail management. Providing research infrastructure that their researchers routinely use will lead to efficient research activities and eventually to the institution's greater research capacity. Needless to say, the institution should not take advantage of the digital platform it provides to watch and control its researchers' activities more than necessary.

To make its digital platform a viable place for researchers' daily routine, an institution must ensure that the platform has essential functions, namely those to set an access level for each user and to share data only with specific internal and external co-researchers. It is also necessary to have a system for functional authentication and permission in order to control access by users. Given that a digital platform has external users as well, each institution should also make adjustments to its information security policy with those users in mind. Moreover, the institution must consider introducing a system of persistent identifiers for users and research data, among others, to ensure that research trails are managed.

Researchers' activities will be even more efficient when the digital platform provided by the researchers' institutions can connect to internal and external databases and measuring devices and to a supercomputer or any extensive analysis platform. Since this type of digital platform acts as a medium throughout the process of the input, analysis, and output of research data, it contributes to even more sophisticated recording of research trails. It is also advised that, to ensure the reusability and reproducibility of research data, the platform is designed to save research data coupled with a tool to analyze them and an execution environment for the tool.

In the context of Open Science, research data fully serve their purpose only when they are found by a third party to be reused, or used for joint research. Given that research data can be found through metadata about them (e.g., keywords, the generator of the data, the date/location of generation) [23], a digital platform for open research data should be designed in the way that enables the addition of these metadata. The platform should also be designed to authorize users other than researchers to edit because research support staff, instead of researchers, may add metadata. Research data may contain something confidential such as personal information, in which case only metadata about the research data, instead of these data themselves, may be published. Research data should be openly available not just for external use, but to meet requests from international academic journals for data as supporting evidence for posted papers [10].

Open research data can be found by a search engine. To improve the accuracy of a search, it is advisable to make research data searchable by not only standard metadata, but information related to individual research dataset. For example, information such as academic papers that use the research data, the name of the researcher who has generated the data and the institution where he/she works, and the research funder and the project may be provided in connection with the research data, so that the data can be found by a full-text search.

All these functions of a digital platform must be customizable for each institution according to its

purpose of research data management. Furthermore, checking how the platform is used ensures proper operation.

Functions of a digital platform for research data management lead to partnerships between researchers across different institutions. Hence, cross-institutional standardization and the development and provision of common institutional infrastructure are encouraged. It is also advisable to create an environment that suits any institutions regardless of their size and purpose of research data management. Digital platforms for research data management are still at a developmental stage around the world. It is hoped that research institutes, including the National Institute of Informatics, will continue to advance the use of these platforms, taking account of academic needs. The establishment of these platforms as Japan's research data infrastructure will serve as an efficient way to urge institutions across the country to adopt a research data management system, thereby giving Japan a greater competitive edge.

VII. Research data management personnel training

> a. Basics of research data management (graduate students, researchers, research support staff, institution's office in charge, etc.)
> b. Hands-on training for advanced research data management in each specialized field (graduate students, researchers, research support staff)
> c. Technical knowledge and skills of research data management (research support staff)
> d. Basics of data science for research data analysis and processing (undergraduate students, graduate students, researchers, and research support staff)
> e. Hands-on training in data science for research data analysis and processing (graduate students, researchers, and research support staff)

Research data have conventionally been managed by researchers, and each researcher and each research field have used a different way of data management. Advances in digitization in recent years have enabled large volumes of research data to be retrieved with increasing frequency, necessitating advanced skills in saving, managing, and analyzing research data. It has also created more opportunities to collaborate through research data, which is accompanied by the challenges of managing versions of those data and of properly retaining and managing the data after projects end. Moreover, international academic journals, research funders, and the public demand that research data be open and shared to ensure the reproducibility and transparency of research. This has created the worldwide movement toward open research data that are generated and processed under proper management, accompanied by explanatory information. Given that research data are now shared more frequently during and after projects, it is advised that research data be retained, managed, shared, and made open using a standardized means.

It is hoped that all members of each academic institution will be aware of how research data should be managed in this digital era and follows it as a basic institutional practice [24][25][26]. It would be ideal if the members understand why research data should be managed that way in this age and feel motivated to work together toward "open scholarship" in the digital era, rather than just adopt the practice. Moreover, through graduate students entering the workforce, the idea and practice of societally workable research data management should become norms. It should also be expected that academic work will move forward hand in hand with society.

Research data management also requires technical knowledge and skills that are specific to an area of expertise or to research data management. Some areas of expertise have developed their practices over years, while others have not. In the areas that already practice research data management, their ways and techniques of managing data should continue to evolve in line with progress in infrastructure so that they will be commonly used. As for the areas that have yet to develop their ways of research data

management, it is advised that the academic community in each of these areas work out a proper approach to research data management that will be shared across the community, so that the area will continue to grow and its academic assets will be preserved for posterity [27].

