

**Guidelines for Management, Disclosure, Sharing, and Use of Research Data  
in RIKEN Spring-8 Center (RSC), RIKEN  
(RSC Research Data Guidelines) (draft)**

March 31<sup>st</sup>, 2021

RIKEN Spring-8 Center (RSC)

1. Purpose

The primary mission of the RIKEN Spring-8 Center (RSC) of RIKEN (hereinafter referred to as “Institute”) is to maintain stable and reliable operation of the large-scale research facilities Spring-8 and SACLA and related facilities (hereinafter referred to as “User Facilities”), and to provide brilliant X-rays for a broad range of researchers from universities, research institutes, and industries. The RSC develops state-of-the-art technologies for the production and utilization of brilliant X-rays and creates new science through the synergistic use of the two research facilities. The RSC also has an organizational function to produce its own research data.

The RSC shall contribute to promoting the Institute’s research activities through information technology and maximizing the results of such research, in compliance with the Basic Policy on Management, Disclosure, Sharing, and Use of Research Data (hereinafter “the Basic Policy”) set forth by the Institute. The purpose of the RSC’s Research Data Guidelines (hereinafter “these Guidelines”) is to define matters necessary for handling research data in the RSC so that the researchers, etc. who belong to the RSC can manage research data in an appropriate and reliable manner.

At the User Facilities, users who belongs to the institutions other than the Institute (hereinafter “the External Users”) also generates research data. These guidelines do not apply to these External Users’ research data, because all rights to these research data belong to the External Users.

2. Definitions of Research Data and Useful Data

[Research data]

In the Basic Policy, “research data” means all kinds of data acquired or generated in the course of research undertaken at the Institute. All research data in the RSC shall be managed by electromagnetic means. Print and other media that cannot be managed electromagnetically shall be converted appropriately so that they can be managed by electromagnetic means. Research data does not include research notes, research papers, or programs that are executed on a computer, such as database systems or software components. Research data shall be classified as below, depending on how it is used, how it is collected, or its source.

[Primary data]

Research data newly generated in the course of research shall be called “primary data.” Primary data includes the data, including metadata, that is generated and collected from observation

instruments, measurement instruments, imaging devices or other research instruments, and the research data measured and created by researchers.

[Secondary data]

“Secondary data” means the data generated by processing primary data, secondary data, or both.

[Useful data]

According to the Basic Policy, “Useful data” means the research data, including metadata, which shall be disclosed along with publication of research papers (hereinafter “Data of Research Papers”) and other academic documents or research data that the RSC has determined on behalf of the Institute will contribute to the development of science and technology.

[Local research data]

“Local research data” means research data that is used exclusively by researchers or research projects and that is placed in a data management area that cannot be accessed by any third party. Data management media include fixed storage devices connected to computers exclusively used by researchers, storage media provided with various discs or non-volatile memory, and storage areas that can be accessed through a network and accessed exclusively by users.

[Useful local data]

“Useful local data” means local research data that is categorized as useful data.

[Disclosure/sharing flag]

“Disclosure/sharing flag” means the information that indicates where to disclose useful data registered in the repository. In accordance with the Basic Policy, the types of disclosure/sharing flags include those for “undisclosed data,” “data with restricted internal access,” “data with restricted internal/external access,” and “publicly disclosed data.”

[Data generator]

“Data generator” means a person who has generated research data.

[Data manager]

“Data manager” means a person who manages research data generated by data generators, using information systems infrastructure or information media. The data manager may be either the same as or different from the data generator.

[Data user]

“Data user” means a person who receives and uses useful data.

### 3. Selection and Use of Research Data Repository

The research data repository developed by the Institute (hereinafter referred to as "Institution repository") shall, in principle, be used for the management of research data as well as useful data.

In some cases, such as the generation of large-volume data, however, it may be difficult to manage data effectively, for example, due to limitations of repository capacity or transfer network

performance. In such a case, it is necessary to request the Institute to increase repository functions, which must be accompanied by objective and quantitative evidence. The RSC shall endeavor to precisely measure the quality and quantity of research data generated in the RSC in order to ensure the effective operation of the repository.

At the User Facilities, most of the research data and operation logs (hereinafter referred to as "User-Facility research data") generated by the operation of the User Facilities and the use of the User Facilities are stored automatically into the data systems associated with the User Facilities, such as the accelerator control systems, and user experiment-data systems (hereinafter referred to as "User-Facility repositories"), which are operated by the RSC. Because these User-Facility research data is a large-volume data, storing these User-Facility research data into Institution repository in addition to the User-Facility repositories is not rational in terms of the cost. The use of the User-facility repositories is granted for the User-Facility research data. However, even if the User-Facility research data are stored in the User-Facility repositories, the handling of research data should be in compliance with the Basic Policy.

