



Legal Interoperability of Research Data: Principles and Implementation Guidelines

The Challenge:

Providing researchers with guidance for dealing with the legal aspects entailed when working with data from diverse sources.

What is the solution?

The CODATA-RDA Legal Interoperability Interest Group has studied the issues related to the intellectual property of data: the resulting outcome is a set of principles and practical implementation guidelines. They are offered as high-level guidance to all members of the research community—the funders, managers of data centers, librarians, archivists, publishers, policymakers, university administrators, individual researchers, and their legal counsel—who are engaged in activities that involve the access to and reuse of research data from diverse sources. The Principles are synergistic, so their greatest benefit is realized when they are considered together.



Produced by: **RDA/CODATA Legal Interoperability Interest Group**

<https://www.rd-alliance.org/groups/rdacodata-legal-interoperability-ig.html>

What is the impact?

The ability of the research community to share, access, and reuse data, as well as to integrate data from diverse sources for research, education, and other purposes requires effective technical, syntactic, semantic, and legal interoperability rules and practices. The Principles and guidelines produced by the RDA/CODATA Legal Interoperability Interest Group focus on legal interoperability because there tends to be misunderstanding and lack of knowledge and guidance about legal issues concerning research data generally.

The scope is limited to laws and policies pertaining to access and reuse of data, either produced by researchers themselves or used in research and other activities. It is specifically focused on the ability to integrate those data for research purposes in the context of intellectual property law.

RDA/CODATA Legal
Interoperability Interest
Group Supporting Output



March 2020



23 Things For Research Data Management

The Challenge:

Build capacity within the library community around research data management, ensure libraries are aligned with other research data management service providers, and share practices as new library services evolve and mature.

What is the solution?

23 Things: Libraries for Research Data provides an overview of practical, free, online resources and tools that librarians can begin using today to incorporate research data management into the daily practice of librarianship. Available in 12 languages, the resource groups 23 free, online resources to help librarians address issues on data management such as Metadata, Digital Preservation, Data Repositories, Data Licencing and Privacy or Data Citation.



Produced by: **Libraries for Research Data Interest Group**

<https://www.rd-alliance.org/groups/libraries-research-data.html>

What is the impact?

23 Things is a recognized training concept that supports flexible, self-paced learning designed to progressively build participants' understanding of key concepts and gain experience in research data management. Libraries and librarians hold an immense responsibility and are adapting their function to making data reliably accessible and re-usable. 23 Things will help them in applying the principles of library science to solve problems and provide new services related to research data. It could also support them in their potentially daily work of organizing, classifying, and describing research data and developing standards for metadata to help make data more easily discovered, understood, and preserved or in that of promoting the scholarship of data by encouraging and enabling data citation, assigning identifiers to datasets, creating links between documents and data, and helping users properly attribute credit to data producers.

**23 Things: Libraries for
Research Data Supporting
Output**



March 2020



Sustainable Business Models for Brokering Middleware to support Research Interoperability

The Challenge:

Broker software now has a level of maturity that requires addressing sustainment of the middleware, including the question of which business models could be adopted to best serve this need.

What is the solution?

Bridging scientific disciplines through the introduction of a broker or mediation capability has been demonstrated for a number of cross-discipline applications [Nativi, et al 2013]. The development and use of brokering middleware in a research environment is a relatively new development and the sustainability of such a capability has not previously been addressed. The Brokering governance Working Group, in addressing the questions of sustainability, does not presuppose which business/revenue or hybrid model might be embraced to sustain this middleware, but rather examines five classes of business models for consideration.



Produced by: **Brokering Governance WG**

<https://www.rd-alliance.org/groups/brokering-governance.html>

What is the impact?

The Brokering Governance Recommendation discusses the strengths and weakness of different models in the context of long-term sustainment of brokering middleware. Examination of alternatives leads to a prioritization among models with the recognition that no single model by itself is likely to provide the desired sustainment. The analysis suggests that a hybridized model incorporating aspects of three different business models over the lifespan of the brokering middleware, i.e. federally funded data facility guardianship in the establishment stage replaced or supported by a Consortium model and/or Software-as-a-Service as the broker matures, will likely provide the strongest model for sustainment.

the Brokering Governance WG
Recommendation



March 2020



RESEARCH DATA SHARING WITHOUT BARRIERS

The Research Data Alliance is an international organisation focused on the development of infrastructure and community activities aimed to reduce barriers to data sharing and exchange, and promote the acceleration of data driven innovation worldwide.

www.rd-alliance.org



@resdatall



10.000+ MEMBERS FROM
144 COUNTRIES



38 FLAGSHIP
RECOMMENDATIONS & OUTPUTS
100+ ADOPTION &
IMPLEMENTATION CASES



51 ORGANISATIONAL MEMBERS
& **10** AFFILIATE MEMBERS



28 WORKING GROUPS
57 INTEREST GROUPS



15 PLENARY MEETINGS



Workflows for Research Data Publishing: Models and Key Components

The Challenge:

Assisting research communities in understanding options for data publishing workflows and increasing awareness of emerging standards and best practices

What is the solution?

The RDA/WDS Publishing Data Workflows Working Group advocates the implementation of existing good practice for repositories and all parts of the data publishing process and the development of new standards where necessary. The group works on a data publishing reference model comprised of generic components which need to be interoperable and seamlessly accessible.



Produced by: **RDA/WDS Publishing Data Workflows WG**

[https://rd-alliance.org/groups/
rdawds-publishing-data-workflows-wg.html](https://rd-alliance.org/groups/rdawds-publishing-data-workflows-wg.html)

What is the impact?

All stakeholder groups, including publishers, libraries, data centres and most importantly researchers, have access to guidance on how to do professional data publishing using recommended best practices.

This will ensure improved interoperability of the individual elements of data publishing (e.g. journals with repositories) and ultimately drive the practice of their adoption in the research communities, hence fostering open science and data-driven discoveries and reliable reuse of research data.

Find out more about the
RDA/WDS Publishing
Data Workflows WG
Recommendation



March 2020



An open, universal literature-data cross-linking service

The Challenge:

Sharing information about the links between the literature and research data.