In regard to knowledge and skills specific to research data management, research support staff members assigned to the acquisition thereof will likely work as data curators or other equivalent professionals. It is envisioned that URA, technical staff, assistant professors, and assistants working on the ground or somewhere close to it will manage research data during a project and that librarians who have conventionally managed academic research outputs will manage research data to be shared and made open. Ideally, during a project, research data should be structured and saved, and the steps taken to process research data should be described, according to certain common rules. Research data should be accompanied by metadata (e.g., when and where the data were generated by whom; the format of the data) so that they can be easily findable and reusable. Furthermore, personal information should be properly anonymized and confidential information should be handled according to the rules set by each institution. Research data are academic research outputs in the digital era. This means they must be managed with the care that have been exercised to manage print academic papers and books (i.e., academic research outputs in the hard-copy age) so that they will be retained for generations [28]. Research data fully serve their purpose when they are used as well as managed. With research data increasingly becoming extensive and complex in recent years, research data processing and analysis today requires certain skills and knowledge that include those of programming, statistical processing, and machine learning. Countries around the world now urgently need trained data scientists. The skills and knowledge required of a data scientists range from basics taught in undergraduate programs to professional expertise needed for on-the-ground research and development. Hence, it is advised that each academic institution systematically train its data scientists. The training should be provided not only as part of undergraduate and graduate curricula, but as training sessions, seminars, and workshops for faculty and staff as well as students who need to acquire the new skill set. These training programs should not be one-offs. They should be provided as a set of modules for a particular curriculum designed to develop their skills over time. It would also be ideal if those who meet skill standards can be certified.

[Definitions of Terms]

○ **Academic institutions**

Institutions of higher education or academic research. The Recommendations are intended for these institutions and refer to them as "academic institutions" or "institutions."

○ **Multi-stakeholder approach**

In the multi-stakeholder approach, three or more stakeholders meet and discuss matters on an equal footing in order to reach consensus and resolve issues that cannot be settled by one stakeholder or between two stakeholders [20].

A research data management system at an academic institution requires the involvement of multiple stakeholders within the institution (e.g., the administrative office responsible for communication with officers in the university administration that determines the purpose of research data management; users of the system and those who support the use of the system; the unit in charge of the digital platform that supports infrastructure; the unit that assists with the operation of the institution's research data management system). This means these stakeholders need to agree on the direction of the institution's research data management so that they create a framework that enables them to fulfill their own roles.

○ **Research data**

Data that are generated in the process of research. These data include not only raw data as research materials, but data that are processed in the middle of research, analytical data that lead to the publication of research outputs, and data as supporting evidence for published research outputs, coupled with documents that explain these data. They may also include information related to a research environment designed to make the generation of research data reproducible (e.g., data acquisition requirements, source code). "Research data" may also be used as a collective term that refers to all types of materials created to accompany academic research activities (e.g., research outputs including presentations at academic conferences and research project evaluation forms; documents about research grants and various procurement related to the research; records of communication with co-researchers).

Research data may take all types of forms, including numbers, images, and text. These Recommendations are made on the assumption that research data management will likely make progress in the digital era, and thus the focus of the Recommendations is on data in digital format. Physical research materials and documents may also be included if there are digital data that explain them.

Each institution needs to define what qualifies as research data in accordance with the purpose and scope of research data management it sets out.

○ **Evidence data of research outputs**

Data that support the argument made in published research outputs (e.g., an academic paper) and ensure the reproducibility of the outputs. More and more international academic journals demand that supporting data be submitted with papers in order to ensure the reproducibility of research outputs. In Japan, institutions are obliged to retain these data for a certain period [6][7].

○ **Publicly-funded research outputs**

Research outputs from research activities funded by public research funders. These outputs include not only the final outputs such as academic papers, but various research data that are generated, processed, and analyzed during the research activities. These publicly funded outputs are expected to benefit society [3][4][5]. For this reason, many research funders around the world require or suggest that academic papers as publicly funded outputs be published. They also suggest that research data generated in the process of publicly funded research activities be shared and open, and they have begun to mandate the submission of a DMP [8][9].

○ **Research data profile (metadata)**

Information including who has generated the research data, the date and location of generation, the date of registration, where the data are retained, how to use the data, and any related academic papers. The technical term for it in research data management is "metadata." In Japan, the Japan Consortium for Open Access Repository (JPCOAR) has developed and published a new metadata schema titled "JPCOAR schema" designed for research data and international interoperability of repositories [23].

Given that metadata about research data are key to the findability and reuse of the research data, it is internationally acknowledged that metadata should follow the FAIR Data Principles [29][30]. "FAIR" stands for "Findable, Accessible, Interoperable, Reusable."