#### 4. Handling of Personal Information, etc. in Medical Research Involving People

The research data including personal information that is generated in the RSC or received from outside the Institute when conducting medical research involving people shall be handled strictly as highly sensitive data. Specifically, such data shall be handled in compliance with the regulations for the handling of personal information established by the Institute, as well as relevant laws, regulations, and guidelines, and ethical guidelines.

Furthermore, data received from outside the Institute will be treated as primary data for research, after being processed in accordance with the guidelines for information management including the regulations described above.

#### 5. Registration of Research Data in Repository

The research data generated in the course of research and managed by the RSC typically has a life cycle from the generation of primary data, to registration of data in a repository for sharing and storage, data release, and data deletion. Specifically, the life cycle of research data can be described as follows. In this case, the data generator and the data manager may be either the same or different.

##### a. Generation of primary data

The primary data generated by a data generator is managed as local research data by a data manager.

##### b. Generation of secondary data

The data generator generates secondary data by analyzing and processing primary data or secondary data. The data manager manages the generated secondary data as local research data.

##### c. Selection of useful data

The data manager selects the data to be stored, shared, and disclosed as useful data from the primary or secondary data managed as local research data, and generates useful local data.

#### d. Registration in the repository

The data manager uploads useful local data to the repository for registration by using the upload functionality provided by the repository. At the time of registration, it is desirable to set the disclosure/sharing flag status as “undisclosed data” to prevent the data from being disclosed accidentally.

#### e. Setting of disclosure/sharing flag status

The disclosure/sharing flag for data is set according to the procedure described below.

#### f. Deletion of data

Data is deleted according to the procedure described below.

In selecting useful data from research data, the data generator or data manager comprehensively evaluates and determines the need for data storage or data sharing with others, from various perspectives—including the status of research, the management of intellectual property in research, and a basis for the research paper to be written—and according to his/her conscience as a researcher.

Specifically, the types of data selected as useful data include:

- Data recorded in a laboratory notebook and which needs to be stored;
- Data to be shared in collaborative research;
- Data needed in the next phase of research; and
- Data needed for publication of and as a basis for a research paper that is planned to be written.

Consideration shall also be given to the fact that the operation and maintenance of the repository necessary for the storage of research data involves a significant cost. Efforts shall be made to reduce costs and improve the effectiveness of useful data, including omitting some data to an extent that does not hinder the circulation or use of data, compressing data, and limiting the scope of the data that can be generated in a program to the original data.

The data manager shall be responsible for managing local research data and useful local data. In a case where such data is managed by using information systems infrastructure or media other than the repository, necessary measures shall be taken to prevent the theft or loss of, or unauthorized access to the information media, and the leakage of data at the time of disposal.

## 6. Changes of Disclosure/Sharing Flag Status

For useful data registered in the repository with the status of disclosure/sharing flag as “undisclosed data,” the data generator or data manager shall change the disclosure/sharing flag status by analyzing research trends and data users and by obtaining approval for change of

disclosure/sharing flag status from the head of the division. The following is the disclosure/sharing flag status listed from smallest to largest in terms of the scope of eligible users.

[Undisclosed data] This flag status is given to data accessible only by the data generator or data manager.

[Data with restricted internal access] This flag status is given to data which data users in the Institute can read. It is possible to target all those in the Institute or only some of them by restricting access e.g. by password or IP address.

[Data with restricted internal/external access] This flag status is given to data targeted at a single data user or multiple data users, regardless of whether inside or outside the Institute.

[Publicly disclosed data] This flag status is given to data that can be read by anyone, without limiting the scope of eligible data users.

The means in which data is provided to users should be individualized, according to data characteristics. For example, there is no need to constantly keep data readable through networks, etc., and it is possible to read the data and offer it using an information medium at the written request of a user.

After the setting of disclosure/sharing flag status, the data manager shall endeavor to ensure the setting of appropriate status, comprehensively considering research trends, the value of useful data, the progress of collaborative research, and the effectiveness of data release, and shall change the disclosure/sharing flag status as needed with the approval of the head of the division.

## 7. Period of Grace for Disclosing Useful Data

Useful data shall, in principle, be disclosed within two years from the time when the data is generated or within one year from the time when the data is selected as useful data, whichever comes earlier. However, it may be handled case-by-case, if there is any contract or agreement at the time of carrying out research associated with the useful data. In a case where the period of grace for disclosing Data of Research Papers is stipulated in the conditions for publication of the research paper, such data shall be disclosed in accordance with such provisions.

## 8. Licensing Use of Useful Data

Efforts shall be made to appropriately present the license to use useful data, along with the scope of disclosing and sharing the data, or its disclosure/sharing flag status to the recipient of the data in order to prevent an infringement on the rights of the holder of the copyright of the data. In a case where a collaborative research agreement or a Material Transfer Agreement (MTA) has been concluded, a license to use useful data shall be sought according to such an agreement.