What is the solution?

Building on pre-existing components and international initiatives, the RDA/WDS Publishing Data Services Working Group is one of the drivers behind the “Data Literature Interlinking Service” (DLI), developed in a synergy with OpenAIRE & PANGAEA. DLI is aimed at improving visibility, discoverability, re-use and reproducibility by bringing 2M+ existing article/data links together, normalize them using a common schema, and exposing the full set as an open service.



Produced by: **RDA/WDS Publishing Data Services WG**

<https://rd-alliance.org/groups/rdawds-publishing-data-services-wg.html>

What is the impact?

Accessing and using literature-data links at large scale in an efficient and reliable way allows different stakeholders in the data publishing landscape to improve their services, increase data discoverability and usability.

Data centres will be able to assess much better how often their data is used in the literature, and present their users with links to relevant publications.

Find out more about the RDA/
WDS Publishing Data Services
WG Recommendation



March 2020



Repository Audit and Certification Catalogues

The Challenge:

Creating harmonized Common Procedures for certification of repositories

What is the solution?

Drawing from the procedures already put in place by the Data Seal of Approval (DSA) and the ICSU World Data System (ICSU-WDS), the RDA Repository Audit and Certification DSA-WDS Partnership Working Group has produced a convergent DSA-WDS certification standard aiming to eliminate duplication of effort, increase certification procedure coherence and compatibility thus benefitting researchers, data managers, librarians and scientific communities.



Produced by: **The Repository Audit and Certification DSA-WDS Partnership WG**

<https://rd-alliance.org/groups/rdawds-certification-digital-repositories-ig.html>

What is the impact?

The convergent DSA-WDS certification standard marks a step towards having more coherent, increasingly stringent and compatible standards for repository certification.

The harmonization of repository certification criteria will increase adoption and the number of certified repositories building trust for data generators, data consumers and funding bodies.

Find out more about
the Repository Audit
and Certification DSA-
WDS Partnership WG
Recommendation



March 2020



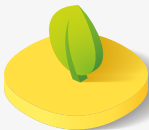
Wheat Data Interoperability Guidelines, Ontologies and User Cases

The Challenge:

Helping researchers create, manage and exchange wheat data

What is the solution?

To answer the needs of the Wheat Data community the RDA Wheat Data Interoperability Working group has built an interactive cookbook with recommendations and guidelines on wheat data use, set up a repository with wheat-related linked vocabularies and built a set of prototypes designed to test interoperability across datasets and services.



Produced by: **Wheat Data Interoperability (WDI) WG**

[https://rd-alliance.org/groups/
wheat-data-interoperability-wg.html](https://rd-alliance.org/groups/wheat-data-interoperability-wg.html)

What is the impact?

The Recommendations of the Wheat Data Interoperability WG are aimed at improving the management, sharing, discovery, reusability, and interoperability of data and datasets within the wheat community. The standardization and harmonization of wheat data will reduce variability and increase the relevance of wheat data related tools.

The outputs of this group are used as a building block in the Wheat Information System (Wheat IS) a framework to establish a global wheat information system, supported by the Wheat Initiative, a global initiative that brings together national and international research programmes to increase food security and nutritional value while taking into account societal demands for sustainable and resilient agricultural production systems.

Find out more about the
Wheat Data Interoperability
WG Recommendation



March 2020



Metadata Standards Directory

The Challenge:

Enabling discovery of metadata standards for documenting research data, regardless of academic discipline, and addresses issues related to coverage, ease of maintenance and sustainability.

What is the solution?

The Metadata Standards Directory Working Group adopted, enriched, & expanded the Disciplinary Metadata Standards Catalogue set up by the UK Digital Curation Centre (DCC), and developed a functional prototype directory based around the GitHub infrastructure, that places the information from the DCC directory into an environment where it can be maintained transparently with full version control.



Produced by: **Metadata Standards Directory (MSD) WG**

<https://rd-alliance.org/groups/metadata-standards-directory-working-group.html>

What is the impact?

The Metadata Standards Catalogue was built to guide researchers towards the metadata standards and tools relevant for their discipline. The directory drives up adoption of those standards, improving the chances of future researchers finding, accessing, and reusing the associated data.

By raising awareness of existing standards, the directory reduces the proliferation of ad hoc metadata formats and helps direct future standards development efforts towards those areas that most need it.



Find out more about the MSD
WG Recommendation

March 2020



Data Description Registry Interoperability Model

The Challenge:

Providing researchers with a mechanism to connect datasets in different data repositories based on various models such as co-authorship, joint funding, grants, etc.

What is the solution?

The Data Description Registry Interoperability Working Group is engaged in a series of bi-lateral information exchange projects and an open, extensible, and flexible cross-platform research data discovery software solutions.



Produced by: **Data Description Registry Interoperability WG**

[https://rd-alliance.org/groups/
data-description-registry-interoperability.html](https://rd-alliance.org/groups/data-description-registry-interoperability.html)

What is the impact?

Researchers will be able to use simple enabling infrastructures, such as the Research Data Switchboard, based on existing open protocols and standards with a flexible and extensible approach, to retrieve data relevant to their studies from multiple heterogeneous repositories.

Repositories will be able to plug-into these infrastructures enabling context based retrieval of the data, reducing fragmentation across institutions, countries and research domains.

Find out more about the DDRI
WG Recommendation



March 2020



Scalable dynamic-data Citation Methodology

The Challenge:

Supporting accurate citation of data subjected to change, for the efficient processing of data and linking from publications.

What is the solution?

The Dynamic Data Citation Working Group developed a simple, scalable mechanism that allows the precise, machine-actionable identification of arbitrary sub selections of data at a given point in time irrespective of any subsequent addition, deletion or modification.



Produced by: **Data Citation (WGDC) WG**

[https://rd-alliance.org/groups/
data-citation-wg.html](https://rd-alliance.org/groups/data-citation-wg.html)

What is the impact?

The main impact is on the reproducibility of work: it allows databases to be dynamically updated as needed, while users can still retrieve the data as it was at a specific point in time, providing valuable semantic information on the way the specific dataset was constructed.

