Report

March 17, 2025

Recommendations on Evaluation Items for Open Science Activities

Interim Report

(English Translation)

RIKEN

The Open Life Science Project (OLSP)

The Working Group on the Evaluation of Open Science Activities

RIKEN, The Open Life Science Project
The Working Group on the Evaluation of Open Science Activities
List of Members　(as of March 17, 2025)

RIKEN	
KASUKAWA Takeya	RIKEN Center for Integrative Medical Sciences
KOBAYASHI Norio	RIKEN Information R&D and Strategy Headquarters
KUSHIDA Tatsuya	RIKEN BioResource Research Center
MASUYA Hiroshi	RIKEN BioResource Research Center
NIKAIDO Itoshi	RIKEN Center for Biosystems Dynamics Research
OKUMURA Yoshihiro	RIKEN Center for Brain Science
ONAMI Shuichi	RIKEN Center for Biosystems Dynamics Research
YAMAGATA Yuki	RIKEN Information R&D and Strategy Headquarters
YAMAMOTO Yasunori	RIKEN Information R&D and Strategy Headquarters
Kyushu University	
ITO Takashi	Faculty of Medical Sciences
National Agriculture and Food Research Organization	
KAWAMURA Takahiro	Research Center for Agricultural Information Technology
Research Organization of Information and Systems	
GOTO Susumu	Database Center for Life Science
National Institute of Informatics	
KOMIYAMA Yusuke	Digital Content and Media Sciences Research Division
National Institute of Genetics	
FUJISAWA Takatomo	Bioinformation and DDBJ Center
The University of Osaka	
MIZUGUCHI Kenji	Institute for Protein Research

Total: 15 members

Summary

Advancements in cutting-edge information and communication technologies (ICT) have enhanced collaboration among researchers and expanded opportunities for sharing and disseminating research outputs. In this evolving environment, science is expected to progress through mutual recognition of contributions while respecting the diversity of research. Ideally, open science should function as an ecosystem in which researchers are motivated by incentives and where the benefits of their efforts are returned to them. However, in practice, a significant gap remains between this ideal and the current reality.

The Working Group has developed this recommendation on the Evaluation of Open Science Activities, composed of researchers from both within and outside RIKEN. It proposes evaluation items intended to serve as one of the incentives for promoting open science in the life sciences field. Chapter 1 provides background information. Chapter 2 defines the scope of open science addressed in this recommendation as "open data science," and discusses its current state in Japan as well as the ideal form of open data science led by researchers. Chapter 3 examines the handling of "research data," which plays a central role in open data science, analyzing its relationship with scientific publications in the life sciences, the benefits and risks of data sharing and dissemination, and issues concerning data circulation and persistence. Based on this analysis, Chapter 4 presents specific evaluation items.
These are as follows:
1. Realization of high-quality research data,
2. Data dissemination,
3. Outreach activities for data utilization,
4. Data management and persistence,
5. Contributions to the research community, and
6. Human resource development.

The aim is to encourage the scientific community to implement multifaceted and appropriate evaluations based on these six criteria.
This document is an interim report, and the evaluation items will continue to be reviewed with input from researchers and other stakeholders. The final report will present specific methods for evaluating each item, along with definitions of quantitative and qualitative measures.

Background to This Recommendation

Within RIKEN, the Open Life Science Project (OLSP) was launched as a research initiative to promote open science activities in the life sciences. As part of this effort, the Working Group on the Evaluation of Open Science Activities was established to develop criteria for evaluating such activities.

The first meeting of the working group was held in August 2021 and was attended by a total of 11 participants, including 10 researchers from RIKEN—some of whom were affiliated with life science-related centers—and one external researcher. A brainstorming session was conducted during this meeting. Following the second meeting, which was held on November 1, 2021, six external researchers were invited to exchange views, and a workshop (review meeting) was held on January 18, 2022, to facilitate further discussion.

Subsequently, additional review meetings with invited external researchers were held on March 14, 2023, and September 2, 2024. In parallel, draft recommendations were reviewed via social media as needed, and numerous discussions were held among the core members of the working group. Throughout this process, it was determined that extensive deliberation would be required to establish concrete methods for measuring and calculating the evaluation indicators. Accordingly, the working group decided to compile an interim report that presents the proposed evaluation indicators as the first step. The interim report was approved for release by the OLSP Steering Committee on March 17, 2025.