To grasp the status of data use without identifying recipients, it is necessary to ask the receivers to provide their information (such as the name and affiliation of a receiver), together with IP address and other records before downloading the data, in recognition of its being personal information. When the data is presented as secondary-use results in research data, the license to use the useful data, which requires the user to provide information on the data providing source

in the body or acknowledgement of the user's research paper, shall be specified.

In preparing a license to use useful data, consideration should be given—such as including a provision to request that identifiers for references to source data or the author of the data should be used in the research paper—so that the readers of the research paper can refer to reference source data. For providing data to more recipients by enabling them to freely read the data, including the use of data from a program, the type of authorization to use the data that is most suitable for open data, such as a Creative Commons licence, should be used.

## 9. Storage Period of Useful Data and How to Manage in Repository

As set forth in the Basic Policy, the useful data to be registered in the repository shall be kept for 10 years or longer. Due to various circumstances, such as the advance of science and technology and changes in trends, the value of using the useful data may change during the storage period of the useful data. The data manager shall endeavor to check the useful data registered in the repository on a regular basis and set the disclosure/sharing flag status and license for use that are most suitable for the data.

## 10. Use of Public Repositories

In the handling of publicly disclosed data, there are some cases where it is recommended to use public repositories depending on the research field or the type of data. Publicly disclosed data should be registered in a public repository when it is deemed suitable to use public repositories, after comprehensive consideration of the conditions for disclosing data set by the public repository, such as the period for disclosing data or licensing use of data. At that time, metadata shall be registered and disclosed in the repository developed by the Institute.

## 11. Providing Metadata

Research institutes of all sizes in Japan and abroad have started to publicly disclose metadata on the Web, thereby leading to the formation of a global network of research data linked by metadata. In response to this trend, the ISC shall endeavor to provide high-quality metadata from very useful research data in order to promote open science and the use of research data. For describing metadata, it is desirable to use standardized ontology and data items to ensure mutual interoperability on a global scale. Researchers who have generated research data shall seek to create metadata of the world's highest quality by using the ontology and tools provided or recommended by and gaining technological support from the RSC's division in charge.

In view of the costs and convenience of data management, it is most desirable to generate metadata when target data is produced. However, metadata is often provided at the time of selecting useful data or sharing/disclosing data, thereby requiring more efforts to generate metadata. To address this issue, efforts shall be made to work on the development of a plan for handling research data, considering, at the stage of developing a research plan, the level of detail of metadata and the timing of providing it.

When the description of metadata refers to details on scientific findings, including experimental conditions and a phenomenon seen in research, the metadata itself can become high-value

research data. In such a case, efforts shall be made to provide an appropriate disclosure/sharing flag status or a license to use the metadata, as for useful data other than metadata. However, considering the spread of the standard technology used for queries, which enables users to automatically search for specific metadata from programs, etc., the metadata created as a data catalog shall be disclosed by providing a Creative Commons license with the approval of the head of the division to lower barriers hindering access.

Although research data and the metadata provided to the research data are closely connected, consideration shall also be given to the case where they differ in the status of the disclosure/sharing flag to be set. It is possible to disclose metadata but keep research data undisclosed, for example in the case where the content of research is proactively disclosed by using metadata while preserving its intellectual property rights. In addition, in a case where data is disclosed in an urgent manner and it takes some time to provide detailed metadata, it is acceptable to disclose data provided with a minimum level of metadata, such as the generator of the data and the location of the data.

## 12. Response to the Transfer of Registrant

The useful data registered in the repository shall, in principle, be managed responsibly by each registrant or the head of the division to which the registrant belongs to. When a registrant is transferred from the RSC to another center, etc., the head of the division to which the registrant belongs shall take on the responsibility for management. The head of the division who takes on the responsibility for management may request another person in charge to manage the useful data registered in the repository. If there is any useful data for which nobody can take on management responsibility, the head of the division shall consult the RSC's division in charge on how to manage the data.

## 13. Deletion of Data

As stipulated in the Basic Policy, useful data, including metadata, to be registered in the repository shall be kept for 10 years or longer. After expiration of the storage period, the data shall be deleted. In a case where it is deemed desirable to prolong the storage period of data owing to its scientific importance, however, the storage period may be prolonged on a temporary basis after approval by the head of the division.

Useful data that is deemed appropriate to delete prior to the expiration of the storage period may be deleted from the repository after approval by the head of the division, which requires justification for deletion of the data. Once the data is removed, the reason for the deletion of the data shall be registered in the repository, together with the setting of the same disclosure/sharing flag as that prior to the data deletion.

For the deletion of data, there is no need to delete both research data and its metadata at the same time, and an appropriate plan should be developed to delete each.

## 14. Evaluation of Researchers, etc. and Reward for Their Accomplishments

A researcher who generates, acquires, and actively shares research data with others as useful data

may be entitled to receive an appropriate evaluation and reward for his/her accomplishments in this regard.