As data gets migrated to new representations, the queries to extract the specific cited subset can be migrated accordingly, ensuring stability across technological changes enhancing data interoperability, sharing and re-use.

Find out more about the Data
Citation WG Recommendation



March 2020



Machine Actionable Policy Templates

The Challenge:

Data repositories need to publish their policies & procedures to build trust in their operation.

What is the solution?

The Practical Policy Working Group produced a document describing best practices and a set of machine actionable policy templates covering 11 highly relevant policy areas including contextual metadata extraction, data access control, data backup, data formal control, data retention, disposition, integrity (incl. replication), notification, restricted searching, storage cost reports, and use agreements.



Produced by: **Practical Policy (PP) WG**

<https://rd-alliance.org/groups/practical-policy-wg.html>

What is the impact?

Computer actionable policies are used to enforce management, automate administrative tasks, validate assessment criteria, and automate scientific analyses.

Benefits of using policies include minimization of the amount of labor needed to manage a collection, the ability to publish to the users the rules that are being used, and the ability to automate process management.

Find out more about the PP
WG Recommendation



March 2020



Persistent Identifier Type Registry

The Challenge:

Defining standard core PID information types to enable simplified verification of data identity and integrity

What is the solution?

The PID Information Types Working Group identified the essential types of information associated with persistent identifiers, defined and registered a number of core persistent identifiers (PIDs) information types. The working group developed a conceptual model for structuring the typed information, an application programming interface to access to typed information and a demonstrator implementing the interface.



Produced by: **PID Information Types (PIT) WG**

<https://rd-alliance.org/groups/pid-information-types-wg.html>

What is the impact?

In complex data domains, unique and persistent identifiers (PIDs) are at the core of proper data management and access, providing a unified access method to all PID service users and the automatic processes to deal with heterogeneous data allowing them to fully exploit the huge amount of data available. Registered PID Information Types are one of the fundamental building blocks to enable semantic interoperability of data.

Developers are able use a simple API to deal with PID related information, such as checksum verification, to assess data identity and integrity, thus enabling software development to become more efficient.



Find out more about the PIT
WG Recommendation

March 2020



Data Type Model & Registry

The Challenge:

Ensuring data producers classify their data sets in standard data types, allowing data users to automatically identify instruments to process and visualise the data

What is the solution?

The Data Type Registries Working Group established open Data Type Registries (DTRs) as public repositories where developers and researchers can add their data types and references to tools that can operate on them. Datasets are associated to data types defined in DTRs, thus enabling researchers to access and visualise data regardless of its specific type, facilitating sharing and re-use.



Produced by: **Data Type Registries (DTR) WG**

[www.rd-alliance.org/groups/
data-type-registries-wg.html](http://www.rd-alliance.org/groups/data-type-registries-wg.html)

What is the impact?

Researchers are able to process and/or visualize the data types registered in the DTRs without any prior knowledge of the data types. This increases efficiency in time and the level of cross-domain data interoperability and re-use.

Machines will be able to automatically process the data types registered in the DTRs. Machine readable DTRs pave the way to automatic processing in the data domain, which is becoming increasingly complex, without putting an additional load on the researchers.



Find out more about the DTR
WG Recommendation

March 2020



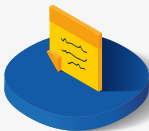
Basic Vocabulary of Foundational Terminology Query Tool

The Challenge:

Ensuring researchers apply a common core data model when organising their data and thus making data accessible and re-usable.

What is the solution?

Based on 21 data models presented by experts coming from different disciplines and 120 interviews and interactions with different scientists and scientific departments, the Data Foundation & Terminology Working Group has produced 5 inter-related reports and extracted a common core model for the proper organisation of data. Based on this common conceptualization a number of core terms were defined and registered to also help harmonisation of terminology.



Produced by: **Data Foundation and Terminology (DFT) WG**

[www.rd-alliance.org/groups/
data-foundation-and-terminology-wg.html](http://www.rd-alliance.org/groups/data-foundation-and-terminology-wg.html)

What is the impact?

Researchers from different disciplines will be able to more easily implement proper and FAIR compliant data organisations for example in repositories and also interact more easily with each other and come to a common understanding more rapidly, through increased cross disciplinary data exchange & interoperability. Developers will be able to design interoperable data management and processing software systems allowing data to be exchanged and integrated easily in particular in a cross-disciplinary setting. It will be easier to specify standard APIs to request useful information related to a specific Digital Object, thus accelerating the development time & enabling extensive data & software re-use.

Find out more about the DFT
WG Recommendation



March 2020



Summer Schools in Data Science and data sharing framework

The Challenge:

A framework to run a series of Summer Schools in Data Science and data sharing in low and middle income countries (LMICs).

What is the solution?

The CODATA/RDA Summer School in Data Science and Cloud Computing in the Developing World WG provides a framework to run a series of Summer Schools in Data Science and data sharing in low and middle income countries (LMICs) with the goal of addressing the gap in research data science skills that stops researchers from reaping the benefits of the data revolution.



Produced by: RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World WG

<https://www.rd-alliance.org/groups/rdacodata-summer-schools-data-science-and-cloud-computing-developing-world.html>

What is the impact?

The RDA/CODATA Working group is focused on consolidating the framework and the continued collaboration for the development and improvement of the curriculum and training materials, which include the principles and practice of Open Science, research data management and curation, the use of a range of data platforms and infrastructures, large scale analysis, statistics, visualisation and modelling techniques, software development.

The need for a consistent education in Research Data Science is increasingly paramount for many stakeholders, from scientists to funders to policy makers from many nations. All disciplines need to ensure that research is reproducible and that provenance is documented reliably and this requires a transformation in practice and the promotion of the necessary culture, practice and skills.

**RDA/CODATA Summer
Schools in Data Science
and Cloud Computing in
the Developing World WG
Recommendation**



March 2020



Research Data Repository Interoperability Primer

The Challenge:

Huge amounts of research data stored in a multitude of research data repository platforms can often only be used by a comparably small audience.

What is the solution?

