

An Earth and Space Science Rubric and ePortfolio for Research Data Management

**Fifth Research Data Management Workshop
March 30, 2021**

Janice Ann Smith, Ph.D., ePortfolium
Toshihiko Iyemori, Ph.D., Kyoto University
Masahito Nose, Ph.D., Nagoya University
Jacques Raynauld, Ph.D., ePortfolium
Shoji Kajita, Ph.D., Kyoto University

KARUTA
open source portfolio

ePortFolium

Overview

1. The RDM Project
2. The CDL Rubric
3. The Earth and Space Science Rubric
4. The Karuta Open Source Portfolio Project
5. An RDM ePortfolio Prototype
6. Next Steps
7. Questions

Overview of the Research Data Management (RDM) Project

Overview

- Teams of disciplinary faculty create RDM rubrics for the physical, biological, and social sciences, as well as for interdisciplinary research.
- A RDM ePortfolio is designed, customized, and tested using the Karuta Open Source Portfolio (version 3.0).
- The rubric and ePortfolio guide instruction in RDM seminars for new researchers and postdocs.

The CDL Rubric

The CDL Rubric

	Ad Hoc	One-Time	Active and Informative	Optimized for Re-Use
Planning your project	When it comes to my data, I have a "way of doing things" but no standard or documented plans.	I create some formal plans about how I will manage my data at the start of a project, but I generally don't refer back to them.	I develop detailed plans about how I will manage my data that I actively revisit and revise over the course of a project.	I have created plans for managing my data that are designed to streamline its future use by myself or others.
Organizing your data	I don't follow a consistent approach for keeping my data organized, so it often takes time to find things.	I have an approach for organizing my data, but I only put it into action after my project is complete.	I have an approach for organizing my data that I implement prospectively, but it not necessarily standardized.	I organize my data so that others can navigate, understand, and use it without me being present.
Saving and backing up your data	I decide what data is important while I am working on it and typically save it in a single location.	I know what data needs to be saved and I back it up after I'm done working on it to reduce the risk of loss.	I have a system for regularly saving important data while I am working on it. I have multiple backups.	I save my data in a manner and location designed maximize opportunities for re-use by myself and others.
Getting your data ready for analysis	I don't have a standardized or well documented process for preparing my data for analysis.	I have thought about how I will need to prepare my data, but I handle each case in a different manner.	My process for preparing data is standardized and well documented.	I prepare my data in such a way as to facilitate use by both myself and others in the future.
Analyzing your data and handling the outputs	I often have to redo my analyses or examine their products to determine what procedures or parameters were applied.	After I finish my analysis, I document the specific parameters, procedures, and protocols applied.	I regularly document the specifics of both my analysis workflow and decision-making process while I am analyzing my data.	I have ensured that the specifics of my analysis workflow and decision-making process can be understood and put into action by others.
Sharing and publishing your data	I share the results of my research, but generally I do not share the underlying data.	I share my data only when I'm required to do so or in response to direct requests from other researchers.	I regularly share the data that underlies my results and conclusions in a form that enables use by others.	Because of my excellent data management practices, I am able to efficiently share my data whenever I need to with whomever I need to.

The Earth and Space Science RDM Rubric

Characteristics

- A rubric should point to specific learning outcomes in the syllabus and/or curriculum of any course or program in which it is to be used.
- Effective rubrics require many redrafts and testing / review by potential users.
- Each cell in the rubric must be measurable in some way.

Purposes

- A rubric should be usable as a chart or map describing:
 - The typical progression for learning the skills and achievements it contains.
 - A step-by-step process for teaching those skills and achievements.
 - A means to assess each student's progress in mastering those skills and achievements.

RDM Physical Science Rubric Phases

- Planning for Data
- Organizing Data
- Analyzing Data
- Sharing and Publishing Data

RDM Physical Science Rubric Dimensions

- Planning for Data

- Data to be Obtained
- Methodology for Obtaining Data
- Team Members
- Funding
- Re-Use of Data
- Open / Closed Data

- Organizing Data

- Data Collection
- Storage
- File Format
- Versioning
- Back-up

RDM Physical Science Rubric Dimensions

- Analyzing Data

- Software for Analysis
- Data Processing
- Intermediate Data
- Data Cleaning

- Sharing and Publishing Data

- Licensing
- Metadata Schema
- Data Completeness, Validity, Reliability, and Ethics
- Sharing or Publishing the Data