The final report is scheduled for publication at a later date. It will incorporate the latest trends in open science both domestically and internationally, propose specific methods for measuring and calculating the evaluation indicators, and reflect feedback from researchers in Japan and abroad, thereby further refining the recommendations.

1. Background of This Recommendation

The primary objective of science lies in discovering and understanding the laws of nature. Knowledge gained by humanity through scientific endeavors has traditionally been documented and shared in the form of scholarly articles, significantly contributing to the advancement of science. Scientists have published their research findings, novel perspectives, and methodologies in academic papers, thereby promoting the creation of new knowledge and the deepening of scholarly discourse. In principle, scientific knowledge has been shared in a manner that is accessible to all, fostering research driven by free and creative thought. Since the advent of academic journals, the dissemination of research findings through publications has become standard practice, and science has come to be regarded as an open endeavor. However, in reality, challenges related to sharing research data remain. Specifically, a lack of incentives, physical and legal barriers, and the high cost of subscriptions to academic journals continue to restrict access to scientific knowledge. In recent years, the open science movement has aimed to reduce these access barriers by leveraging modern media and infrastructure, while upholding the fundamental principle of knowledge sharing in science. The goal is to improve the quality, speed, and reliability of research.

In the life sciences, the sharing of research data has been emphasized even before the concept of open science became firmly established. This is because the field not only encompasses the biological diversity of organisms but also requires a wide range of research approaches. The generation of research data necessitates a unique scientific problem setting, advanced experimental design, and sophisticated methods of data analysis, making collaboration among researchers essential. The advancement of omics technologies has enabled the comprehensive acquisition of data related to genes, proteins, transcription, metabolism, phenotypes, and other biological processes, thereby accelerating the progress of data-driven science. Furthermore, rapid developments in measurement technologies have enabled the generation of vast and highly diverse datasets. By integrating such data and utilizing artificial intelligence, more complex and advanced analyses have become achievable.

On the other hand, the sharing and publication of research data involve numerous challenges. It is necessary to organize and convert data into formats that facilitate reuse and application. Through the development of an Open and Closed Strategy [7], appropriate management is required to distinguish between data that should be made publicly available and data that should remain restricted. Additionally, the proper maintenance and management of shared data are crucial, requiring ongoing operational support. From the perspective of data security, it is also necessary to address emerging regulations on data management, such as those under the Foreign Exchange and Foreign Trade Act (FEFTA) and controls related to technologies with potential military applications.

The realization of open data relies heavily on adherence to the FAIR principles [5]—Findable, Accessible, Interoperable, and Re-usable. To meet these principles, a wide range of efforts is required, including the standardization of data sharing and reuse practices, data format conversion and restructuring, the assignment of metadata, validation of data accuracy, and the selection of appropriate data licenses. However, these tasks represent an additional burden for researchers and can pose significant barriers, particularly for those newly engaging in open science. Moreover, existing evaluation indicators are primarily based on quantitative metrics such as the volume of published data and the number of data accesses. These do not directly assess the effort involved in data sharing, and thus often fail to serve as effective incentives for researchers. To overcome these challenges, it is essential to establish new evaluation items and criteria that provide appropriate incentives for researchers. Furthermore, it is necessary to develop evaluation systems not only at the individual level but also at the organizational level to build mechanisms that support the promotion of open science.

The Working Group on the Evaluation of Open Science Activities was established in August 2021, with its inaugural meeting held the same month. Its purpose was to address the aforementioned challenges by focusing on the life sciences and attempting to define the ideal form of open data-centered activities—one of the key elements of open science—as "Open Data Science." At the same time, the group aimed to formulate evaluation items for Open Data Science activities that would serve as a basis for appropriate incentives, through a process led by researchers themselves. Subsequently, the group engaged in numerous discussions, both in-person and online, ultimately compiling the outcomes on March 17, 2025. As a result of these efforts, this recommendation has been formally released.