On the one hand, this is caused by differences in semantics, underlying data models and metadata schemas, whose complexity and number prevents scientists from taking advantage of them. On the other hand, a lack of interoperability between research data repository platforms causes research data not to be used to their full potential. The goal of the RDA Research Data Repository Interoperability WG (RDRIWG) and of the Primer is to achieve consensus on an adoptable approach to facilitating research data repository interoperability for a defined set of initial use cases.



Produced by: **Research Data Repository Interoperability WG**

www.rd-alliance.org/groups/research-data-repository-interoperability-wg.html

What is the impact?

The Research Data Repository Interoperability Primer presents a set of initial use cases, as well as an overview of standards, technologies and tools that could be components of an agreed adoptable approach to facilitating interoperability between different research data repository platforms. Main focus is on machine-machine communication with the primary goal of enabling migration and replication.



Research Data Repository
Interoperability WG

March 2020



Recommendation on Research Data Collections

The Challenge:

Research data management practice requires not only to describe collections, but to make them actionable by automated processes to be able to cope with ever increasing amounts and volumes of data.

What is the solution?

The RDA Research Data Collection Recommendation provides a comprehensive model for actionable collections and a technical interface specification to enable client-server interaction. It also reports on first adoption and implementation efforts across communities and institutions and provides perspectives on the use of data types in connection with collection structures, highlighting pathways for possible future work.

Produced by: **Research Data Collections WG**

www.rd-alliance.org/groups/research-data-collections-wg.html

What is the impact?

The recommendation allows building collections within diverse domains and then sharing or expanding them across disciplines.

This should enable common tools for end-users and e-infrastructure providers. Individual disciplinary communities can directly benefit if such tools are made widely available, and cross-community data sharing can benefit from increased unification between collection models and implementations.



Research Data Collections WG

March 2020



Matrix of use cases and functional requirements for research data repository platform

The Challenge:

Gather and analyze research data use cases in the context of repository platform requirements

What is the solution?

Based on use cases, the matrix describes forty-four functional requirements identified for research data repository platforms and provides a score identifying relative importance. These functional requirement scores can be used to assess research data repository platforms and to prioritize functional requirements for development and adoption.



Produced by: **Repository Platforms for Research Data IG**

<https://www.rd-alliance.org/groups/repository-platforms-research-data.html>

What is the impact?

The matrix was envisioned to support the work of developers and service providers of repository software, repository implementers and managers, funding agencies, grant applicants, and researchers generating and using scientific data.

The RPRD IG's focus is the collection of experiences with selecting, implementing, and using specific research data repository platforms/products across institutions and domains, with reference to the aforementioned functional requirements. From this, the IG may identify potential gaps between needs and solutions in the current research data repository environment, and will seek opportunities to propose the formation of new RDA working groups to address them.

Matrix of use cases and functional requirements for research data repository platform supporting output



March 2020



The FAIRsharing Registry and Recommendations: Interlinking Standards, Databases and Data Policies

The Challenge:

With thousands of interoperability standards, databases and policies, researchers - and those involved in curating, preserving, publishing or regulating data - need help to make informed decisions on which metadata standards to use, which are implemented in a repository and/or recommended by a journal.

What is the solution?

Recommendations to guide producers of standards, databases, repositories and policies, and a registry to make their resources discoverable to prospective users, called FAIRsharing (fairsharing.org). This registry tracks the development and evolution of standards, their implementation in databases, and adoption in data policies from funders, journals and other organizations.



Produced by: **Force11 / RDA FAIRsharing WG**

[https://rd-alliance.org/group/
fairsharing-registry-connecting-data-policies-standards-databases.html](https://rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases.html)

What is the impact?

As of March 2018, we have 1168 standards, 1050 databases and 111 policies. Our aim is to ensure quality rather than quantity. A key part of this is working directly with the community, engaging the maintainers of each standard, database and policy to actively update the record for their resource on FAIRsharing. FAIRsharing has already been used and adopted by a number of journal publishers (e.g. PLOS, Springer Nature, F1000), research data management programmes (e.g. ELIXIR and NIH), community and infrastructure groups (e.g. Biodiversity TDWG), and research and funding programmes (e.g. JISC) (see more at fairsharing.org/communities).

This WG is a joint effort between the RDA and Force11, the initiative that fostered the development of the FAIR principles. Starting in the life sciences domain, this RDA/Force11 WG has now expanded to be cross-discipline/domain. Join us!

Find out more about the
FAIRSharing Registry
recommendations



March 2020



RDA 11 Plenary Meeting

21-23 March 2018
Berlin, Germany

From Data to Knowledge

www.rd-alliance.org/plenaries/rda-eleventh-plenary-meeting-berlin-germany

The Eleventh RDA Plenary Meeting took place from the 21st to the 23rd March 2018 in Berlin, Germany.

Under the theme "From Data to Knowledge", the Plenary meeting welcomed the participation of all data scientists, experts and practitioners engaged in the advancement of data-driven science and economy.



COLLECT THEM ALL!

RDA Plenary





Data Discovery Paradigms: User Requirements and Recommendations for Data Repositories

The Challenge:

When there are more and more data open and available through data repositories, it becomes ever more challenging for researchers to find relevant data. This supporting output presents a work that addresses the data findability problem.

What is the solution?

By gathering and analysing use cases, the IG has identified a core set of functional requirements for a data repository and proposed ten recommendations with exemplar implementations to meet the requirements. This work can help data repositories improve search and discovery of their data.



Produced by: **Data Discovery Paradigms Interest Group**
www.rd-alliance.org/groups/data-discovery-paradigms-ig

What is the impact?

The work focuses on enabling and improving the methods and tools by which users find data in repositories. Recommendations can be used as guidelines or as a reference for developers, project and product managers at data repositories, as well as researchers who are involved in developing and improving data repositories, community platforms, or interfaces to data collections.

Find out more about the Data
Discovery Paradigms Interest
Group



March 2020



Eleven Quick Tips for Finding Research Data

The Challenge:

Considering the vast amount and variety of data are made available, finding relevant data to meet a research need is increasingly a challenge. Over the past decades, science has experienced rapid growth in the volume of data available for research - from a relative paucity of data in many areas to what has been recently described as a data deluge.