RDM Physical Science Rubric Levels

- Beginning
- Developing
- Improving
- Completed

PLANNING FOR DATA	Beginning	Developing	Improving	Completed
Data to Be Obtained	The essential data to be produced have been identified.	The types and usage of the essential data to be produced have been identified.	The resolution, quality, and minimum period (or area) of the essential data to be produced have been identified.	The researcher is confident that the essential data, as identified by data types, usage, resolution, quality, and minimum period (or area), will be produced.
Methodology for Obtaining Data	The methodology required to obtain the data has been determined.	The instruments, network, and storage device required to obtain the data have been identified.	Preparation of the instruments, network, and storage device has begun.	Preparation of the instruments, network, and storage device is complete.
Team Members	Team members for producing the data have been identified.	Roles for team members with specific skills have been assigned.	A plan for effective communication among team members has been prepared and collaboration has begun.	Members of the team understand their roles and those of their team mates and are fully prepared to work together.
Funding	Funding required for the research has been estimated and possible funding agencies have been identified.	A budget for funding team members, instruments, equipment, procedures, and/or travel has been prepared.	One or more proposals to funding sources have been prepared and submitted.	Sufficient funding to support the research project in obtaining the required data has been secured.
Data Re-Use	It is understood that plans for re-use of the data to be produced may change ways in which it will be managed.	The availability of the data for re-use has been determined.	The embargo period for re-use of the data has been determined.	Plans for re-use of the data, including necessary changes in its management, are complete and accepted by the team.
Open / Closed Data	Requirements or requests for data publishing by desired journals have been identified. Options for sharing data in open science repositories have been considered.	Decisions as to which data will be open for publication and which data will be closed have been made.	Possible licenses and repositories for publishing and sharing the data have been identified.	A variety of options for open and closed data have been documented and are understood and accepted by all team members.

ORGANIZING DATA	Beginning	Developing	Improving	Completed
Data Collection	Planned observations and associated measurements are listed.	Listed items are documented in language easily understood by other researchers.	Plans have been made to convert observations and associated measurements to machine readable form.	Plans have been made to regularly upload the data to a secure server.
Storage	The necessary volume for data storage has been estimated.	A minimum amount of necessary data storage has been arranged.	Sufficient storage to accommodate additional data has been identified.	The researcher is confident that the data to be produced will be safely stored for a sufficiently long period of time.
File Format	Internationally-accepted file formats commonly used by the discipline have been considered.	An easily-understandable file format for effective data analysis has been selected.	The selected data format has been documented.	The data have been formatted in ways that will be understood internationally.
Version Control	A plan to use versioning to process complicated data over time has been prepared.	Descriptors to distinguish the most recent version of the data from previous versions have been identified.	An archival system has been prepared to store previous versions of the data.	The researcher is confident that all existing versions of the data will be available to researchers who choose to access them.
Backup	Plans for backing up the data have been made.	The process for backing up the data is ready to use but is not automatic.	The data are backed up automatically on a regular schedule.	The data are backed up in more than one location or repository.

ANALYZING DATA	Beginning	Developing	Improving	Completed
Software for Analysis	Appropriate software with needed functionality has been identified and obtained for data analysis.	Missing functionality in available software has been identified and "home grown" software developed to meet the need.	The chosen software has been used to analyze the data.	Documentation of the software has been prepared. If the software was originally developed by the project, considerations are in process to make it open.
Data Processing	Research instruments have been set up to record observations in numerical values.	Numerical values have been converted to physical quantities in order to extract meaningful results.	The converted data have been formatted with a format commonly used by the discipline.	The formatted data have been transferred to a computer for cleaning, analyzing, and sharing.
Intermediate Data	It has been determined whether or not intermediate data generated during the research process merits retention and storage.	Plans have been made for formatting and storing intermediate data.	Documentation explaining the content of intermediate data has been prepared.	Intermediate data is retained and stored with along with its documentation.
Data Cleaning	Data plots have been created for visual scanning.	Visual scans of rough data checks are used to identify noise.	Data files are duplicated with the originals preserved. Obvious noise (e.g., errors, spikes, steplike jumps) is removed	Data are rechecked so that deleted noise determined to be valuable can be restored.

SHARING OR PUBLISHING DATA	Beginning	Developing	Improving	Completed
Licensing	The rules for use of the data by other researchers have been determined.	Reasons for licensing the data are understood and possible licenses have been identified.	A suitable internationally-recognized license has been selected.	Indication of the chosen license has been attached to the published data.
Metadata Schema	Observations and associated measurements have been reviewed for completeness.	The researcher has acquired an understanding of metadata and available software to facilitate its creation.	A metadata schema used internationally by the discipline has been selected.	Metadata has been prepared according to the selected metadata schema.
Data Completeness, Validity, Reliability, and Ethics	The researcher is satisfied with the completeness (integrity), validity, and reliability of the data.	At least one expert in the discipline is satisfied with the completeness, validity, and reliability of the data.	Evidence for the integrity, validity, and reliability of the data has been documented.	Issues with data ethics have been assessed and addressed.
Sharing or Publishing the Data	A suitable repository for publishing and sharing the data has been secured and a digital object identifier (DOI) has been obtained.	The method and schedule for sharing the data have been fixed.	Data ownership, network security, and the performance of the repository backup system have been confirmed and documented.	The researcher has confirmed that the data, along with its metadata and documentation, have been published in the repository and are available to other researchers.