○ **Research data management (RDM) at academic institutions**

Research data are essentially managed by the researchers who generate, process, and analyze them. The academic institutions where these researchers belong provide an environment needed to manage their research data as many of these data are generated in publicly funded projects and thus they should be properly managed, shared, and preserved for posterity. Those data also give a clear picture of the institutions' research capacity, which will be publicly presented and enhanced [11][12]. More specifically, each institution defines its own purpose of research data management and then establishes a research data management policy, digital platform, and system and framework to manage research data accordingly.

○ **Open data and data sharing**

The term "open data" refers to research data that are made freely available online. "Data sharing," on the other hand, does not always mean making data freely shareable online. It refers to providing research data in some form for the use of anyone who needs them. Since some research data require confidentiality and/or need to be accompanied by an explanation as to how to use them, the idea of "data sharing" has come into being [8].

○ **Data management plan (DMP)**

In recent years, research funders have begun to ask for a DMP when an application for a research grant is submitted or accepted. A DMP should describe how data from the research project will be acquired, retained, and managed, along with whether the data will be shared and, if they will, how they will be used. Applicants may choose not to share any data from their projects, in which case they will often be asked to provide a legitimate reason [8][9].

○ **Embargo period**

A period of time after which research data are made open. Publicly funded research outputs are expected to be promptly made open for external use in the context of Open Science. Yet if research data are made openly available as soon as they are created, another researcher may use the data to write an academic paper before the researcher who has created the data does. If this happens, the researcher as the creator of the data may not be rewarded for the work he/she has put in to generate the data. This is why a certain period of time may be given between the creation of research data or the completion of a research project and the publication of the research data.

An embargo period may be set by a research funder, international academic journal, or the academic institution where researchers who create data work.

○ **Open Science**

The pursuit of a new form of academic research activities driven by advances in information and communications technology [13]. There is no official definition of Open Science. It may be explained as a collective term for open data, open codes, preprints, and electronic lab notebooks, among others [31]. The "open" in these terms does not indicate that these data and codes must be fully made available to the public. The needs to protect personal and/or confidential information, as well as to set an embargo period so that researchers who have created data will be the first to write a paper using the data, are duly noted. "Open Science" is not a term used to demand that academic work be open to the public. It implies that academic work should have dynamic connections with stakeholders around the world in order to make progress, as the digital era has

made academic information more shareable than ever and communication and collaboration with anyone around the world far easier than ever, and a demand for "academic work open to society" has been growing.

Science and technology policies across the globe state actions to take toward Open Science, which include 1) Providing open access to academic papers, 2) Making publicly funded research data open, and 3) Making any other academic work open, which includes ensuring open collaboration and open peer review [32]. Efforts toward the "open access to academic papers" stated in 1) have been made since the 1990s [33][34]. In recent years, the world has been increasingly focusing its efforts to make "research data openly available" stated in 2). In Japan, since the G8 Science and Technology Ministers' Meeting issued a joint statement about making publicly funded research data open in 2013 [4], the Cabinet Office [14], the Science Council of Japan [15], and other organizations have submitted reports regarding Open Science. Furthermore, once the Integrated Innovation Strategy defined "data infrastructure for open science" as one of the three key infrastructures for innovation in Japan [16], the country began to focus its efforts on research data management based on the open science policy. As for "making any other academic work open" stated in 3), certain efforts have been made toward open academic work in Europe and North America, whereas no specific initiatives for it are in progress as part of a national policy in Japan though it is included in the scope described in the country's policy document.

○ **Digital era**
Digital data in general refer to discretized data as opposed to analog data. The term "digital era" in the contexts of research data management in these Recommendations and of the advancement of open science refers to the age of advances in not only digitization but also the internet as a catalyst for dramatic progress in information sharing and global communication. In the English-speaking world, the digital era is also referred to as the "digitally-networked world."

[References]

[1] European Commission, "Realising the European Open Science Cloud" (2016)

[2] German Council for Scientific Information Infrastructures (RfII), "An International Comparison of the Development of Research Data Infrastructures. Report and Suggestions" (2017)

[3] OECD, "Declaration on Access to Research Data from Public Funding" (2004)

[4] G8 UK 2013, "G8 Science Ministers Statement London UK" (June 2013)

[5] OECD, "Making Open Science a Reality" (2015)

[6] Ministry of Education, Culture, Sports, Science and Technology, "Guidelines for Responding to Misconduct in Research" (Adopted in 2014)

[7] Science Council of Japan, "Response: Improving Soundness in Scientific Research" (March 6, 2015)

[8] NIH, "Final NIH Statement on Sharing Research Data" (2003)

[9] Japan Science and Technology Agency, "Implementation Guidelines: JST Policy on Open Access to Research Publications and Research Data Management" (2017)