What is the solution?

The Data Discovery Paradigms IG output offers eleven quick tips that researchers can follow to more effectively and precisely discover data that meets their specific needs. The articles include an list of useful resources and examples that can be easily leveraged by all types of users.



Produced by: **Data Discovery Paradigms Interest Group**

<https://www.rd-alliance.org/group/data-discovery-paradigms-ig/outcomes/ten-quick-tips-finding-research-data>

What is the impact?

Encouraging wide adoption to educate and train research students and early career researchers, and to help researchers more effectively and precisely discover data that meets their specific needs.

Find out more about the Eleven
Quick Tips for Finding Research
Data



March 2020



Income Streams for Data Repositories

The Challenge:

Basic funding of data infrastructure may not keep pace with increasing costs. There is a need, therefore, to consider alternative cost recovery options and a diversification of revenue streams. In short: who will pay for public access to research data?

What is the solution?

The RDA/WDS Interest Group Publishing Data Cost Recovery for Data Centres aims to contribute to strategic thinking on cost recovery by conducting research to understand current and possible cost recovery strategies for data centres. As the main body of work, the Interest Group has conducted a survey of over twenty data centres around the globe, and in different domains.



Produced by: **RDA/WDS Publishing Data Cost Recovery for Data Centres IG**

www.rd-alliance.org/groups/rdawds-publishing-data-cost-recovery-data-centres.html

What is the impact?

The survey consisted of structured in-depth interviews with high-level representatives of these organisations and was focused on identifying various existing approaches to cost recovery, the range of income streams available and current and possible business models.

The principal beneficiaries of this work will be data centre managers on the one hand and research infrastructures on the other. Each of these stakeholders will gain a valuable insight into alternative options for cost recovery, substantiated by the survey results.



Find out more about the Income
Streams for Data Repositories

March 2020

The Digital Frontiers of Global Science

INTERNATIONAL
DATA WEEK 2018
5 - 8 NOVEMBER
WWW.INTERNATIONALDATAWEEK.ORG

Co-organized by the ICSU World Data System (WDS), the ICSU Committee on Data for Science and Technology (CODATA) and the Research Data Alliance (RDA), IDW 2018 combines the 12th RDA Plenary Meeting, the bi-annual meeting of the research data community, and SciDataCon 2018, the scientific conference addressing the frontiers of data in research.



With the theme of 'The Digital Frontiers of Global Science', this landmark event was a rich week of science and data, featuring world renowned keynote speakers, plenary panels and discussions, and the presentation of high quality research and practical working sessions for international collaborations.

www.internationaldataweek.org



COLLECT THEM ALL!





Research Data Repository Interoperability WG Final Recommendations

The Challenge:

Huge amounts of research data stored in a multitude of research data repository platforms can often only be used by a comparably small audience. This is caused by differences in semantics, underlying data models and metadata schemas, whose complexity and number prevents scientists from taking advantage of them. As well as a lack of interoperability between research data repository platforms. The goal of the RDA Research Data Repository Interoperability WG (RDRIWG) was to achieve consensus on an adoptable approach to facilitating research data repository interoperability for a defined set of initial use cases.



Produced by: **Research Data Repository Interoperability WG**

<https://www.rd-alliance.org/groups/research-data-repository-interoperability-wg.html>

What is the solution?

This output provides recommendations with respect to an interoperable packaging and exchange format (based on the BagIt specification) for digital content. In order to achieve a basic level of interoperability of such exchange packages, DataCite metadata are recommended to be included in each package. The presented recommendations document describes the exchange format itself together with adoption guidelines and information about related efforts.

What is the impact?

Once implemented, compliant packages can be used to migrate or replicate digital content between research data repository platforms or for preservation purposes.

Find out more about this
recommendation:



March 2020



A survey of current practices in data search services

The Challenge:

A critical element of any data repository is its search system, which determines the discoverability of a repository's holdings. However, it is unknown how the catalogues of each have been configured and what the benchmarks are of the services used for making data searchable.

What is the solution?

This survey was conducted between June and July 2018. It examined the practices that data repositories employ in helping users search their holdings and demonstrate the value of looking across disciplines at the common data discovery issues, such as relevancy, that all data repositories face.



Produced by: **Data Discovery Paradigms IG**

<https://www.rd-alliance.org/groups/data-discovery-paradigms-ig>

What is the impact?

The study finds that repositories desire guidelines for improving relevancy ranking in their systems, with small repositories having the greatest need. Repositories understand that their search systems need to be evaluated and improved, but often lack the resources (time and/or expertise) to explore and evaluate the available options. The study concludes that there is an opportunity for people working in the search space to collaborate, to build test collections and other efforts that offer the greatest improvements in search services at the lowest cost.



Find out more about this survey:

March 2020



Addressing the Gaps: Recommendations for Supporting the Long Tail of Research Data

The Challenge:

Large structured datasets in well-established disciplines are more likely to adopt unified and standardized formats that are disciplinarily defined and accepted. Similarly well established disciplines tend to have common and understood workflows, where long tail datasets, which vary radically in source, discipline, size, subject, provenance, funding, format, longevity, location and complexity, are less likely to adhere to common standards. Ensuring that wide distributed and diverse long-tail data is discoverable and stored in appropriate formats to facilitate reuse is challenging.



Produced by: **Long tail of research data IG**

<https://www.rd-alliance.org/groups/long-tail-research-data-ig.html>

What is the solution?

The RDA “Long Tail of Research Data Interest Group” has been assessing the situation of long tail data. This document provides seven recommendations for a variety of stakeholders, including governments, funders, research institutions and researchers to help improve the current approach to managing long tail data.

What is the impact?

“Long Tail of Research Data Interest Group” calls on the community to work together to create necessary and sufficient conditions to properly steward long tail valuable research outputs for future generations of researchers.

Find out more about this
recommendation:



March 2020



Persistent identifiers: Consolidated assertions

The Challenge:

Despite 20 years of discussions about PIDs, there still exists much confusion and lack of knowledge about persistent identifiers (PIDs) in many research communities. In part, this is the consequence of the vast number of documents and reports on the PID topic in existence, many of which focus on very specific aspects, or use different and sometimes incorrect terminologies.

