



# Les recommandations et produits de la RDA dans le cycle de vie des données

Françoise Genova

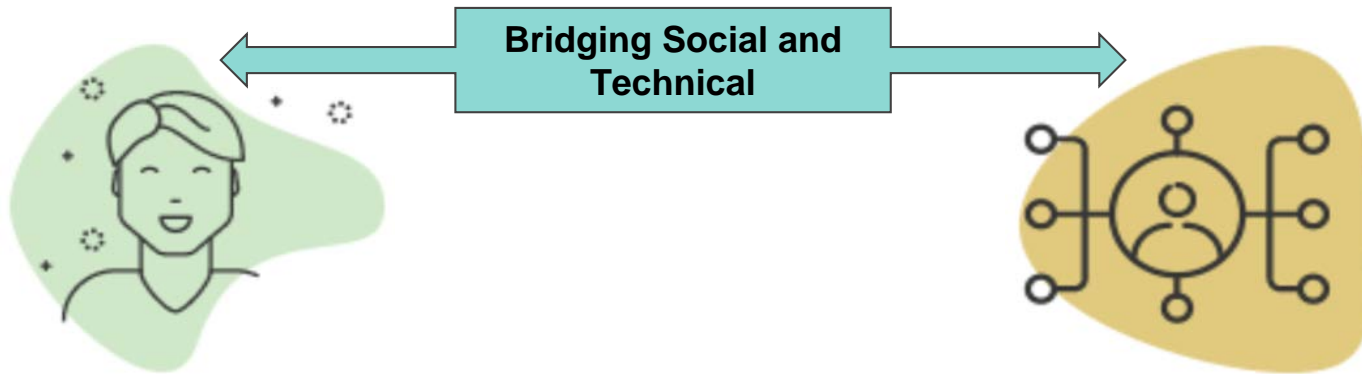
Exposé repris pour l'essentiel de l'exposé de A. Juehne – Secretariat

[https://www.rd-alliance.org/sites/default/files/attachment/RDA%20Outputs\\_RDA%20Adoption%20Business%20Session\\_final-2.pptx](https://www.rd-alliance.org/sites/default/files/attachment/RDA%20Outputs_RDA%20Adoption%20Business%20Session_final-2.pptx)



# ***Adopting RDA Recommendations and Outputs Across the Research Data Lifecycle***

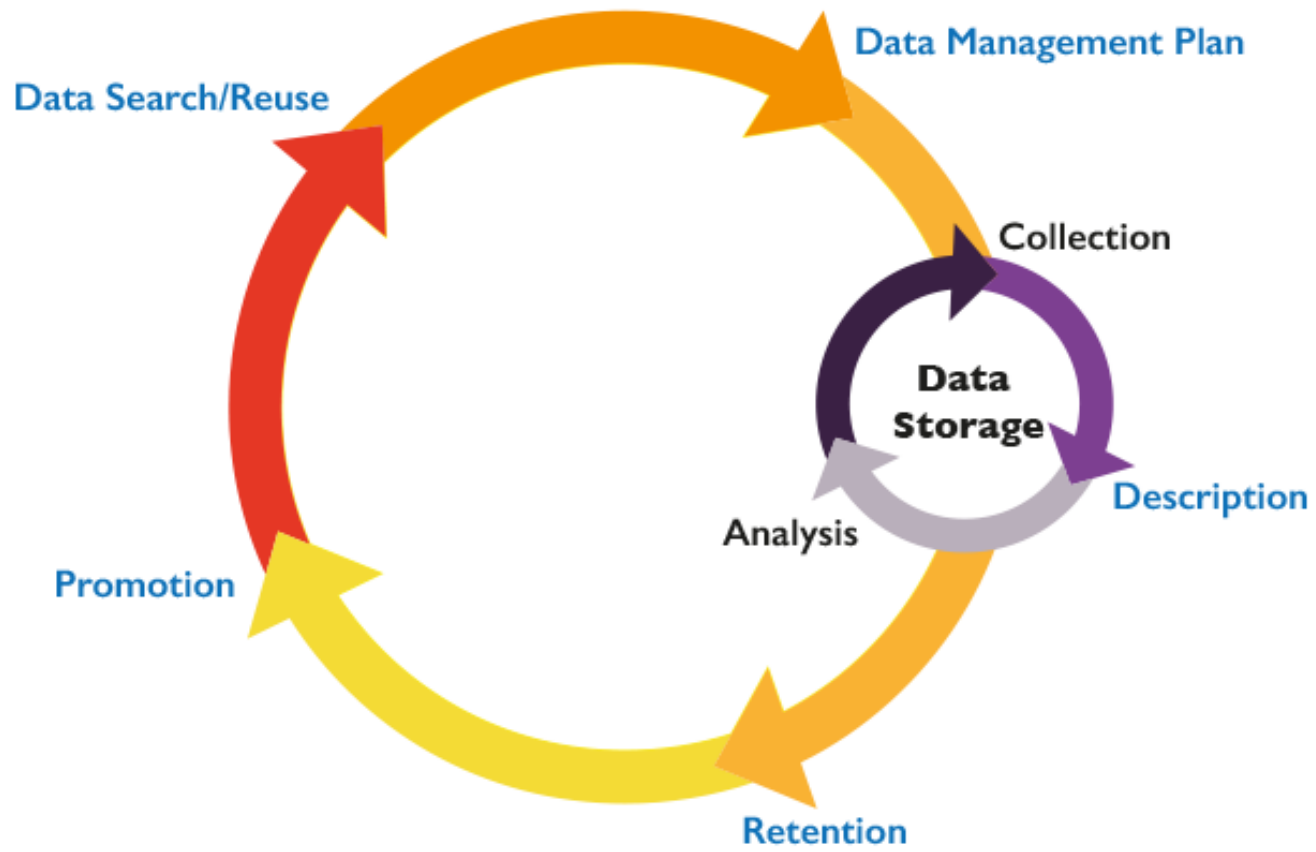
RDA Secretariat  
RDA P13 Business and Outputs Session  
April 2, 2019



- Preservation and stewardship
- Improving usefulness of data
- Data sharing culture
- Education and training
- Data security, legal interoperability, and compliance

- Data discovery
- Provenance, reuse, and interconnection
- Issues of scale
- Governing and maintaining infrastructure
- Data access and interoperability

# The Research Data Management Lifecycle



---

# Data Management

---

- DMP Common Standards
- 23 Choses: des bibliothèques pour les données de recherche



# DMP Common Standards

*Problem:* Data management plans are currently static objects which cannot evolve throughout the course of a research data lifecycle.

*Goal:* Develop best practices and tools for **non-static, machine-readable data management plans** which can evolve throughout the research data lifecycle, as well as be machine-readable by collaborators and stored with the data



## 23 Choses: des bibliothèques pour les données de la recherche

Un aperçu des ressources et des outils pratiques, gratuits et accessibles en ligne que vous pouvez commencer à utiliser dès aujourd'hui pour les intégrer dans vos activités de bibliothécaire.

Partage sans barrières des données de la recherche

### Ressources d'apprentissage

Les bibliothécaires apprennent comment appliquer les principes de la bibliothéconomie à la résolution des problèmes et à fournir de nouveaux services autour des données de la recherche.