2. Open Science Addressed in This Recommendation: Open Data Science

To clarify the scope of discussion in this recommendation, the range of open science addressed herein is explicitly defined. Research activities are multifaceted, and the concept of open science is similarly diverse; therefore, it is challenging to implement all aspects uniformly. As such, definitions should be established with due consideration to the specific circumstances of each country, organization, and research field. In this recommendation, the discussion is limited to the life sciences, and a definition is provided that reflects the specific conditions in Japan. Based on this definition, evaluation items are examined. This approach is not intended to hinder international dialogue; rather, it presumes that further detailed discussion will take place in consideration of the varied contexts of the global research community.

2.1. Current Status in Japan and Worldwide

Full-scale discussions on open science began in earnest following the G8 Science Ministers Meeting held in the United Kingdom in June 2013, where the open access of scholarly publications and the public availability of research data were recommended [2]. In Japan as well, deliberations began on a new approach to scientific research in which the results of publicly funded research, such as publications and research data, would be accessible and usable not only by the scientific community but also by society at large. In March 2015, the Cabinet Office of Japan released a report from the "Expert Panel on Open Science Based on Global Trends" [3]. Since then, the Japanese government, the Science Council of Japan, and other relevant bodies have promoted open science as part of their science policy, presenting a range of strategies. Meanwhile, as a means of managing and sharing research data, public repositories tailored to specific data types and research fields are being operated. In some cases, the deposition of research data in public repositories is required as a condition for submitting academic papers. Furthermore, academic journal publishers have begun collecting and managing not only articles but also related data, raising concerns about the potential monopolization of research data by commercial entities. Thus, the current development of open science is being led primarily by governments and commercial organizations, and it cannot be said that researchers' leadership and autonomy are being fully exercised.

2.2. A Vision for Researcher-Led Open Science

Open science is one of the strategic approaches for advancing scientific research more efficiently and diversely, and it is something that researchers themselves should actively promote. It is essential to establish an ecosystem in which researchers who engage in open science can receive tangible benefits in return for their contributions. Strengthened collaboration and the development of mutually beneficial relationships among researchers are also desirable outcomes.

In particular, life science research is characterized by the following features:

- Diversity of biological species

- Diversity of research methodologies and approaches
- Emphasis on data comprehensiveness (e.g., omics data)
- A variety of measurement instruments
- Diversity and the vast scale of research data
- Diversity of applied research areas (e.g., agriculture, materials and substances, drug discovery, medical science)

With recent advances in information and communication technology (ICT) and intelligent information science, it has become increasingly necessary to integrate ICT into conventional research methodologies to enable the efficient sharing of large and diverse datasets. Furthermore, transforming research practices through the utilization of artificial intelligence and the promotion of data-driven research is an urgent challenge. Researchers are also expected to demonstrate, through objective means, that their findings have been obtained through appropriate methods, and to ensure that these outcomes are disseminated not only within the life sciences research community but also to society at large. In this recommendation, open science is framed from the perspective of data science as "Open Data Science," and concrete discussions are developed based on the current state of the life sciences and related considerations.

2.3. Open Data Science Based on the Current State of the Life Sciences

Based on the above, this recommendation presents a definition of open science that, while focusing on the life sciences, is as broadly applicable as possible.

Definition of Open Data Science:

Open Data Science refers to the advancement of science through the sharing and dissemination of research outcomes while respecting the diversity of research, by leveraging advanced information and communication technologies (ICT) and strengthening collaboration among researchers. This definition assumes the utilization of cutting-edge ICT to unlock the potential of science fully and regards open science not merely as a methodology, but as science itself.

Strategies for Realizing Open Data Science:

To realise Open Data Science, the following strategies are required:

1. Accelerate research activities and respond rapidly to the increasing sophistication and diversification of research
2. Return research outcomes to society and promote industry-academia collaboration

through joint and applied research

3. From a long-term perspective, promote the sharing of research results that are transparent, reproducible, and sustainable

Researchers are expected not only to transform research methodologies through the incorporation of ICT but also to actively return research outcomes to the academic community and society while ensuring the reliability of their findings. Moreover, it is necessary to establish a framework in which contributions to open science itself, in addition to individual research achievements, are appropriately recognized and evaluated.