The Karuta Open Source Portfolio Project

- Capture learning processes
- For learning, assessment, and showcasing
- Customizable for specific purposes
- Incorporating rubrics, dashboards, and reports
- Templates for sharing processes

Under the Umbrella of the Apereo Foundation

“Apereo is a [open source] global network, with member institutions on six continents. We actively seek to develop partnerships to further our mission of creating and sustaining software supporting learning, teaching and research.”

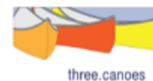


An History of Responding to User Needs

- Founding Partners: HEC Montréal, Kyoto University, IUT-2 Grenoble (France), Three Canoes (USA), ePortfolium (Canada)
- International Governance Board - International Community

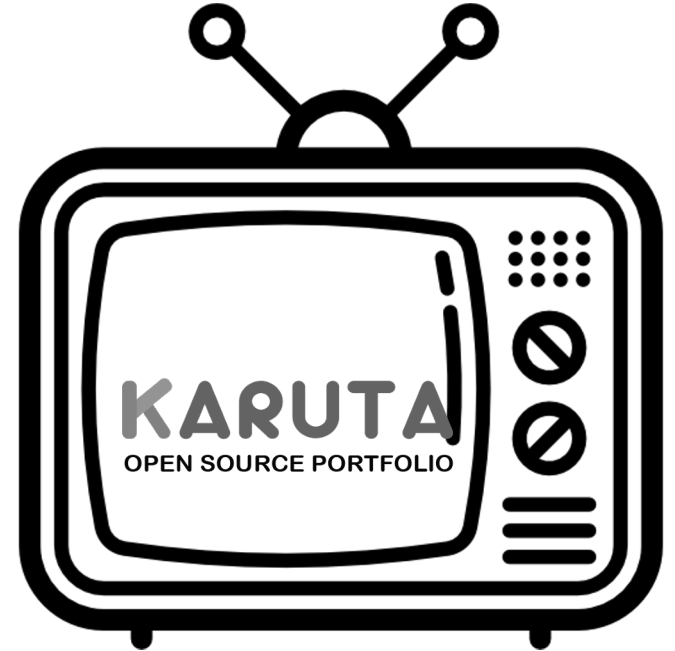


ePortFolium

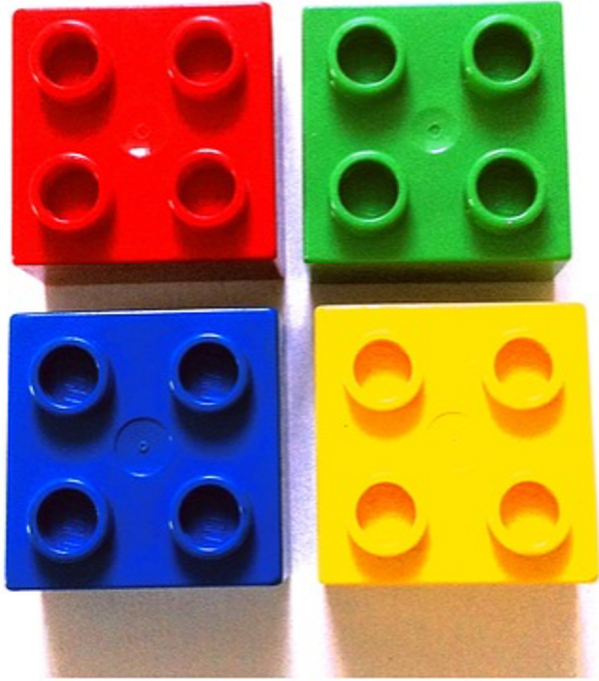


Karuta's Main Features

- Open Source
- LTI ready
- Responsive Design (tablets)
- Multilingual
- Presentation, learning, or accreditation ePortfolios
- Support from a commercial partner (ePortfolium)



Karuta's Flexibility Can Capture any ePortfolio Process



- Organize **different resources** (text, documents, rubrics, comments, etc.)...
- ... according to a **workflow** for different users (students, evaluators, mentors, instructors, etc.).
- with a complete control over the **UI** presentation

Listen to Users and Start an Iterative Design Process

**KARUTA
Designer**

**Prototype
– Pilot**

**KARUTA
Production**

Meet the stakeholders, listen to their objectives, show them a prototype, and immediately incorporate their comments.