1. "Top ten" des recommandations du LIBER destinées aux bibliothèques qui débutent dans la gestion des données de la recherche, <http://bit.ly/RDAthing1>
2. Présentation et cartographie de concepts pertinents dans e-Science Thesaurus, <http://bit.ly/RDAthing2>
3. Comprendre la vie des données de la recherche grâce au DCC Curation Lifecycle Model, <http://bit.ly/RDAthing3>
4. MANTRA, tutoriels en ligne pour les bibliothécaires,

Ressources d'apprentissage  
Référénts et sensibilisation pour les données  
Plans de gestion des données  
Littératie des données  
Citer les données  
Attribution de licences pour les données et protection de la vie privée  
Conservation numérique  
Entrepôts de données  
et Communauté de pratique

*...pour aider les bibliothécaires à s'engager dans la gestion des données de la recherche !*

### Référénts & sensibilisation pour les données

Les bibliothécaires répondent aux questions des usagers concernant les données et mènent des actions de sensibilisation pour évaluer les besoins de leurs chercheurs et étudiants.

7. Initier une conversation avec un chercheur sur les données en

Librarian Outreach Kit pour aider vos bibliothécaires,

10. Réponses des experts, sur le forum DataQ, aux questions sur les données, <http://bit.ly/RDAthing10>

### Plans de gestion de données

Les bibliothécaires se familiarisent avec les exigences des financeurs et avec les actions de conseil à mener auprès des chercheurs pour les aider à écrire et à mettre en œuvre des plans de gestion de données appropriés.

11. Un exemple est le DMPTool qui recense les exigences des financeurs aux États-Unis et qui construit un plan en demandant au chercheur de répondre à une série de questions. D'autres pays comme le Royaume-Uni et le Canada ont des outils semblables, <http://bit.ly/RDAthing11>

### Littératie des données

11 Available Translations: <https://rd-alliance.org/group/libraries-research-data-ig/outcomes/23-things-libraries-research-data-supporting-output>










---

# Collecting Data

---

- › Eleven Quick Tips for Finding Research Data
- › Dynamic Data Citation
- › Reproducible Workflows for Health Data

# Eleven quick tips for finding research data

Kathleen Gregory , Siri Jodha Khalsa , William K. Michener , Fotis E. Psomopoulos , Anita de Waard ,  
Mingfang Wu  

Published: April 12, 2018 • <https://doi.org/10.1371/journal.pcbi.1006038>

**Many levels of entry!**

**Start where you are  
comfortable and  
grow from there.**

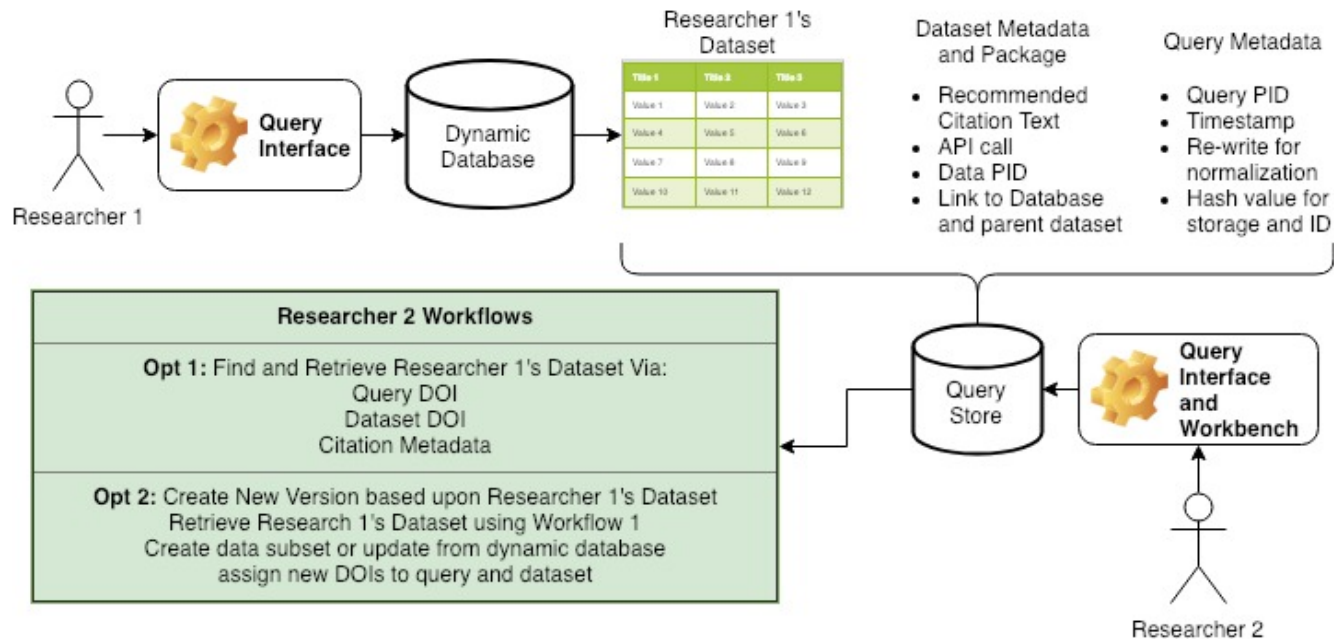
- › Tip 1: Think about the data you need and why you need them.
- › Tip 2: Select the most appropriate resource.
- › Tip 3: Construct your query strategically.
- › Tip 4: Make the repository work for you.
- › Tip 5: Refine your search.
- › Tip 6: Assess data relevance and fitness -for -use.
- › Tip 7: Save your search and data- source details.
- › Tip 8: Look for data services, not just data.
- › Tip 9: Monitor the latest data.
- › Tip 10: Treat sensitive data responsibly.
- › Tip 11: Give back (cite and share data).



## Dynamic Data Citation

- » Data Versioning: For retrieving earlier states of datasets the data needs to be versioned. Markers shall indicate inserts, updates, and deletes of data in the database.
- » Data Timestamping: Ensure that operations on data are time-stamped, i.e. any additions, deletions are marked with a timestamp.
- » Data Identification: The data used shall be identified via a PID pointing to a time-stamped query, resolving to a landing page

# Scalable Dynamic Data Citation



RDA Data Citation WG: <https://rd-alliance.org/groups/data-citation-wg.html>

RDA WG Case Statement: <https://www.rd-alliance.org/filedepot?fid=102>

Executive Summary of Output:

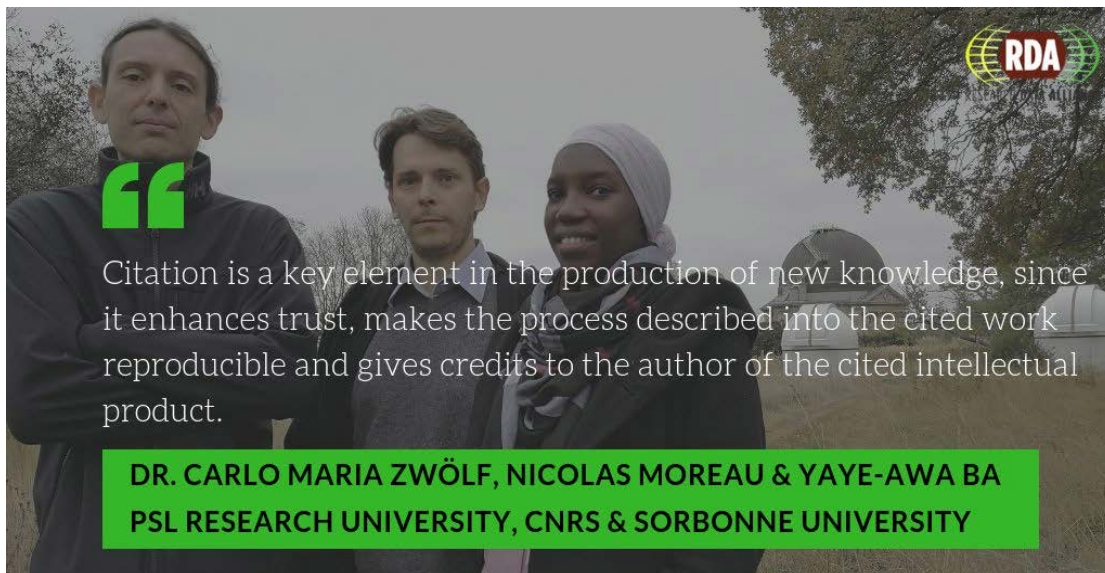
<https://docs.google.com/document/d/1SUr28B30Gg4yNNHfmztRawP7zLJqmx2lpEZSceeBMs/edit?usp=sharing>

# Adoption et Implémentation

*“Pour résoudre le problème, l’implémentation doit être incluse dans le processus, afin de s’assurer que les vrais problèmes sont traités. La clé, c’est la résolution ouverte des problèmes.”*

Nous avons réussi à mettre en œuvre les nouveaux paradigmes de citation des données établis par la RDA (Data Citation et Scholix) sur l’infrastructure distribuée VAMDC. Cela a permis d’éliminer les obstacles techniques qui empêchaient la citation automatique des données et la transmission des crédits bibliographiques pour les données extraites par le VAMDC.

Amélioration de la citation des données dynamiques dans des infrastructures asynchrones distribuées



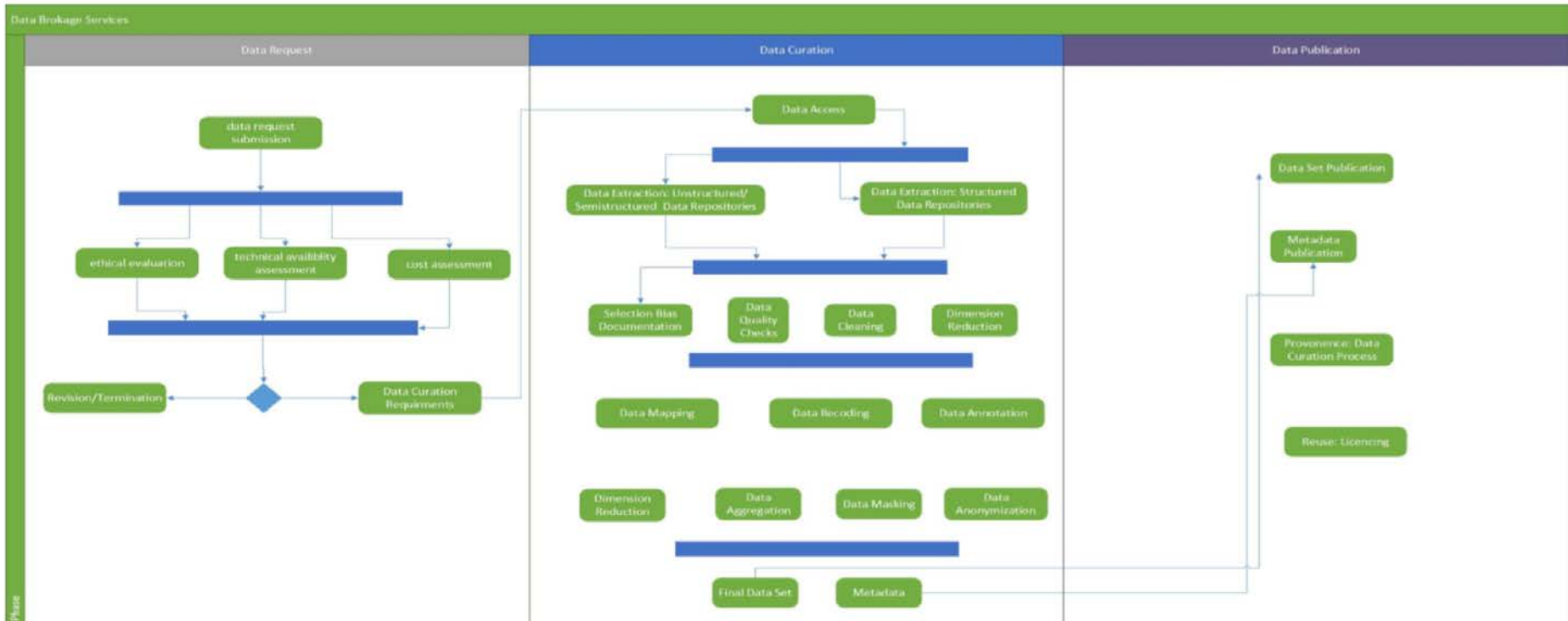
Le **Virtual Atomic and Molecular Data Centre (VAMDC)** a mis en œuvre d’une part les recommandations de la RDA en matière de citation de données pour identifier et citer les données dynamiques, et d’autre part, les **recommandations Scholix** permettant de créer des liens avec les jeux de données du VAMDC.

# Reproducible Workflows for Data Services Centers

## Data Request

## Data Curation

## Data Sharing



# Health Data WG












CEDAR

CENTER FOR EXPANDED DATA  
ANNOTATION AND RETRIEVAL

## ▼ Data Extraction - Structured

- Query Script\_Structured  
Data\_File Name
- Link to Shared Query  
Script\_Structured Data
- Query Script\_Structured  
Data\_File Format
- Query Script\_Structured  
Data\_Software Language
- Query Script\_Structured  
Data\_File Version #
- Query Script\_Structured  
Data\_Author(s)
- Query\_Structured  
Data\_Execution Date

## ▶ Data Extraction\_Unstructured Data

	Title	Created
	Cost Assessment	1/9/18 2:15 AM
	Data Access	1/10/18 7:00 AM
	Data Extraction - Structured	1/10/18 6:13 PM
	Data Extraction_Unstructured Data	1/10/18 6:16 PM
	Ethical Assessment Review	1/9/18 1:50 AM
	Participant Exclusion Criteria	2/9/18 8:55 PM
	Participant Inclusion Criteria	1/11/18 2:32 AM
	Project Metadata	1/11/18 6:59 AM
	RDA_Health Data Service Center_Data Brokerage Workflow	1/11/18 7:21 AM



---

# Describe Data

---

- › Metadata Standards Directory
- › Wheat Data Interoperability Guidelines
- › 39 Hints to Facilitate the Use of Semantics for Health Data and Nutrition
- › Data Type Registries 1 & 2
- › WG RDA/TDWG Metadata Standards for attribution of physical and digital collections stewardship
- › Brokering Governance



## Metadata

RDA | Metadata Directory

[Edit this page](#)

[View the standards](#)

[View the extensions](#)

[View the tools](#)

[View the use cases](#)

[Browse by subject areas](#)

[Contribute](#)

[Add standards](#)

[Add extensions](#)

[Add tools](#)

[Add use cases](#)

## Arts and Humanities

[FISH Interoperability Toolkit](#) [Edit](#)

A suite of tools using the [MIDAS](#) Heritage metadata standard to facilitate used to record the historic environment.

