2. Research Data Management

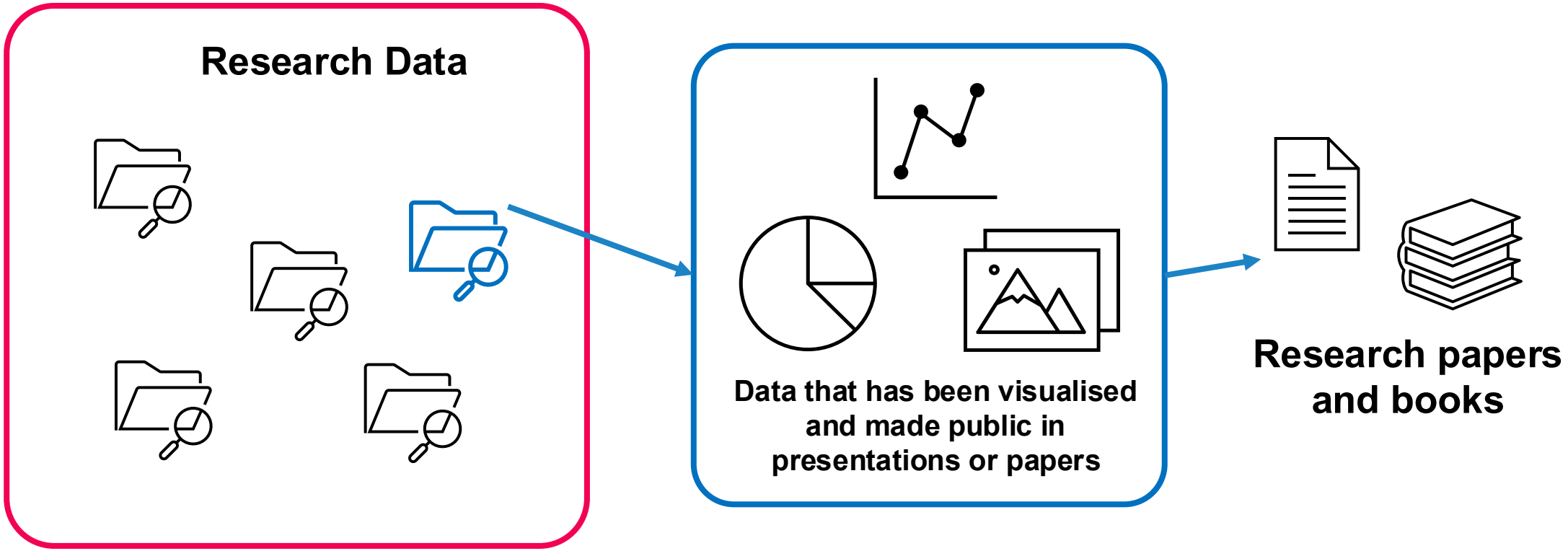
What are Research Data?

All data and records collected by researchers for a research project:

- Research notes (written text)
- Photographs
- Diagrams
- Videos
- Audio recordings
- Data and measurements entered in spreadsheets etc.
- Presentation slides
- Website links, etc.

What are Research Output Data?

- Data and records that were used and published in presentations or papers
- Created in accordance with the customs and policies of the research field



Why Manage Research Output Data?

1. To preserve the evidence supporting research findings

   Especially from 2024 onward, when applying for government-funded grants such as KAKENHI, researchers will be required to submit a research data management plan.

2. To be able to provide the data ASAP when required

   Researchers have a responsibility to promptly provide research output data when requested

3. To preserve research output data for over 10 years and ensure they can be disclosed if needed

Basic Management of Supporting Data

Basic information necessary for data management:

1. Data name: What kind of data is it?
2. Creator: Who created it?
3. Creation date: When was it created?
4. Storage: Where is it stored?
5. Citation: Which part of which paper (or research topic) is it used in?
6. Creation method: How was it created or collected?
7. Scope of disclosure
8. Usage permission (if it can be made public)

◎ Ideally, record items 1 - 8 in a spreadsheet (or similar format)

Basic Management of Supporting Data

Basic information necessary for data management:

1. Data name: What kind of data is it?
2. Creator: Who created it?
3. Creation date: When was it created?
4. Storage: Where is it stored?
5. Citation: Which part of which paper (or research topic) is it used in?
6. Creation method: How was it created or collected?
7. Scope of disclosure
8. Usage permission (if it can be made public)

◎ Ideally, record items 1 - 8 in a spreadsheet (or similar format)

Why Is It Necessary to Set the Scope of Disclosure and Usage Permissions for Research Output Data?

1. To promote the reuse of research data and the sharing of research insights
2. To allow authors to indicate how their research output data should be handled, while protecting interests, such as personal information and patents

Why Is It Necessary to Set the Scope of Disclosure and Usage Permissions for Research Output Data?

1. To promote the reuse of research data and the sharing of research insights
2. To allow authors to indicate how their research output data should be handled, while protecting interests, such as personal information and patents

Scope of Disclosure for Research Output Data

Points to consider regarding the scope of data disclosure:

- **Make decisions in line with legal and ethical requirements:**
  - Personal information
  - Copyright
  - National security
  - International relations
  - Research-related contracts
  - Research data management policies of affiliated institutions and funding agencies, etc.
- **Ensure data quality at the time of publication:**
  - Resolution, format and size, color, annotations, etc.