[10] DRYAD, "Joint Data Archiving Policy (JDAP)" (2011)

[11] Hochschulrektorenkonferenz, "Management von Forschungsdaten - eine zentrale strategische Herausforderung für Hochschulleitungen" (2014) (German Rectors' Conference, "Management of research data – a key strategic challenge for university management" (2014))

[12] Allianz der deutschen Wissenschaftsorganisationen, "Positionspapier,, Research data at your fingertips "der Arbeitsgruppe Forschungsdaten" (2015) (Alliance of Science Organizations in Germany, "Research Data at Your Fingertips -- A Position Paper by the Research Data Working Group" (2015)

[13] European Commission "Open Innovation, Open Science, Open to the World"(2016)

[14] Cabinet Office, Government of Japan, Report by The Expert Panel on Open Science based on Global Perspectives (2015)

[15] Committee on Open Science, Science Council of Japan, "Recommendations Concerning an Approach to Open Science That Will Contribute to Open Innovation" (2016)

[16] Cabinet Office, Government of Japan" Integrated Innovation Strategy" (2018)

[17] Royal Society, "Science as an open enterprise" (2012)

[18] Science International, "Open Data in a Big Data World" (2015)

[19] Expert Panel on Open Science based on Global Perspectives, "Guideline for Establishing Data Policy at National Research and Development Agencies" (2018)

[20] Cabinet Office, Government of Japan "Definition and Characteristics of a Multi-stakeholder Process" (2008) (Last accessed September 30, 2018)

[21] Michael Nielsen, "Reinventing Discovery: The New Era of Networked Science" (2013)

[22] Center for Information on Security Trade Control, "Export Control at Universities" (Last

accessed September 30, 2018)

[23] Japan Consortium for Open Access Repository, "JPCOAR Schema Guidelines" (first published 2017 and revised thereafter)

[24] Japan Consortium for Open Access Repository, "RDM Training Tool" (2017)

[25] gacco, "Research Data Management in the Age of Open Science" (2017)

[26] EDiNA, "MANTRA—Research Data Management Training" (Last accessed 2018.9.30)

[27] Digital Curation Center, "Disciplinary RDM Training" (Last accessed 2018.9.30)

[28] Digital Curation Center, "Data management and curation education and training"(Last accessed September 30, 2018)

[29] FORCE11, "FAIR Data Principles" (2016)

[30] National Bioscience Database Center, "The FAIR Data Principles as Data Sharing Standards" (2018)

[31] FOSTER, "What is Open Science? Introduction" (Last accessed 2018.9.30)

[32] European Union, "Open Science Monitor" (2018)

[33] Budapest Open Access Initiative, "Ten years on from the Budapest Open Access Initiative: setting the default to open" (2012)

[34] Japan Society for the Promotion of Science, "Implementation Policy on Open Access to Research Papers in Projects of the Japan Society for the Promotion of Science" (2017)

Members of Academic eXchange for Information Environment and Strategy (AXIES) Research Data
Management Subcommittee (as of May 1, 2019)

担当理事
相原 玲二 (広島大学)

部会主査
青木 学聡 (京都大学)

部会副査
船守 美穂 (国立情報学研究所)　　　　　脇昌 弘 (ソニービジネスソリューション)
松原 茂樹 (名古屋大学)　　　　　　　　朝原 啓太 (日立製作所)

部会員(順不同)
梶田 将司 (京都大学)　　　　　　　　　橋本 耕平 (日立製作所)
野間口 真裕 (京都大学)　　　　　　　　荒木 豊 (日立製作所)
金子 康樹 (慶応義塾大学)　　　　　　　酒匂 正昭 (日立製作所)
尾城 孝一 (国立情報学研究所)　　　　　森 幹生 (日立製作所)
河野 明日路 (国立情報学研究所)　　　　川上 晋 (日立製作所)
込山 悠介 (国立情報学研究所)　　　　　渡部 真 (日立製作所)
山地 一禎 (国立情報学研究所)　　　　　箕輪 修一 (日立製作所)
林 正治 (国立情報学研究所)　　　　　　野村典文 (伊藤忠テクノソリューショ
林 洋平 (国立情報学研究所)
嶋田 晋 (筑波大学)
土屋 雅稔 (豊橋科学技術大学)
辻井 高浩 (奈良先端科学技術大学院大学)
西村 浩二 (広島大学)
緒方 淳 (武蔵大学)
北條 優 (琉球大学)
松平 拓也 (金沢大学)
高田 良宏 (金沢大学)
宮本 貴朗 (大阪府立大学)
外山 勝彦 (名古屋大学)
荒牧真之 (日立製作所)
中町喜好 (日立製作所)
宮村秀雄 (日立製作所)