## Engineering

[CIF2Cell](#) [Edit](#)

A tool to generate the geometrical setup for various electronic structur

[ICATLite](#) [Edit](#)

A sister project of ICAT, consisting of a suite of [CSMD](#)-based software

[IUCr checkCIF](#) [Edit](#)

A tool used to check the integrity and consistency of crystal structure e

[Software for CIF](#) [Edit](#)

The International Union of Crystallography's list of programs and librari

## Life Sciences

[Bio-Formats](#) [Edit](#)

Bio-Formats reads proprietary microscopy image data and metadata,

[Darwin Core Archive Assistant](#) [Edit](#)

A web application that offers data publishers wishing to serve to the C  
composing an appropriate XML [Darwin Core](#) descriptor file to accomp



## Metadata Standards Directory: Impact

- *Learn about the various metadata standards applicable to your research.*
- *Learn about controlled vocabularies used by research communities and understand the elements that comprise these standards and vocabularies.*
- *Map between elements when combining data from different sources.*

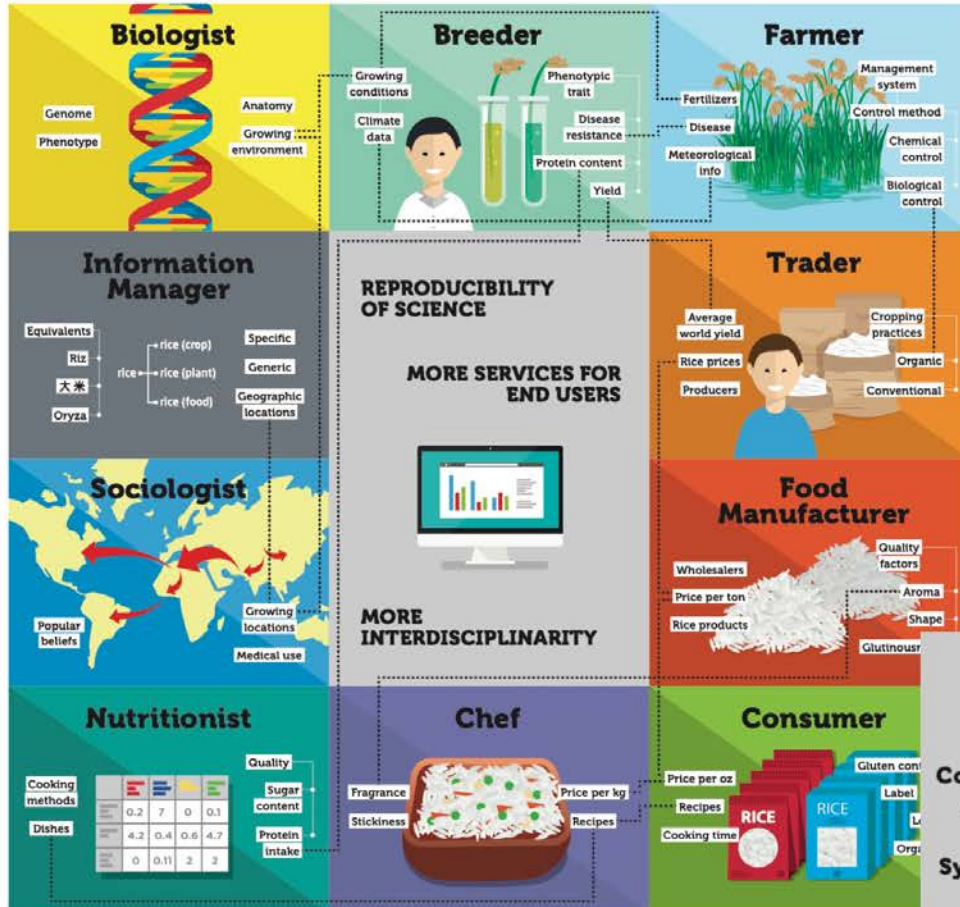
# Wheat Data Interoperability Guidelines



[Home](#) [Guidelines](#) ▾ [Ontologies & Vocabularies](#) [Use cases](#) ▾ [Getting involved](#) [About](#) ▾

- Define international metadata and ontology standards for interoperable wheat and rice data.
- Train researchers, publishers, and repositories in its use.
- Develop a interface portal for exploring the applicability of metadata and ontologies for physical and digital objects.
- Increase interoperability, discovery, and reuse of wheat and rice data.

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



## ISSUES

- Diversity of focus
- Conflicting view points
- Scale / granularity
- Language
- Synonymy & ambiguity
- Silos

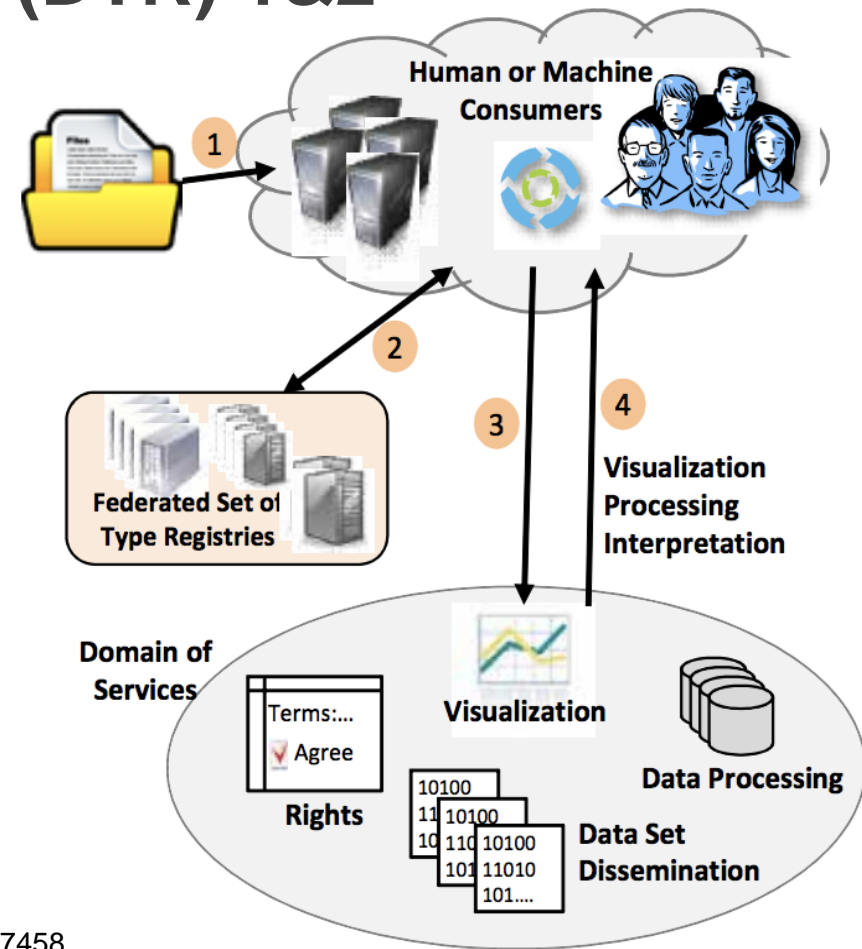
## SOLUTIONS

- Ontologies & skos resources
- Network of ontologies
- Documentation
- Standards (RDF... W3C)
- Persistent identifiers
- Shared infrastructures

# Data Type Registries (DTR) 1&2

*Problem:* Without knowing a type of data one is receiving, it is difficult to know how to analyze, store, or in other ways fit the received data within a data-centric workflow

*Solution:* DTR built a federated registry of defined data types to support automatically identifying a data's type and the associated sources and platforms which generate the defined types of data.