3. Handling of Research Data in the Life Sciences

This chapter clearly defines the scope of research data to be addressed in the discussions. First, although this recommendation does not provide a detailed discussion on the management of research records, it is assumed that research data must be properly managed to enable a prompt response in cases where questions or concerns arise regarding published papers. Researchers are considered responsible for managing such data over the long term.
On the other hand, this recommendation emphasizes research data that supports the advancement of open science. In other words, the scope of research data considered here includes data and datasets that are disseminated and utilised for the advancement of science.

3.1. Research Papers and Research Data

The relationship between research papers and research data in the life sciences is discussed from two perspectives:
A. Experimental (wet) researchers and data (dry) researchers, and
B. Advances in ICT and measurement technologies.

A. Experimental Researchers and Data Researchers

Data scientists (or data researchers) aim to elucidate biological phenomena through data analysis and simulation using computational methods. Because their work involves data and software, the collection and generation of research data, as well as the publication of those data and associated programs, are relatively straightforward.
In contrast, experimental researchers investigate biological phenomena using reagents and

biological samples. Although they are often perceived as being on the opposite end of the spectrum from data researchers, they play a crucial role in setting research questions and exploring solutions—activities that result in the generation of research data. Therefore, sharing and publishing research data, together with papers that describe the experiments in detail and accompanying metadata, can significantly contribute to the advancement of life sciences as a whole.

To promote the sharing and publication of research data, effective measures include: implementing functions that automatically record the settings and conditions of measurement instruments; standardizing data formats to reduce barriers to data distribution; and providing tools that facilitate the reuse and application of research data.

B. Advances in ICT and Measurement Technologies

A typical research paper consists of an introduction, methods, results, and discussion. Among these, the "results" section is written most objectively and holds universal value. Research data are likewise objective in nature and, in principle, should be documented in the results section. However, due to recent advancements in measurement technologies, it has become possible to generate vast and highly diverse datasets. Documenting all such data within a single paper is no longer realistic.

As a result, it has become common practice for authors to compile the datasets that underpin their results and to publish them by submitting them simultaneously with the manuscript to the journal, or by depositing them in a data repository.

3.2. Benefits and Risks of Data Sharing and Publication for Researchers

When sharing and publishing research data, it is essential to formulate an appropriate *Open and Closed Strategy* that takes into account the management of personal and confidential information, as well as the contractual conditions associated with the research. Simply making data open is not sufficient; researchers must handle data following a carefully considered strategy that weighs both the risks and benefits.

The following risks are associated with the sharing and publication of research data:

Risk 1: Misuse of Data

There is a risk that data may be falsified, misinterpreted, subjected to inappropriate analysis, or used for improper purposes, potentially damaging the credibility of both the researcher and the research field.

Risk 2: Issues Related to Intellectual Property Rights
If the copyright of research datasets (or databases) is not correctly handled, the original authors may suffer harm, and the researcher responsible may lose professional trust.

Risk 3: Loss of Research Competitiveness
By making research data publicly available, there is a risk that other researchers may use the same data to conduct new studies, thereby reducing the original researcher's competitive advantage.

Conversely, the following benefits can be gained:

Benefit 1: Increased Transparency and Reliability of Research
Open access to research data enables the reproduction and verification of research results, leading to improvements in research quality, reliability, and the evaluation of researchers.

Benefit 2: Greater Citation of Research
When publicly available data are cited by other researchers, the original researcher's influence expands. This can enhance their leadership in setting research agendas, establishing standards, securing funding, and attracting talent.

Benefit 3: Generation of New Research and Innovation
The reuse of open data by other researchers can lead to the emergence of new studies and innovations, thereby advancing scientific progress.

Benefit 4: Improved Efficiency through Data Reuse
Sharing data among researchers reduces redundant experimentation and enhances the overall efficiency of research activities.

In light of these risks and benefits, it is essential for research institutions and individual researchers to develop Open and Closed Strategies from a multifaceted perspective. Designing such strategies is inherently complex and requires a high level of awareness and sound judgment from researchers. Furthermore, the formulation of such a strategy should itself be recognized as a critical component of open science activities and evaluated accordingly.

3.3. Research Activities and Research Data

Research data are the outcomes of research activities, and the two are deeply intertwined. This section discusses research activities in the life sciences from six perspectives, serving as a foundation for defining evaluation criteria for open science activities.