An RDM ePortfolio Prototype

An RDM ePortfolio Prototype

- The RDM ePortfolio maps a step-by-step learning process:
 - The learning process is guided by an RDM Rubric.
 - Participants upload and document evidence for each level of each dimension.
 - Participants self-evaluate mastery of each level.
 - Instructors assess participant evidence and self-evaluation using a shared dashboard.

1. Access the RDM seminar rationale and instructions on the Welcome page.

My Research Portfolio

Planning for Data



Organizing Data



Analyzing Data



Sharing and Publishing Data



Research Portfolio Test

[Welcome](#) [Dashboard](#) [Phases](#)

Planning for Data

- + Data to be obtained
- Methodology for Obtaining Data
- Team Members
- Funding
- Data Re-Use
- Open / Closed Data

PLANNING FOR DATA	Beginning	Developing	Improving	Completed
Data to Be Obtained	The essential data to be produced have been identified.	The types and usage of the essential data to be produced have been identified.	The resolution, quality, and minimum period (or area) of the essential data to be produced have been identified.	The researcher is confident that the essential data, as identified by data types, usage, resolution, quality, and minimum period (or area), will be produced.
Methodology for Obtaining Data	The methodology required to obtain the data has been determined.	The instruments, network, and storage device required to obtain the data have been identified.	Preparation of the instruments, network, and storage device has begun.	Preparation of the instruments, network, and storage device is complete.
Team Members	Team members for producing the data have been identified.	Roles for team members with specific skills have been assigned.	A plan for effective communication among team members has been prepared and collaboration has begun.	Members of the team understand their roles and those of their team mates and are fully prepared to work together.
Funding	Funding required for the research has been estimated and possible funding agencies have been identified.	A budget for funding team members, instruments, equipment, procedures, and/or travel has been prepared.	One or more proposals to funding sources have been prepared and submitted.	Sufficient funding to support the research project in obtaining the required data has been secured.
Data Re-Use	It is understood that plans for re-use of the data to be produced may change ways in which it will be managed.	The availability of the data for re-use has been determined.	The embargo period for re-use of the data has been determined.	Plans for re-use of the data, including necessary changes in its management, are complete and accepted by the team.

2. Access the phases and dimensions of preparing and curating research data.

3. Select a phase and dimension.

Evidence

Provide evidence (text, document, images, videos) to support your self-assessment and link it to the different levels.



Menu ▾

Level

Beginning ▾

Essential Data to be Produced

Baseline data from previous study

 <https://agupubs.onlinelibrary.wiley.com/doi/full/2005GL024083>

Add text

Add document

Add url

Add image

4. Add evidence to demonstrate mastery of each level.

Self-evaluation

Self-evaluate (yes or no) as you progress in the development of this dimension.

Beginning

The methodology required to obtain the data has been determined.

Status

Yes



Date

2021/03/17

5. Self-evaluate (based on uploaded evidence) at each level of the dimension.

6. View instructor evaluation via a shared student / instructor dashboard.

Planning for Data	Beginning		Improving		Developed		Completed		Evaluation professor
	Evidence	Status	Evidence	Status	Evidence	Status	Evidence	Status	
Data to Be Obtained	✓✓✓	Yes	✓✓✓	Yes	✓✓✓	Yes	✓✓✓	Yes	Completed
Methodology for Obtaining Data	✓✓✓	Yes	✓✓✓	Yes	✓✓✓	No	✓✓✓	Yes	Completed
Team Members	✓✓✓	Yes	✓✓✓	Yes	✓✓✓	No	✓✓✓	Yes	Completed
Funding	✓✓✓	Yes	✓✓✓	Yes	✓✓✓	No	✓✓✓	Yes	Completed
Data Re-Use	✓✓✓	Yes	✓✓✓	Yes	✓✓✓	Yes	✓✓✓	No	Developed
Open / Closed Data	✓✓✓	Yes	✓✓✓	Yes	✓✓✓	Yes	✓✓✓	No	Developed

Next Steps

RDM Rubrics

- Collaborate with additional research faculty to produce RDM rubrics for:
 - Biological Sciences
 - Social Sciences
 - Humanities?
 - Interdisciplinary Disciplines and Projects
- Ask additional researchers to review and refine each RDM rubric.

RDM ePortfolios

- Further develop the RDM ePortfolio prototype.
- Test the ePortfolio prototype with volunteer students and faculty and refine it based on their feedback.
- Adapt the resulting ePortfolio to additional disciplines and test with additional volunteers.

RDM Seminars

- Use the RDM rubrics and ePortfolios to organize and guide seminars for new researchers and postdocs.
- Further refine the rubrics and ePortfolios based on seminar participant feedback.

Questions

Thank you!

For further information, contact me at:

janice.smith@threecanoes.com

KARUTA
open source portfolio

ePortFolium