Determining Whether Data Can be Published

1. **Try to lift any restrictions on publication**
   - Anonymize data containing personal information
   - Obtain permission from data rights holders
   - Set an embargo period to delay publication

2. **If publication is possible: decide where to publish**
   - Storage system or repository is needed
   - Confirm that the data meets requirements such as format, size, retention period, access control, and metadata (bibliographic information)

Determining Whether Data Can be Published

1. Try to lift any restrictions on publication

- Anonymize data containing personal information
- Obtain permission from data rights holders
- Set an embargo period to delay publication

2. If publication is possible: decide where to publish

- Storage system or repository is needed
- Confirm that the data meets requirements such as format, size, retention period, access control, and metadata (bibliographic information)

Usage Permissions

Apart from copyright, clearly indicate the conditions under which users may use your data.

- **Clearly express your intent regarding usage permissions in writing.**
  - Describe metadata such as the data's source, copyright, and relevant information.
- **Using licenses pre-established by organizations like Creative Commons is convenient.**
  - CC licenses allow clear marking with standard icons and are documented in multiple languages.

Examples of Creative Commons License Logos
https://creativecommons.org/share-your-work/cclicenses/

Usage Permissions

Apart from copyright, clearly indicate the conditions under which users may use your data.

- **Clearly express your intent regarding usage permissions in writing.**
  - Describe metadata such as the data's source, copyright, and relevant information.
- **Using licenses pre-established by organizations like Creative Commons is convenient.**
  - CC licenses allow clear marking with standard icons and are documented in multiple languages.

Examples of Creative Commons License Logos
https://creativecommons.org/share-your-work/cclicenses/

Planning for Long-term Data Preservation (10+ Years)

△ Relying only on personal storage carries risks:

- Obsolescence or degradation of storage or software
- Data loss due to changes in affiliation

○ It is recommended to use **both personal storage and a repository**

- Repository = A storage system for data

Examples of repositories capable of long-term preservation and publication:

- **Institutional repositories** (Store and manage papers, images, and publicly available research data created by members, semi-permanently)
- **Discipline-specific repositories**
  - Zenodo (Capable of storing diverse datasets)
  - Informatics Research Data Repository (IDR by NII)
    https://www.nii.ac.jp/dsc/idr/en/

Planning for Long-term Data Preservation (10+ Years)

△ Relying only on personal storage carries risks:

- Obsolescence or degradation of storage or software
- Data loss due to changes in affiliation

○ It is recommended to use **both personal storage and a repository**

- Repository = A storage system for data

Examples of repositories capable of long-term preservation and publication:

- **Institutional repositories** (Store and manage papers, images, and publicly available research data created by members, semi-permanently)
- **Discipline-specific repositories**
  - Zenodo (Capable of storing diverse datasets)
  - Informatics Research Data Repository (IDR by NII)
    https://www.nii.ac.jp/dsc/idr/en/

The University of Osaka Institutional Knowledge Archive OUKA

Provided Services

- Support for open access of papers
- Support for repository registration of published image data
  - High-resolution image publication is possible under the IIIF standard
  - Next, we will introduce the method for registering images in the repository, and the features, uses, and creation methods of the IIIF standard

Supplement: Managing Data Location (1)

[Managing All Research Data, Folders and Files]

1. Gather all research-related data in one place
2. Divide folders by academic year or research project name
   - Do not create too many nested folders (keep it to about two levels)
3. Setting a naming rule for files makes them easier to search
   - Example file name: dataName_YYYYMMDD_version.(file extension)
   - Unify alphanumeric characters and symbols in file names to half-width
   - Do not include spaces in file names

Supplement: Managing Data Location (1)

[Management of Figures and Datasets Used in Publications]

1. Create a project folder, combining the publication year and research project name, distinct from the main research data folder
2. From the main research data folder, copy the data (e.g., images) that were used for the figures of the publication, into the project folder
3. Record the items shown on the slide “Basic Management of Supporting Data” (page 5) in a spreadsheet:
    - Title
    - Repository URL (if accessible)
    - File name and location in the main research data folder
    - Scope of disclosure and usage permissions

References

1. Research Data Utilization Forum (RDUF), Research Data Licensing Subcommittee (2019) “Guidelines for specifying conditions of use in research data publishing, ver.1.0”, p.31. https://ir.soken.ac.jp/record/6591/files/%E7%94%B22366_Appendix%205.%20Guidelines%20for%20specifying%20conditions%20of%20use%20in%20research%20data%20publishing.pdf (in English) https://doi.org/10.11502/rduf_license_guideline (in Japanese)

2. Kyoto University Committee for Researcher Information Infrastructure, Research Data Management Subcommittee (2021) “Guidelines for Establishing Departmental Policies on the Management and Disclosure of Research Data at Kyoto University”, p.15. https://repository.kulib.kyoto-u.ac.jp/dspace/bitstream/2433/262758/5/guideline_20210315.pdf (in Japanese)