What is the solution?

To help bring more clarity, it was therefore agreed in GEDE and the RDA Data Fabric Interest Group to join forces to produce a compilation of relevant assertions (extracted statements from original sources), supplemented with context information and explanations as required, and leading up to a summary of agreements and disagreements reflecting the convergence of understanding.



Produced by: **Data Fabric Interest Group**

<https://www.rd-alliance.org/group/data-fabric-ig.html>

What is the impact?

With this document GEDE is happy to help demystify PIDs, overcome confusion and create bridges between the various discipline

Find out more about Persistent identifiers: Consolidated assertions



March 2020



Summary of Virtual Layer Recommendations

The Challenge:

Data volumes and complexity of scientific data are increasing rapidly, and cross-border re-usage of such data is becoming more common in most research fields. Open data as an agreed default model, backed by FAIR principles to ensure data is 'Findable, Accessible, Interoperable, and Re-usable', together with Data Management Plans to encourage stronger curation and stewardship have all emerged as community policies underpinning research data sharing without barriers. Presently, the number of research and data infrastructures worldwide is growing, with many similar components re-invented in different variations. Due to this fragmentation, the costs of building and maintaining infrastructure are higher than they should be; as are costs for data re-use. This trend must be counteracted by identifying common components and means of interoperability as a new conceptual framework that creates a new global momentum on data infrastructure interoperability, akin to the creation of the Internet itself.



Produced by: **RDA Data Fabric Interest Group**

<https://www.rd-alliance.org/group/data-fabric-ig.html>

What is the solution?

Computing and storage components have existed from the beginning of modern computing. The Internet Protocol Addressing system identifies all devices in the network so they can exchange messages using protocols and registries. Together these two layers provide us today with almost unlimited cloud computing capacity with near universal access to dynamically configurable, shared compute and storage capabilities. Yet, to support data-driven science, this is no longer sufficient. We must address data organisation, typing and re-use facilitation to be able to fully realise the value of data, network, compute and storage. We are now in the phase where we must provide a new layer of network wide virtualisation that interlinks data and other artefacts with the help of PIDs, new protocols and new registries.

What is the impact?

A high-level conceptual framework to support Digital Object management and service development.



**Find out more about this
Supporting Output**

March 2020



Recommendation on PID Kernel Information

The Challenge:

Global middleware infrastructure is insufficient for robust data identification, discovery, and use. While infrastructure is emerging within sub-ecosystems such as the DOI ecosystem of services purposed for data and literature objects (i.e., DataCite, CHORUS, CrossRef), in general the layers of abstraction that have made the Internet so easy to build on, is lacking for data especially for computer (machine) automated services. The PID Kernel Information recommendation wants to advance a small change to middleware infrastructure by injecting a tiny amount of carefully selected metadata into a Persistent ID (PID) record.



Produced by: **PID Kernel Information WG**

<https://www.rd-alliance.org/groups/pid-kernel-information-wg>

What is the solution?

The PID Kernel Information working group determined which from amongst thousands of relevant metadata elements are suitable to embed in the PID record. They produced a set of guiding principles, architectural considerations, use cases and a fundamental metadata schema to manage information in Persistent Identifier records for scalable middleware infrastructure and automated processes.

This recommendation lays out principles to guide in the identification of information suitable for inclusion in the PID record. The recommendation also includes a draft profile with illustrating examples and cases for adoption in practice.

What is the impact?

This carefully chosen and placed information has the potential to stimulate development of an entire ecosystem of third party services that can process the billions of expected PIDs and do so with more information at hand about an object (no need for costly link following) than just a unique ID.



Find out more about this
Recommendation

March 2020



RDA/TDWG Attribution Metadata Working Group: Final Recommendations

The Challenge:

Research collections are an important tool for understanding the Earth, its systems, and human interaction. Despite the importance of collections, many are not maintained or curated as thoroughly as they should. Part of the reason for this is the lack of professional reward for curation, maintenance, or collection.



Produced by: **RDA / TDWG Metadata Standards for attribution
of physical and digital collections stewardship WG**

<https://www.rd-alliance.org/groups/metadata-standards-attribution-physical-and-digital-collections-stewardship.html>

What is the solution?

To address this gap in metadata attribution, this Working Group recommends the use of PROV entities and properties to link people (Agent), the curatorial actions they perform (Activity), and the digital or physical objects they are curating (Entity). A separate technical document gives specific examples of three use cases as RDF Turtle representations and diagrams.

What is the impact?

This Recommendation supports standardized metadata for attributing work and tracking provenance in the curation and maintenance of research collections.

Find out more about this
Recommendation



March 2020



13th Plenary Meeting

Philadelphia, Pennsylvania USA
2-4 April 2019

www.rd-alliance.org/plenaries/rdas-13th-plenary-meeting-philadelphia-us



Alfred P. Sloan
FOUNDATION



National Science
Foundation

GORDON AND BETTY
MOORE
FOUNDATION

NIST

SPRINGER NATURE

With the theme, "With Data Comes Responsibility," P13 explored how the international, inter- and transdisciplinary nature of RDA has helped to lay a foundation for global partnerships and addressing responsibility throughout the data-lifecycle and across all domains.



COLLECT THEM ALL!

RDA Plenary





RESEARCH DATA ALLIANCE

M DATA K THE DIFFERENCE S



RDA
14th Plenary

Helsinki | Espoo
Finland
23-25 October 2019



Aalto University



FEDERATION OF FINNISH LEARNED SOCIETIES



FMI

Data has enormous and diverse potential to change how we work, make decisions and empower citizens. The 14th RDA plenary will explore the extensive ways data can make the difference by bringing together diverse group of experts from all around the globe. During the plenary we will share and learn together and advance our abilities to use data in making a positive impact in the world to tackle its complex challenges.



COLLECT THEM ALL!

RDA Plenary





CODATA Data Science Journal Special Collection on RDA Results

The CODATA Data Science Journal special collection solicits high quality papers describing the latest results of RDA working groups or interest groups that have recently produced an output, including recommendation and associated use cases that could highlight the added value of RDA work in the data related fields.