3.3.1. High-Quality Research Data

Producing high-quality research data involves significant costs. From the standpoint of data utilization, however, the value of the data becomes a critical factor. If the quality and value of research data can be appropriately assessed, such assessments may serve as incentives for the creation of high-quality datasets.

While the quality and value of data depend on scientific and societal needs, originality, and rarity, it is challenging to evaluate these attributes using absolute criteria. Nonetheless, such assessments are essential to promote open science by incentivizing the production of valuable data. In addition, the comprehensiveness of data in specific topics, particularly with the advancement of omics research, is a crucial evaluation axis. Usage metrics such as data access counts are also widely used as indicators.

In the context of open science, standardization of data formats and descriptions contributes significantly to enhancing data reusability. For example, adopting internationally standardized formats and using controlled vocabularies improves interoperability and promotes broader use. It is also essential to ensure that data can be easily read, visualized, and analyzed. Converting data formats dependent on proprietary software into standard formats, without compromising data quality or informational content, facilitates broader dissemination.

Furthermore, the accurate assignment of metadata significantly enhances the quality and value of the data. In managing research data, it is essential to thoroughly understand the research content and manage metadata by considering the necessary and sufficient elements, as well as reusable formats. Recording conditions under which the data were generated, such as creation date, authorship, and geographic information, enhances the reliability of the data. Managing version information and update history also contributes to usability. Such metadata should employ standardized vocabularies, schemas, and data models, just as the data themselves do. Including metadata that specifies the number of data entries and overall data volume can be helpful for secondary use. Considering the increasing use of AI-generated metadata, it is also essential to distinguish whether metadata was assigned manually or automatically for proper accuracy assessment.

Creating high-value data requires an understanding of the associated costs. However, this awareness has not yet taken firm root in Japan. Assigning metadata or annotations to research data should be the responsibility of the researcher. It must be evaluated in proportion to the effort, ultimately contributing to improved data quality and value.

3.3.2. Circulation of Research Data

Research data should be shared and published in accessible and usable formats and methods. The FAIR principles are a key guideline in this regard. To ensure that data and metadata are accessible on the web, they must be assigned appropriate identifiers (e.g., unique web-accessible IDs). Additionally, setting up websites for retrieving data and metadata, and developing application programming interfaces (APIs) for programmatic access, enhances the utility of research data and facilitates its circulation. When sharing and publishing research data and metadata, it is essential to establish appropriate access and usage rights based on an Open and Closed Strategy. Intellectual property rights (IPRs) must be clarified and presented to users, and access should be granted accordingly. Licensing information can be provided on webpages, through APIs or embedded directly into metadata.

Moreover, employing authentication mechanisms to manage user access is also essential. By adjusting access rights based on user authentication and imposing necessary restrictions, data can be shared and published appropriately while maintaining confidentiality and security.

3.3.3. Outreach Activities for Data Reuse

To promote the reuse of research data, it is essential to raise awareness about the availability of shared and published datasets. Traditional dissemination methods include presenting at academic conferences, publishing in journals, and issuing press releases. Additionally, websites and video content that introduce datasets and their applications are effective. Dialogue-based outreach, such as workshops and briefing sessions, can provide tailored communication to meet user needs. Establishing a user support desk also allows for direct feedback. Publishing research data involves considerable effort and cost—compiling and organizing data, assigning metadata and annotations, ensuring data quality, and guaranteeing data persistence. Therefore, a system that offers incentives surpassing these burdens is necessary to encourage researchers to publish their data proactively. Making it possible for researchers to track the utilization of their data and incorporating this into research evaluations can be a practical approach. To further promote data reuse, it is essential to share analytical methods among researchers and cultivate a new publication culture in which the use of published data and the methods applied are clearly stated within research articles. Moreover, outreach also encompasses the use of research data across different fields and in society. In the life sciences, where applications span agriculture, materials, drug discovery, and medical science, research data can play critical roles in various sectors. Cross-disciplinary utilization may also yield new scientific value. Returning research data to society through industry-academia collaboration contributes not only to industrial development but also strengthens the research capabilities and financial base of academic institutions. Data reuse is also vital in the context of citizen science, involving participation by non-professional individuals, young aspiring researchers, and students. For such audiences, communication must be clear and considerate, with explanations that differ from those aimed at expert researchers.