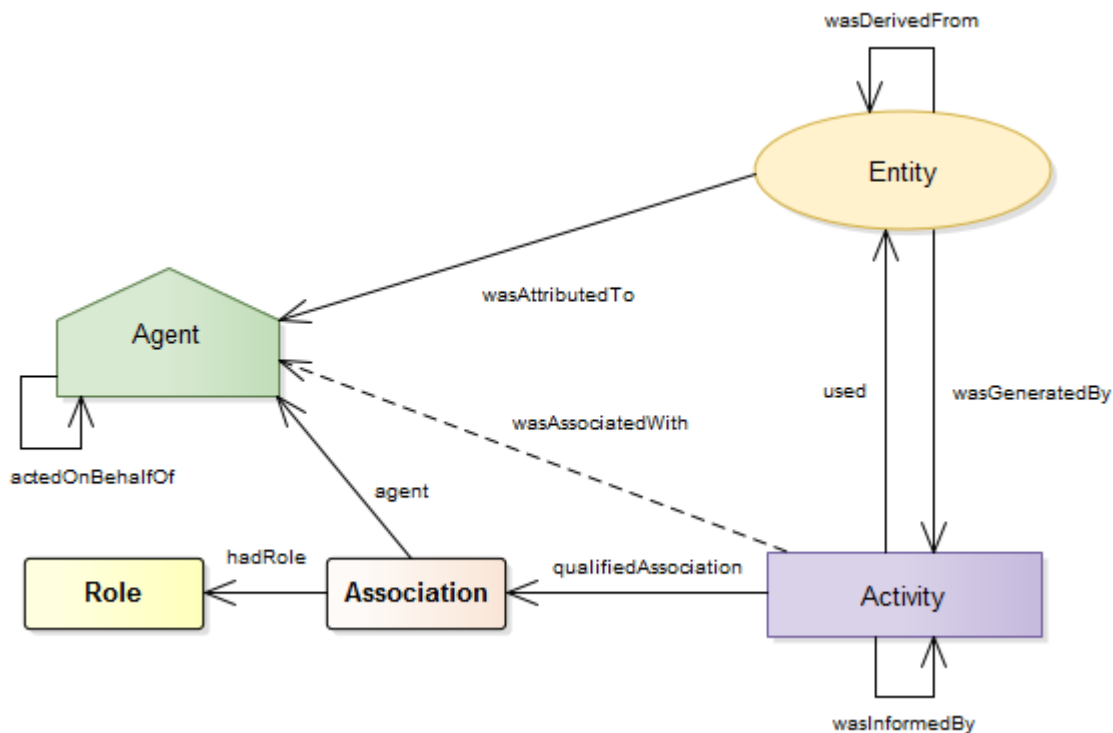




# WG RDA/TDWG Metadata Standards for attribution of physical and digital collections stewardship

*Problem:* Creating standards for giving attribution for the maintenance, curation, and digitization of physical and digital objects with a special emphasis on biodiversity collections.

*Solution:* Developed attribution metadata schema based upon use cases from multiple domains.





## Brokering Governance

Describe, develop, test, and implement a Brokering Framework that allows publication, discovery, and invocation of brokering and mediation components in a standardised manner.

<b>RDA Working Group Outcome</b>	<b>Brokering Framework WG Alignment</b>
New data standards or harmonization of existing standards.	New standard will likely be needed to describe a component within the typical research data infrastructure in a standardised manner. This ensures that its capabilities and role within the architecture is clear and unambiguous.
Greater data sharing, exchange, interoperability, usability and re-usability.	Interoperability automation is strongly dependent on the availability of brokering and mediation components, and on its automated discovery and invocation.
Greater discoverability of research data sets.	Brokering exposes data catalogues that are not completely standardised and thereby increases the scope of discovery.
Better management, stewardship, and preservation of research data.	No direct contribution.

---

# Identify, Store, & Preserve

---

- Data Foundation & Terminology: Basic Vocabulary of Foundational Terminology Query Tool
- PID Information Type Registry
- PID Kernel Information



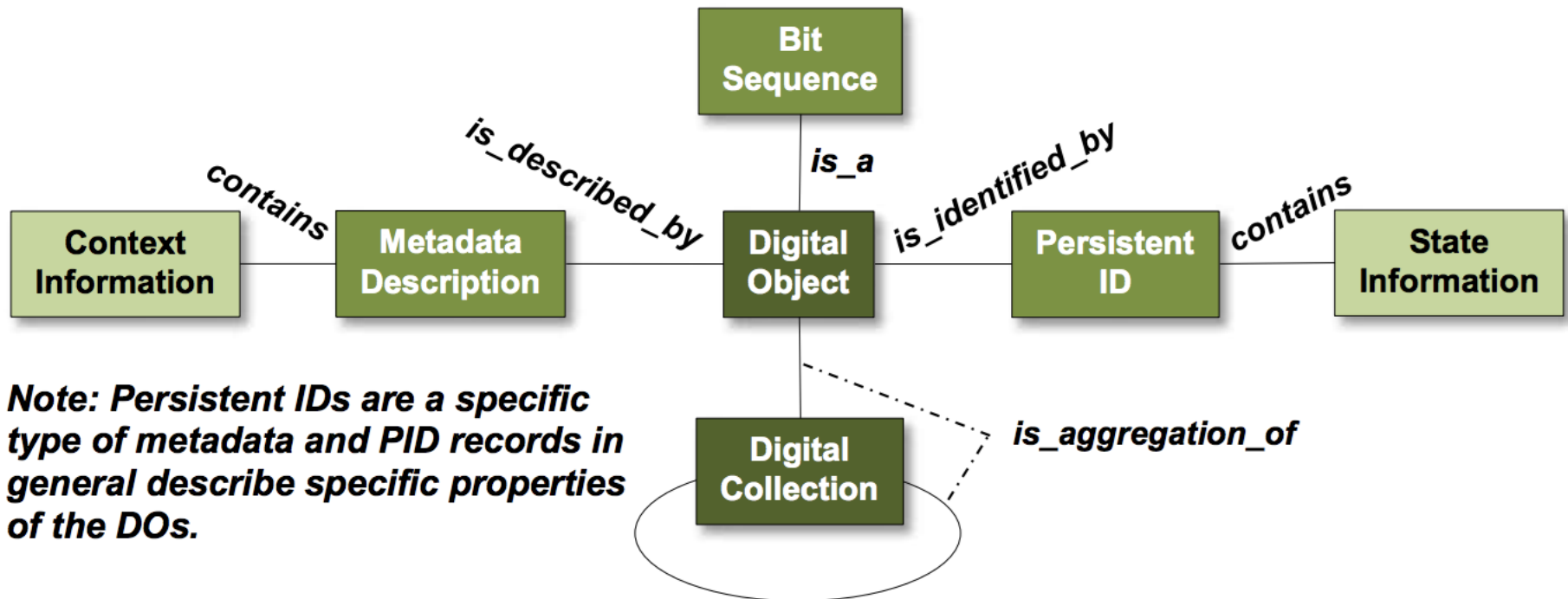
# Data Foundation & Terminology: *Basic Vocabulary of Foundational Terminology Query Tool*

*Problem:* There is no **common core data model** when organizing data to ensure accessibility and reusability.

*Goal:* Conducted interviews with 120 data experts and conducted a gap analysis comparing 21 data models

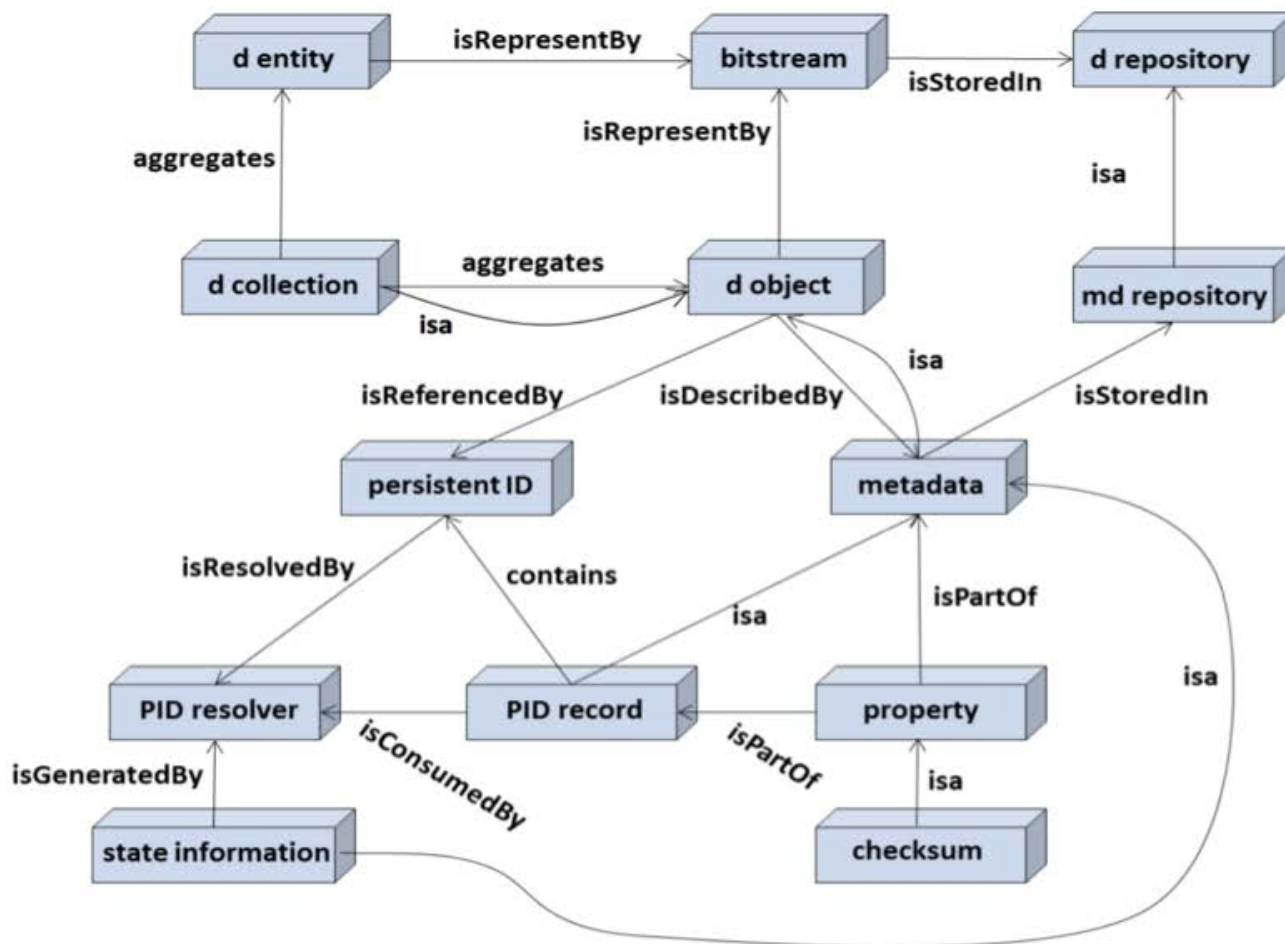
*Solution:* 5 global reports and defined a common core model for organizing data defining core terms to support harmonization of terminology, including PID, Digital Object, Metadata Repository, and Bitstream.

# Data Foundation Terminology Model



**Note: Persistent IDs are a specific type of metadata and PID records in general describe specific properties of the DOs.**

*This diagram describes the essentials of the basic data model that the DFT group worked out in a simplified way. Agreeing on some basic principles and terms would make a lot of difference in data practices.*





# PID Information Type Registry

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

*Goal:* Identify the essential types of information associated with persistent identifiers and register all types of persistent identifiers

*Solution:* Developed a conceptual model of PID types and implemented the model within an API and searchable interface.

*Impact:* Efficiently verify, find, access, and reuse PID related information

# PID Kernel Information

	Property identifier	Content format	Cardinality	Explanation
1	PID	Handle	1..n	Global identifier for the object; external to the PID Kernel Information
2	<u>KernelInformationProfile</u>	Handle	1	Handle to the Kernel Information type profile; serves as pointer to profile in DTR. Address of DTR federation expected to be global (common) knowledge.
3	<u>digitalObjectType</u>	Handle	1	Handle points to type definition in DTR for this type of object. Distinguishing metadata from data objects is a client decision within a particular usage context, which may to some extent rely on the <u>digitalObjectType</u> value provided.
4	<u>digitalObjectLocation</u>	URL	1..n	Pointer to the content object location (pointer to the DO). This may be in addition to a pointer to a human-readable landing page for the object.
5	<u>digitalObjectPolicy</u>	Handle	1	Pointer to a policy object which specifies a model for managing changes to the object or its Kernel Information record, including particularly object access and modification policies. A caller should be able to determine the expected future changes to the object from the policy, which are based on managed processes the object owner maintains.
6	etag	Hex string	1	Checksum of object contents. Checksum format determined via attribute type referenced in a Kernel Information record.
7	dateModified	ISO 8601 Date	0..1	Last date/time of object modification. Mandatory if applicable.

*Problem:* Existing middleware is currently insufficient for linking multiple PID types, consistently growing in volume and complexity.

*Goal:* Inject a tiny amount of carefully selected metadata into a Persistent ID (PID) record.

*Impact:* 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.

<https://docs.google.com/document/d/1EdS5OCoeWd4VY0HNLHkhzdQojsRgc3P8aWXYQKTqs8M/edit>



---

# Disseminate, Link, & Find

---

- FAIRSharing Registry
- Journal and publisher research data policy master framework
- Workflows for Research Data Publishing: Models and Key Components
- Data Description Registry Interoperability: A research Graph dataset for connecting research data repositories using RD-Switchboard
- RDA/WDS Publishing Data Services/SCHOLIX

# FAIRsharing: standards, databases, repositories and policies

## *Joint RDA-Force11 FAIRsharing WG*

- The FAIRsharing resource of interlinked records on standards (for identifying, reporting, citing data and metadata), databases (repositories and knowledge-bases) and data policies (from journals, publishers, funders and other organizations), ranging from the generic and multi-disciplinary, to those from specific domains.
- The FAIRsharing recommendations, to guide the users and producers of standards, databases and repositories on how to best select and describe these resources; and to guide funders and publishers on how to recommend them in data policies.



# Journal and publisher research data policy master framework

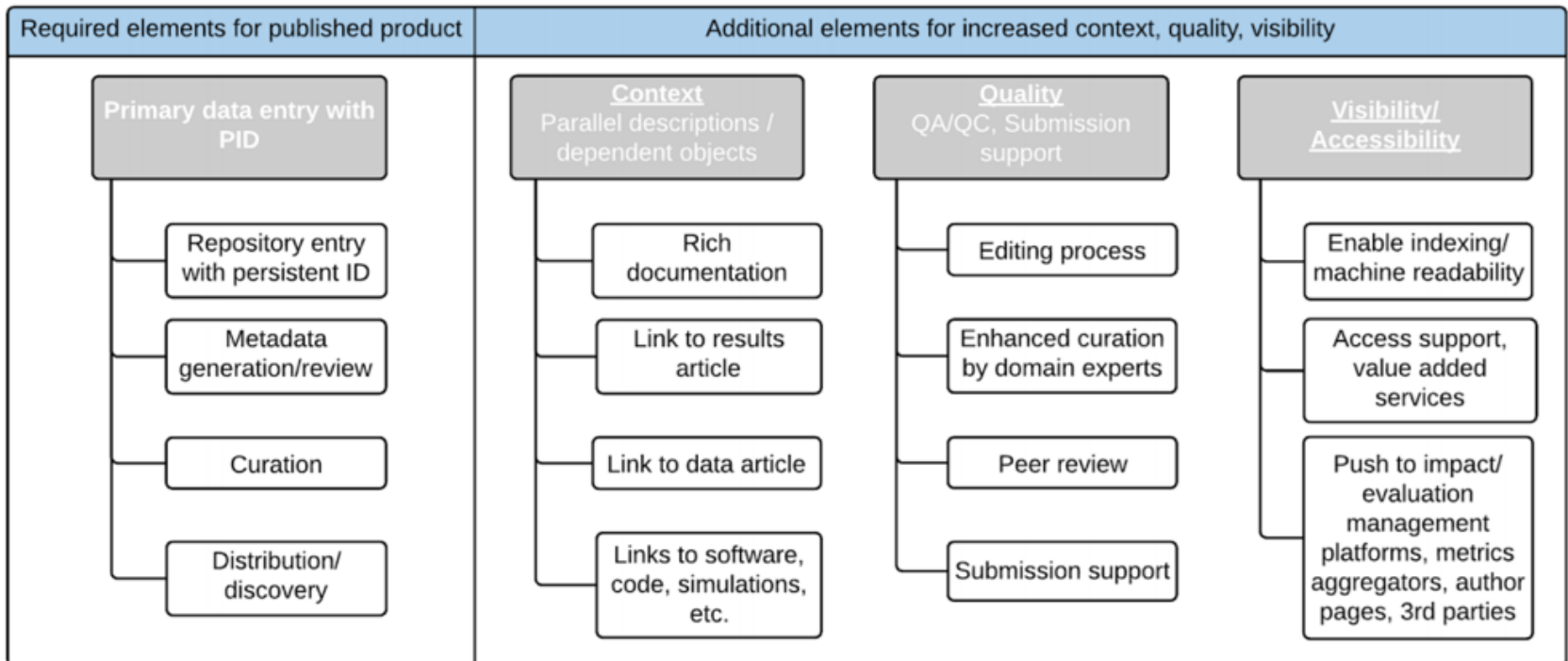
Data citation	●	●	●	●	●	●
Data licensing	●	●	●	●	●	●
Researcher/ author support	◐	●	●	●	●	●
Data availability statements	○	◐	●	●	●	●
Mandatory data sharing (specific papers)	○	◐	◐	●	●	●
Data formats and standards	○	◐	◐	◐	◐	●
Mandatory data sharing (all papers)	○	○	◐	◐	●	●
Mandatory data citation	○	○	◐	◐	◐	●
Peer review of data	○	○	◐	◐	◐	●
Data Management Plans (DMPs)	○	○	◐	◐	◐	●

Draft of framework: [https://docs.google.com/document/d/1DTAfOKkE1a2n2f\\_1hGcrXlrw-5Tq\\_AL5tk-ju8B82\\_E/edit](https://docs.google.com/document/d/1DTAfOKkE1a2n2f_1hGcrXlrw-5Tq_AL5tk-ju8B82_E/edit)



# RDA/WDS Publishing Data Services

## Workflows for Research Data Publishing: Models and Key Components

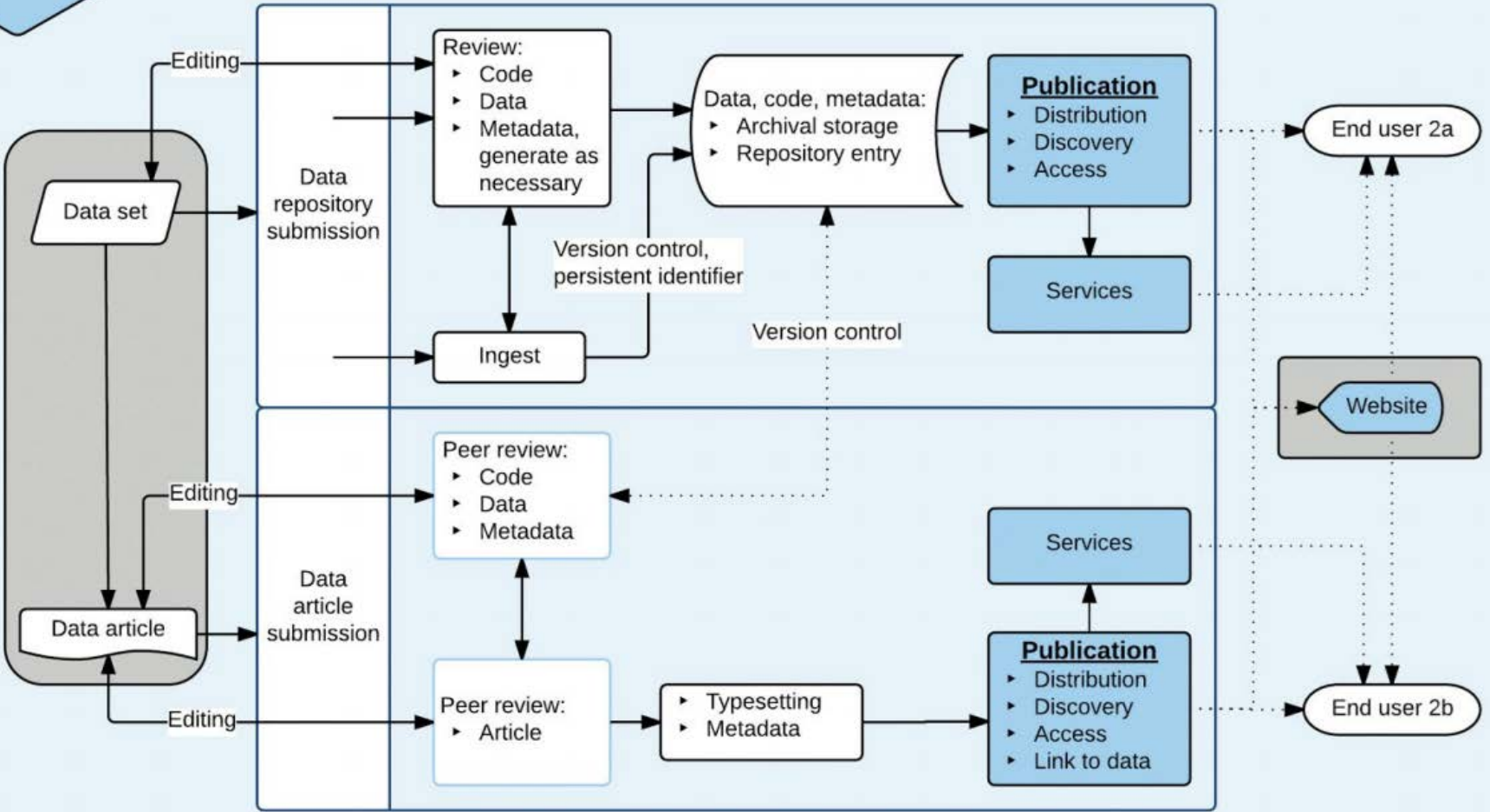


Full Recommendation Package:<http://dx.doi.org/10.15497/RDA00004>

# Data Publication

**Key**

- Process flow
- .....→ Electronic information transfer
- Researcher controlled
- Peer interaction point
- Public exposure of work



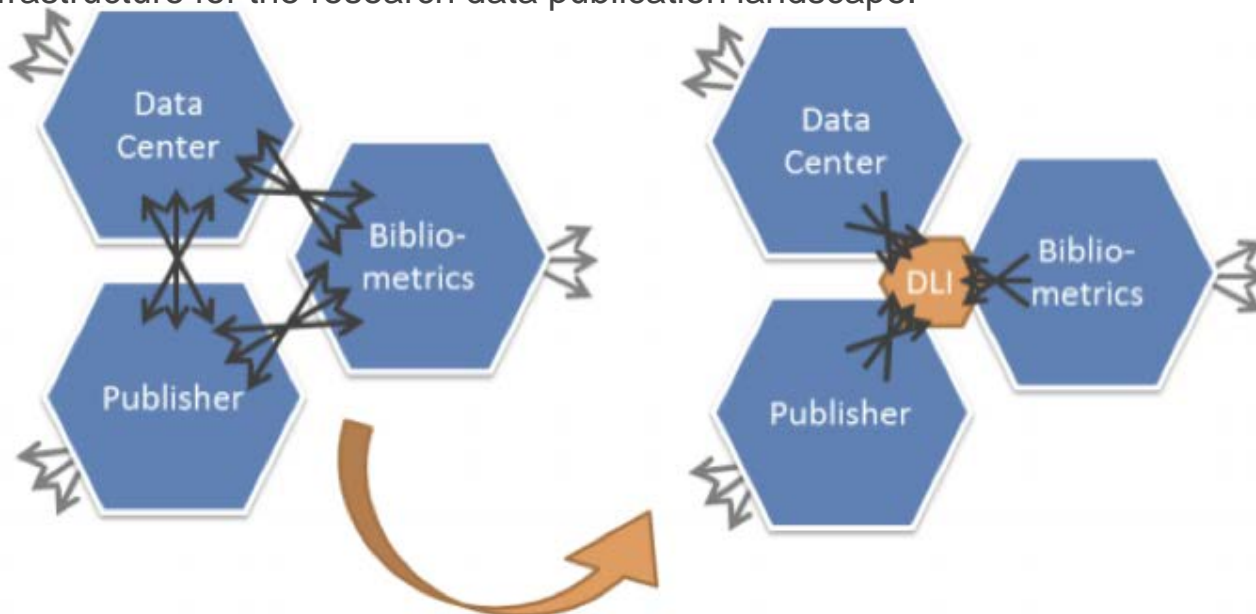


# Data Description Registry Interoperability (DDRI)

- *Problem:* More repositories with more data, yet often siloed and non-interoperable
- *Goal:* Interoperability model for cross-platform discovery by connecting datasets together on the basis of co-authorship or other collaboration models such as joint funding and grants
- *Solution:* Compiles enabling infrastructures based on existing open protocols and standards with a flexible and extensible approach that allows registries to opt-in and enables any third-party to create particular global views of research data.

# RDA/WDS Publishing Data Services

- Problem: There is no standard service for linking publications, research data, and associated metadata and materials.
- Goal: Move from a plethora of (mostly) bilateral arrangements to a one-for-all service model infrastructure for the research data publication landscape.



<http://dx.doi.org/10.15497/RDA00002>





# SCHOLIX

Content Provider	Links	Object			
		Publication	Dataset	Unknown	total
<a href="#">OpenAIRE</a>	29262	14139	4124	472	18735
<a href="#">RCSB</a>	175648	44859	88687	79	133625
<a href="#">Pangaea</a>	859076	14671	289395	37682	341748
<a href="#">Datasets in Datacite</a>	44121092	381351	2646694	1017376	4045421
<a href="#">Cambridge Crystallographic Data Centre</a>	1296158	278005	639698	2306	920009
<a href="#">NCBI Nucleotide</a>	0	0	59200	0	59200

*Open source metadata schema adapted widely by large national and international repositories including Europe PMC, Dryad, and Scopus, as well as providing linkage through CrossRef, OpenAIRE, and DataCite*

RDA/WDS Publishing Data Services WG Output: <http://hdl.handle.net/11304/327e42da-5eb9-4f3d-9e78-5d1301843320>

The Scholix Initiative: [www.scholix.org](http://www.scholix.org)

DLI aggregation and query service fed by Scholix: <https://dliservice-prototype-dli.d4science.org/#/>

---

# Policy, Legal Compliance, and Capacity

---

- Practical Policies
- Repository Audit and Certification/ DSA-WDS partnership >> CoreTrustSeal
- CODATA -RDA Legal Interoperability of Research Data



# Practical Policy

## Machine Actionable Policy Templates for Trusted Research Repositories

- Enforce management
- Automate administrative tasks
- Validate assessment criteria
- Automate scientific analyses
- Ensure policy transparency,  
discoverability, and shareability

Full Recommendation:

<http://dx.doi.org/10.15497/83E1B3F9-7E17-484A-A466-B3E5775121CC>



# Repository Audit and Certification / DSA–WDS Partnership



- *Problem:* Certification is fundamental in guaranteeing the trustworthiness of digital repositories, and thus in sustaining the opportunities for long-term data sharing and corresponding services.



- *Goal:* Develop a **common framework for certification and a service of trusted data repositories** which harmonizes the efforts of long-trusted standards



- *Solution:* Developed Coretrust Seal catalog of common requirements



# CODATA-RDA Legal Interoperability of Research Data

One. Facilitate the lawful access to and reuse of research data.

Two. Determine the rights to and responsibilities for the data.

Three. **Balance the legal interests.**

Four. State the rights transparently and clearly.

Five. Promote the harmonization of rights in research data.

Six. Provide proper attribution and credit for research data.

Full Recommendation: <https://doi.org/10.5281/zenodo.162241>



# RDA outputs across the Workflow

Scientific Community Engagement

Data Search/Reuse

Data Management Plan

Active Data Management Plans

Publisher Commitments

Dynamic Data Citation

Data Sharing Workflows

Develop & Explore Vocabularies

Scholix

Promotion



PID management

Repository Interoperability

Repository Trust

Retention

Reproducibility

Publisher Commitments

Data Sharing Workflows

Scholix

Repository Interoperability

Repository Trust

Reproducibility