The scope of the special collection

The scope of this special collection is to ultimately become a point of reference for experts from the international community that are committed to enabling data sharing, exchange, and interoperability by showcasing the various RDA results and how these have been implemented across disciplines, organisations and geographies.

Specific tools, code, best practices, standards, surveys, recommendations, reports, and other results will be brought to a larger audience via the peer-reviewed, open access Data Science Journal.

Submission process

The Data Science Journal allows for a continuous submission process. Articles are made available as soon as they have gone through the review process.

Publication fees

The “RDA Europe 4.0” project will support the publication of a maximum of 30 articles. Publication fees (i.e. Article Processing Charges) will be covered by the project on a first come first served basis. Please note that funding is available until 15 April 2020.

Continuous submission open at:





A curriculum for foundational Research Data Science skills for Early Career Researchers

The Challenge:

Giving Early Career Researchers (ECR's) the foundational skills in Data Science to work with their data.



Produced by: **RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World WG**

<https://www.rd-alliance.org/groups/rdacodata-summer-schools-data-science-and-cloud-computing-developing-world.html>

What is the solution?

This recommendation describes the curriculum and example materials to give Early Career Researchers (ECR's) the foundational skills in Data Science to work with their data. This curriculum combines technical skills, such as Software Carpentry with responsible research practices such as Open and Responsible Research and Research Data Management. This curriculum includes a set of curriculum specifications for the modules run in this curriculum, an example timetable for a 10 day intensive training event, a diagram to show how these modules are connected, and a spreadsheet of links to example materials that implements this.

What is the impact?

At the end of 2019, approximately 400 students had been taught on four continents using the curriculum.



Find out more about this
Recommendation

March 2020



RDA DMP Common Standard for Machine-actionable Data Management Plans

The Challenge:

Data Management Plans are free-form text documents describing the data that is used and produced during the course of research activities. They specify where the data will be archived, which licenses and constraints apply, and to whom credit should be given, etc. The workload and bureaucracy often associated with traditional DMPs can be reduced when they become machine-actionable.



Produced by: **DMP Common Standards WG**

<https://www.rd-alliance.org/groups/dmp-common-standards-wg>

What is the solution?

The RDA DMP Common Standards Working Group has developed an application profile that allows to express information from traditional DMPs in a machine-actionable way. It allows for automatic exchange, integration, and validation of information provided in DMPs.

The RDA Recommendation document summarises the application profile. The full specification can be found: <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>

What is the impact?

Allows representing Data Management Plans in a machine-actionable way, to enable exchange of information between systems acting on behalf of stakeholders involved in the research lifecycle, such as researchers, funders, repository managers, ICT operators, data stewards, etc. It also helps in automating typical data management tasks, thus contributes to a reduction of workload imposed on the stakeholders.



Find out more about this
Recommendation

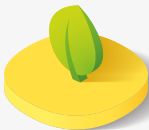
March 2020



39 Hints to Facilitate the Use of Semantics for Data on Agriculture and Nutrition

The Challenge:

Enhancing data interoperability in agriculture by promoting the use of semantics for agricultural data. For this purpose the RDA Agrisemantics Working Group performed an analysis of the current landscape of use of semantic resources with agricultural data, based on first-hand experience of our group members, as well as on bibliometric investigation, and an analysis of existing repositories. Despite the many possible applications of semantics and the interest shown by research and industry alike, their actual implementation and use is lagging behind. A large number of resources are not in machine-readable formats or do not have public APIs, while those available are often used beyond their intended area of use, possibly leading to problems. There is a strong need for “lifting” semantic resources to the web, and create new ones appropriate to their intended use.



Produced by: Agrisemantics WG

<https://www.rd-alliance.org/groups/agrisemantics-wg.html>

What is the solution?

Based on the landscape analysis of use of semantic resources with agricultural data, a set of high level recommendations were defined for future activities and implementations to facilitate the use of semantics for data on agriculture and nutrition. These recommendations were arranged according to the skill sets and roles that are in the best position to implement them. This resulted in 39 hints for researchers and practitioners to extend them according to their area of expertise.

What is the impact?

The document contains a number of highlights on issues to be addressed in order to make the use of semantics more straightforward and widespread. The RDA Agrisemantics Working Group expects the document to help share awareness about semantics among experts in different areas and support strategic allocation of resources in future development.



Find out more about this
Recommendation

March 2020



WDS/RDA Assessment Of Data Fitness For Use WG Outputs

The Challenge:

The increasing availability of research data and its evolving role as a first class scientific output in the scholarly communication requires a better understanding of and the possibility to assess data quality, which in turn can be described as conformance of data properties to data usability or fitness for use. These properties are multifaceted and cover various aspects. The compliance of a data repository or data center providing datasets - for example with certification requirements - could serve as a useful proxy. Currently, there is a fairly good understanding on how to certify the quality of a data center / repository as a whole, but there is no generally acknowledged concept for assessment of data usability of individual datasets. Assessing fitness for purpose and making a decision whether to reuse a dataset is not straightforward. This situation reduces the chances of shared data being reused and in case of reuse could decrease the reliability of research results.



Produced by: **LWDS/RDA Assessment Of Data Fitness For Use WG**

<https://www.rd-alliance.org/groups/assessment-data-fitness-use>

What is the solution?

Criteria for research dataset fitness for use compared against the CoreTrustSeal requirements and FAIR principles. Through this comparison of criteria the WDS/ RDA Assessment Of Data Fitness For Use Working Group determined that a CoreTrustSeal certified repository's data holdings would meet several aspects of dataset fitness for use. Evaluating metadata and data completeness in an automated fashion is not feasible at this time, therefore a manual evaluative process for research datasets that would build on the CoreTrustSeal repository certification process was developed.

A checklist for evaluation of dataset for fitness for use meant to supplement the CoreTrustSeal Repository Certification process, and is based on the data fitness for use criteria. This manual evaluative process would be conducted by a repository manager or an external entity such as a CoreTrustSeal repository evaluator for a sample (6-12) of individual data sets within the repository.