Thus, the advancement of data reuse requires well-designed outreach and incentive systems, and cooperation among researchers, industry, and society is essential to enhancing the value of shared data.

3.3.4. Research Data Management and Sustainability

In the life sciences, a vast amount of data is generated daily. Researchers and research institutions are expected to develop plans for the proper management and utilization of these data as intellectual assets of individuals or institutions. Specifically, data should be managed appropriately according to their characteristics—for example, data to be shared and published as open data, data to be preserved as evidence of research integrity, and data to be retained in a non-public form.

It is also essential to establish a system that enables research data to be stably and securely shared and published over the long term—that is, to ensure sustainability, which requires continuous and proactive efforts.

One effective means of ensuring the long-term sharing and publication of research data is to assign globally persistent identifiers, such as Digital Object Identifiers (DOIs), to prevent link rot. However, the long-term preservation of data cannot be sustained by individual researchers alone; it requires the concerted efforts of their institutions and the broader research community. For example, data management can become difficult when a research project concludes or a researcher changes institutions. Researchers and communities that succeed in enabling long-term data sharing and sustainability should be recognized as contributors to open science. Ensuring data sustainability also requires the construction or utilization of secure data management environments, with appropriate authentication and access controls in place. Research institutions must develop robust research data management infrastructures to support such functionality. This includes not only the development of repositories and network infrastructures but also the creation of middleware and tools that support data management. Because the required infrastructure varies depending on the research field and methodology, it is crucial to develop these systems in collaboration with researchers to meet their specific needs. Moreover, to promote planned and systematic data management, the development and implementation of a Data Management Plan (DMP) is a practical approach. In recent years, an increasing number of research projects have required the formulation of a DMP, and some funding agencies now mandate it as a condition for grant approval. Proper implementation of DMPs can strengthen data management systems and establish a foundation for long-term data reuse. Appropriate management and sustainability of research data are not only essential for improving the reliability and reproducibility of research but also play a vital role in maximizing the value of data and contributing to future research endeavors.

3.3.5. Participation in the Research Community

The management and standardization of research data, as well as the development and sharing of related tools, are supported by the activities of research communities where many researchers gather to exchange ideas and collaborate. While researchers may adopt unique data formats or specifications to demonstrate the novelty of their own research outputs, it is essential that, to the greatest extent possible, widely accepted standard specifications are adopted to ensure that research data are accessible and usable by a broader community. As research advances, updating specifications while maintaining compatibility with existing standards should also be undertaken as part of community activities. Contributions to such efforts should be appropriately recognized within the framework of open science. Specific examples of such activities include participation in international consortia for the standardization, harmonization, design, implementation, and policy formulation related to research data and information infrastructure. The publication and maintenance of analytical programs and data management tools that handle research data, as well as the sharing and updating of open-source tools, are also important activities that merit recognition. Furthermore, it is essential to strengthen compliance with ethical and regulatory norms in research data management to ensure that data can be used safely and responsibly. Activities related to the development, construction, operation, harmonization, and standardization of tools and infrastructure that support these goals should also be recognized as integral to the advancement of open science. In addition to collaboration through data sharing, dialogue among researchers and the exchange of knowledge and techniques through publications are indispensable for promoting open science. While interdisciplinary collaboration is undoubtedly essential, collaboration among researchers and engineers within the same field to deepen the shared research infrastructure should also be regarded as a significant and valuable activity.

3.3.6. Human Resource Development

Human resource development is of critical importance in the open science movement. Securing and nurturing individuals with diverse expertise and technical skills is essential not only for advancing research and contributing to society but also for ensuring the long-term sustainability of open science initiatives. One of the primary objectives of such human resource development is the acquisition and dissemination of the competencies necessary for the creation, compilation, maintenance, management, sharing, and publication of research data. Equally important is the acquisition and dissemination of skills required for the construction and operation of research data infrastructure.

It is also necessary to develop relevant educational content (teaching materials) and to offer training and learning opportunities for their practical application. In addition, the cultivation of individuals capable of contributing to scientific advancement through the adoption of new information and communication technologies (ICT) and innovative research methodologies is increasingly in demand.

The sustained development of open science depends on the availability of skilled individuals. It is also essential to foster researchers who can serve as leaders within the global research community, enhancing international competitiveness and contributing to the advancement of science as a whole.

To achieve these goals, active institutional engagement is required—not only from senior researchers and highly motivated individuals but also from the organizations themselves. For example, the development and dissemination of open science educational content by institutions, along with the implementation of workshops and training sessions, serve as effective measures for human resource development. Furthermore, recognizing institutional contributions to training efforts and incorporating the evaluation items proposed in this recommendation into performance assessments can increase incentives for participation in open science.

In recent years, with the growing prominence of data-driven science, the importance of a professional role known as the data steward has been increasingly recognized. Data stewards are responsible for specialized tasks related to data, including data management, quality assurance, and access control. Their responsibilities may include implementing data governance, maintaining accuracy and consistency, managing data life cycles, ensuring privacy and security, developing data use policies, and supporting data users.

The development of such professional talent is indispensable for promoting open science and should be considered a key area of focus moving forward.

4. Details of Evaluation Items

Among various elements of open science, those most closely related to life science research, particularly from the perspective of data science, and most urgently in need of defined evaluation items are those concerning open data. Accordingly, while taking into account other open science activities such as open access and human resource development, this recommendation focuses on listing evaluation items that are directly related to the sharing and dissemination of research outputs, with open data at the center.
In addition, consideration is given to the FAIR Principles, which specify the standards that research data must meet in data-intensive science enabled by open data. Thus, items such as identifiers, metadata, protocols, vocabularies, and licenses are also defined as components of the evaluation criteria.
Ultimately, the required elements for evaluation will be the evaluation indicators specified for each evaluation item. These indicators should ensure transparency by clearly stating the basis for assessment, being defined in a manner convertible to numerical scores, and being made publicly available. While this interim report is limited to the enumeration of evaluation items, the final report will include appropriate definitions of scoring methods as evaluation indicators, tailored to both domestic and international scientific contexts. The indicators will be examined and optimized through their practical implementation.

Based on the discussions in Section 3.3, this recommendation organizes and classifies evaluation items under six axes:
(1) Activities contributing to the value enhancement of data
(2) Data formats and distribution
(3) Outreach for data reuse
(4) Data management and sustainability
(5) Contributions to the research community
(6) Human resource development

Evaluation items are listed in detail below according to these classifications. Items marked with angle brackets (<>) refer to corresponding FAIR Metrics as described in Appendix 1 [6].

4.1. Evaluation Items Related to the Realization of High-Quality Research Data

This category allows for the scientific value of data, including its generation process, to be evaluated with transparency. The items listed focus on activities that contribute to or enhance the value of the data.

[Items related to the scientific and societal relevance of research data]
(1.a) Originality, rarity, and comprehensiveness of the data
(1.b) Usage metrics, such as the number of accesses to the research data
(1.c) Whether the data format follows standard specifications and uses controlled vocabularies

[Items related to metadata that enhance data quality and value]
(1.d) Whether information about measurement conditions and data provenance (e.g., geographic data, creation date/time, creator) is provided <F2>
(1.e) Whether the data format is correctly described <I1, I2>
(1.f) Whether version and update history are clearly documented <R1.2>
(1.g) Whether standardized vocabularies, schemas, or data models are used for metadata <I1, I2>
(1.h) Whether the accuracy of metadata is described (e.g., whether the metadata was manually curated or generated by AI)
(1.i) Whether the number and volume of data records in the dataset are specified

4.2. Evaluation Items Related to Data Distribution
This category includes attributes and quantitative criteria that data should meet for effective sharing and dissemination. It contains items corresponding to the FAIR Principles, such as data volume, identifier availability, and protocols for access.

(2.a) Whether the data or metadata are assigned a web-accessible identifier (e.g., accession number) <A1+F1>
(2.b) Whether an interface is provided for retrieving and searching data or metadata, including APIs <F4+A1.1>
(2.c) Whether licensing information is clearly presented and whether rights or intellectual property (IP) are specified (i.e., metadata includes IPR information) <R1.1>
(2.d) Whether authentication (including open access where applicable) and access rights are appropriately managed <A1.2>

4.3. Evaluation Items Related to Outreach for Data Reuse
This category includes activities that promote the reuse of shared and published data.