What is the impact?

- Improved communication of data fitness for use
- Improved and standardized data publication services
- Improved reliability and efficiency in the reuse of research data



Find out more about this
Recommendation

March 2020



Engaging Researchers with Data Management: The Cookbook

The Challenge:

Effective Research Data Management (RDM) is a key component of research integrity and reproducible research, and its importance is increasingly emphasised by funding bodies, governments, and research institutions around the world. However, many researchers are unfamiliar with RDM best practices, and research support staff are faced with the difficult task of delivering support to researchers across different disciplines and career stages. What strategies can institutions use to solve these problems?



Produced by: **Libraries for Research Data Interest Group**
<https://www.rd-alliance.org/groups/libraries-research-data.html>

What is the solution?

Engaging Researchers with Data Management is a collection of 24 case studies, drawn from institutions across the globe, that demonstrate clearly and practically how to engage the research community with RDM. These case studies together illustrate the variety of innovative strategies research institutions have developed to engage with their researchers about managing research data. Each study highlights the essential ingredients that led to its success and challenges encountered along the way. By interviewing key staff about their experiences and the organisational context, the authors of this book have created an essential resource for organisations looking to increase engagement with their research communities.

What is the impact?

Provides examples of 24 innovative, searchable and re-usable case studies on how institutions can effectively engage with researchers about research data.



Find out more about this
Recommendation

March 2020



Tromsø recommendations for citation of research data in linguistics

The Challenge:

Language and linguistics datasets are often not cited, or cited imprecisely, because of confusion surrounding the proper methods for citing them.



Produced by: **Linguistics Data IG**

<https://www.rd-alliance.org/groups/linguistics-data-ig>

What is the solution?

For the use of researchers and scholars in the field working with datasets, these recommendations propose components of data citation for referencing language data, both in the bibliography and in the text of linguistics publications. The recommendations are intended to be guidelines rather than strict rules, and are based on the Austin Principles, the FORCE11 and Research Data Alliance Joint Declaration of Data Citation Principles, and the Reproducible Research in Linguistics position statement

What is the impact?

Provides guidance for citing linguistic data in publications to researchers, academic publishers and resource providers.

Find out more about this
Recommendation



March 2020

The logo for FIM4R consists of three horizontal bars: a grey bar on the left, a black bar in the middle, and a blue bar on the right. The text "FIM4R" is positioned above these bars.

FIM4R



Federated Identity Management for Research Collaborations

The Challenge:

Identifying technology-agnostic requirements from an ever-evolving Federated Identity Management (FIM) landscape, which has shifted from its original conception of an arrangement that allows subscribers from multiple organisations to use the same identification data to obtain access to the secured resources of all organisations in the group, to an ecosystem which now comprises an increasing variety of technologies provided by participating organisations.



Produced by: **Federated Identity Management IG**

<https://rd-alliance.org/groups/federated-identity-management>

What is the solution?

Updating a previous white paper on the topic, input from a broad range of research communities on their interaction with FIM systems was sought. A curated list of requirements and recommendations was gathered through discussion with these research communities, together with analysis of the data obtained as part of a survey of contributing communities. This white paper expands on the role of Research Community Proxy and FIM evolution in the context of current proxy architecture.

What is the impact?

FIM stakeholders are able to coordinate activity to enable research use cases, based on the set of recommendations outlined, to achieve the common vision of Federated Identity Management. Each recommendation is expected to have broad positive impact on the viability and value of Federated Identity Management for future support of research communities. A mapping of groups which are mentioned in connection with each recommendation is also provided.



Find out more about this
Recommendation

March 2020



Results of an Analysis of Existing FAIR Assessment Tools

The Challenge:

The aspirational nature of the FAIR data principles and their rapid adoption at international level has led to a wide range of interpretations of FAIRness since the principles do not strictly define how to achieve a state of FAIRness but rather describe a continuum of features, attributes and behaviours that move a digital object closer to that goal. As a result, a number of differing methodologies to assess FAIRness have been developed already and relevant work is under way by various groups. Due to the lack of a common set of core assessment criteria for FAIRness, researchers and organisations cannot evaluate the readiness and implementation level of their datasets vis-à-vis the FAIR data principles in a coherent way. In addition, research performing organisations and data infrastructures cannot develop or follow a minimum set of shared guidelines to climb up the ladder of FAIR because of the increased heterogeneity of the offered FAIR metrics tools.



Produced by: **FAIR Data Maturity Model WG**

<https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>

What is the solution?

As a landscaping exercise, the editorial team of the FAIR Data Maturity Model Working Group analysed current and existing approaches related to FAIR self-assessment tools. The analysis was made based on publicly available documentation and an online survey. Questions and options stemming from these different approaches were classified according to the FAIR principles/facets. Comments were collected and incorporated. This resulted in five slide decks, combined in this output, that make up this preliminary analysis.

What is the impact?

This document provides an overview of a number of existing FAIR assessment tools, listing the indicators used in these tools to assess the FAIRness of a data set. This is useful if you wish to discover and compare existing FAIR assessment tools and the questions that are being asked. The analysis is a first step to identify the FAIR maturity criteria, by the end of 2020.



Find out more about this
Recommendation

March 2020

15th Research Data Alliance Plenary Meeting



RESEARCH DATA ALLIANCE

18-20
MARCH
2020

MELBOURNE
AUSTRALIA



RESEARCH DATA ALLIANCE



Australian Research Data Commons

The Research Data Alliance (RDA) 15th Plenary Meeting “Data for Real-World Impact”, will take place from 18 - 20 March 2020 in Melbourne, Australia. The event will be hosted by CSIRO (Commonwealth Scientific and Industrial Research Organisation) with the support of the Australian Research Data Commons (ARDC). The RDA Plenary meeting will be organised in conjunction with the annual Collaborative Conference on Computational and Data Intensive Science (C3DIS). All of these professionals share a common interest: the production of breakthrough science outcomes that require large scale, complex computations and analyses to be carried out over massive data sets. We are hoping for great interactions between the two communities through the co-location of the two events.



COLLECT THEM ALL!

RDA Plenary