(3.a) Presentations at academic conferences, publications, or press releases about the data or metadata
(3.b) Publication of websites, videos, and holding of workshops or briefing sessions that introduce the data and explain how to use it
(3.c) Activities that encourage citation and proper referencing of reused data in scholarly publications
(3.d) Establishment of a user support desk or inquiry contact point
(3.e) Promotion of data use and collaboration within industry sectors, including agriculture and medicine
(3.f) Outreach to society, students, and non-researchers, including the next generation of researchers

4.4. Evaluation Items Related to Data Management and Sustainability
This category includes items related to compliance with data management practices.

(4.a) For ensuring data sustainability:
- Persistence of identifiers assigned to research data
- Existence of a framework for long-term sharing and publication of data

(4.b) Use of secure data management environments and proper implementation of authentication and access control in research data management
(4.c) Creation, periodic review, and adherence to a Data Management Plan (DMP)

4.5. Evaluation Items Related to Contributions to the Research Community
This category encompasses contributions to frameworks and implementations that facilitate the dissemination of broad data, including guidelines, tools, and systems.

(5.a) Participation in collaborative activities, including international consortia, for standardization, harmonization, design, implementation, and rulemaking related to research data and information infrastructures
(5.b) Sharing, publishing, updating, and maintaining open-source data management tools that support open science
(5.c) Development, construction, operation, harmonization, and standardization of tools and infrastructure to improve data management compliance
(5.d) Collaborative activities with other researchers and engineers engaged in similar fields, to advance open science initiatives

4.6. Evaluation Items Related to Human Resource Development

This category includes items essential for the sustainable promotion and development of open science.

(6.a) Development of personnel capable of creating, compiling, maintaining, managing, sharing, and publishing research data
(6.b) Development of personnel capable of constructing and operating infrastructures that support research data
(6.c) Creation and publication of educational content (teaching materials) related to open science
(6.d) Implementation of training programs and study groups to support human resource development in open science
(6.e) Investigation and introduction of new technologies and trends into research practices

5. Discussion and Conclusion

The evaluation items presented in this recommendation have been compiled into an interim report through discussions aimed at building consensus among a broad range of researchers, both in Japan and internationally. Moving forward, we plan to incorporate feedback from researchers and other stakeholders regarding this interim evaluation and to define concrete methods, as well as both quantitative and qualitative criteria, for each evaluation item. These will be compiled into a comprehensive set of Guidelines for the Evaluation of Open Science Activities. We, the authors, sincerely hope that these guidelines will be adopted by research institutions, universities, and other organizations and will be established as a standard and authoritative method for evaluating research. Through such efforts, we strongly aspire to foster greater interest and participation in open science among researchers, thereby contributing to the further advancement of open science.

End of Document

References

[1] Jeroen Bosman, and Bianca Kramer. Defining Open Science Definitions.
https://im2punt0.wordpress.com/2017/03/27/defining-open-science-definitions/

[2] G8 Science Ministers Statement (2013 年 6 月 13 日)
https://www.gov.uk/government/news/g8-science-ministers-statement

[3] Cabinet Office, Expert Panel on Open Science Based on Global Trends
(内閣府 国際的動向を踏まえたオープンサイエンスに関する検討会)
https://www8.cao.go.jp/cstp/sonota/openscience/

[4] openscienceASAP. Was ist Open Science?
http://openscienceasap.org/open-science/

[5] FORCE11, The FAIR Data Principles.
https://force11.org/info/the-fair-data-principles/

[6] FAIR Metrics
https://fairsharing.org/search?fairsharingRegistry=Standard&recordType=metric&page=1

[7] Cabinet Office, The 5th Science and Technology Basic Plan – Main Text
(内閣府 第 5 期科学技術基本計画 本文)
https://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf

Appendix 1. FAIR Metrics

Findable

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata

F3. Metadata clearly and explicitly include the identifier of the data they describe

F 4 . (Meta)data are registered or indexed in a searchable resource

Accessible

A1. (Meta)data are retrievable by their identifier using a standardized communications protocol

A1.1. The protocol is open, free, and universally implementable

A1.2. The protocol allows for an authentication and authorization procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

Interoperable

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles

I3. (Meta)data include qualified references to other (meta)data

Reusable

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards