RESEARCH DATA ALLIANCE

Adopting Outputs from the Research Data Alliance

Stefanie Kethers, RDA Director of Operations
Andrew Treloar, ANDS Director of Technology

research data sharing without barriers
rd-alliance.org

Agenda

- Introduction and brief overview of current RDA Outputs and their adopters
- An organizational perspective on RDA outputs adoption - Malcolm Wolski
- An infrastructure provider's perspective on RDA outputs adoption – Lesley Wyborn
- A researcher's perspective on the value and uptake of the Biosharing registry - Jeff Christiansen
- Discussion
  - How would you like to find out about RDA Outputs?
  - What would help you adopt RDA Outputs?
  - …

Introduction and Brief Overview of Current RDA Outputs and their Adopters

Stefanie Kethers / Andrew Treloar

research data sharing without barriers

rd-alliance.org

RDA Output Types

- RDA Recommendations
  - “Flagship outputs” - official, endorsed results of RDA
  - Every WG should develop at least one
  - May include specifications, taxonomies or ontologies, workflows, schemas, data models, etc.
  - Comparable to other organisations’ “specifications” or “standards”
  - Undergo formal review
- Supporting Outputs
  - Useful solutions from RDA WGs and IGs
  - May not be as clearly adoptable by organisations as RDA Recommendations.
  - Undergo a community review
- Other Outputs
  - No endorsement process

Sample Recommendation 1: Repository Audit and Certification Catalogues

- Repository Audit and Certification Catalogues
- dois:
  - Common Requirements: 10.17026/dans-22n-gk35
  - Common Procedures: 10.15497/rda00019
- Set of harmonized Common Requirements and Common Procedures for certification of repositories at the core level, drawing on criteria and procedures already put in place by the Data Seal of Approval (DSA) and the ICSU World Data System (ICSU-WDS).

Sample Recommendation 2: Data Description Registry Interoperability Model

- Data Description Registry Interoperability Model
- doi: 10.15497/RDA00003

- Provides a mechanism to connect datasets in various data repositories based on various models such as co-authorship, joint funding, grants, etc.
- Implemented as RD-Switchboard.

Sample Recommendation 3: BioSharing Recommendations

- BioSharing Recommendations: data repositories, standards and policies in the life, biomedical and environmental sciences
  doi: http://dx.doi.org/10.15497/RDA00017
- Set of recommendations to guide users and producers of databases and content standards to select and describe them, or recommend them in data policies.
- Curated registry enacting the recommendations and assisting a variety of end users through well described, interlinked, and cross-searchable records on content standards, databases and data policies.
- Now broadening scope.

Sample Supporting Outputs

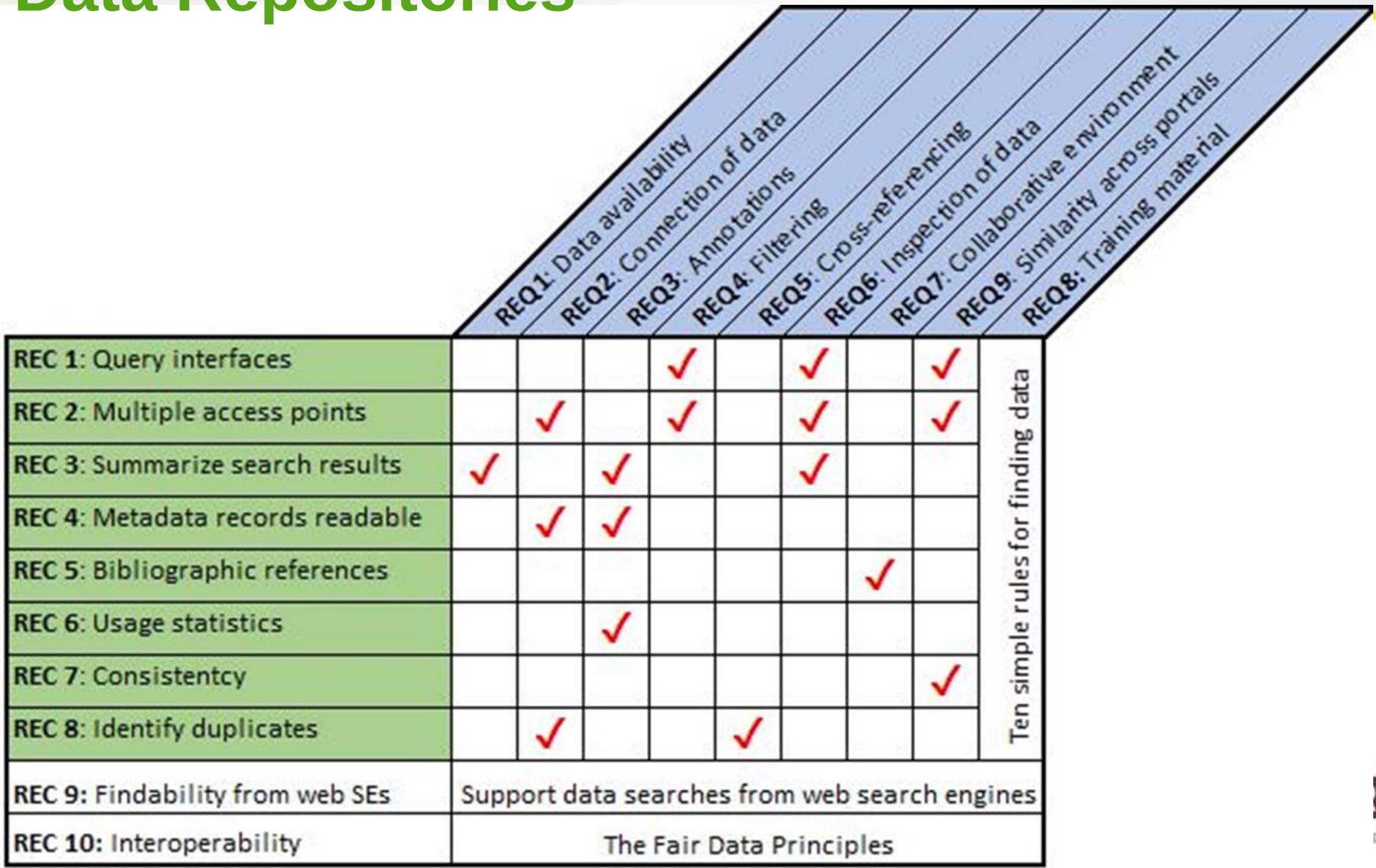
- 23 Things: Libraries For Research Data. doi: dx.doi.org/10.15497/RDA00005 (English, available in 11 languages in total)
- Data Discovery Paradigms IG emerging output:
  1. Ten simple rules for finding research data.
  2. User requirements and recommendations for data repositories.

  Contact: mingfang.wu@ands.org.au

1. Ten simple rules for finding research data

1. Think about the data you need and why you need them.
2. Select the most appropriate resource.
3. Construct your query.
4. Make the repository work for you.
5. Refine your search.
6. Assess data relevance and fitness-for-use.
7. Save your search and data source details.
8. Look for data services, not just data.
9. Monitor the latest published data.
10. Give back.

2. Requirements and Recommendations for Data Repositories



	REQ1: Data availability	REQ2: Connection of data	REQ3: Annotations	REQ4: Filtering	REQ5: Cross-referencing	REQ6: Inspection of data	REQ7: Collaborative environment	REQ9: Similarity across portals	REQ8: Training material
REC 1: Query interfaces				✓		✓		✓	Ten simple rules for finding data
REC 2: Multiple access points		✓		✓		✓		✓	
REC 3: Summarize search results	✓		✓			✓			
REC 4: Metadata records readable		✓	✓						
REC 5: Bibliographic references							✓		
REC 6: Usage statistics			✓						
REC 7: Consistentcy								✓	
REC 8: Identify duplicates		✓			✓				
REC 9: Findability from web SEs	Support data searches from web search engines								
REC 10: Interoperability	The Fair Data Principles								

DATA ALLIANCE

List of all Current Outputs

https://www.rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs

… more to come after March 2018 (RDA Plenary 11)

RESEARCH DATA ALLIANCE

An organizational perspective on RDA outputs adoption

Malcolm Wolski

research data sharing without barriers

rd-alliance.org

Malcolm Wolski
Director, eResearch Services
Griffith University

RDA Organisational Advisory Board Member

CC BY-SA 4.0

Adoption of RDA Outputs

https://www.rd-alliance.org/recommendations-outputs/adoption-recommendations

Example

Data Type Registries	**Data Type Model and Registry** Ensures data producers classify their data sets in standard data types, allowing data users to automatically identify instruments to process and visualise the data	**Recommendation page:** https://rd-alliance.org/group/data-type-registries-wg/outcomes/data-type-registries **DOI:** http://dx.doi.org/10.15497/A5BCD108-ECC4-41BE-91A7-20112FF77458	• Materials Genome Initiative (MGI) • DataFed & the Air Quality Community Catalog • Deep Carbon Observatory Data Portal • Opening up Northern Forest Research Data – Improving Citation and Documentation Systems to Increase Participation in Publishing Data • EUDAT: the data type registry for describing & sharing scientific datasets • Bringing Visibility to Food Security Data Results: Harvests of PRAGMA and RDA

CC BY-SA 4.0

Griffith's Interest in RDA Outputs

Repository Audit and Certification / DSA–WDS Partnership Working Group	Repository Audit and Certification Catalogues dois: 10.17026/dans-22n-gk35 and 10.15497/rda00019	Creates harmonized Common Procedures for certification of repositories at the basic level, drawing from the procedures already put in place by the Data Seal of Approval (DSA) and the ICSU World Data System (ICSU-WDS)
WG OUTPUT		
Repository Platforms for Research Data Interest Group	Matrix of use cases and functional requirements for research data repository platforms	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.
IG SUPPORTING OUTPUT		

CC BY-SA 4.0

Reviewers Needed for RDA Outputs and Case Statements To Provide Organisational Perspective

Are they adoptable as written?

Evidence of sufficient testing?

Examples of early adopters or pilots?

How hard is it to implement?

Why do I care:

- Is there a cost benefit?
- Is there a productivity benefit?
- Are there other reasons eg compliance?
- Is there a strategic or tactical benefit?

Timing: when is the right time to adopt. Eg easily applied to an existing service/system or during next major upgrade?

CC BY-SA 4.0

An infrastructure provider's perspective on RDA outputs adoption

Lesley Wyborn / Jingbo Wang

research data sharing without barriers

rd-alliance.org

NCI
NATIONAL COMPUTATIONAL INFRASTRUCTURE

Adopting Outputs from the Research Data Alliance at NCI

Lesley Wyborn, Jingbo Wang, Ben Evans

NCI is actively engaged in several RDA Working & Interest Groups

Interest Groups:

- Active Data Management Plans IG
- Big Data IG*
- Data Rescue IG*
- Mapping the Landscape IG*
- Physical Samples and Collections in the Research Data Ecosystem IG*
- Research Data Provenance IG
- RDA/WDS Certification of Digital Repositories IG
- Software Source Code IG
- Virtual Research Environments IG*
- Vocabulary Services Interest Group

Working Groups

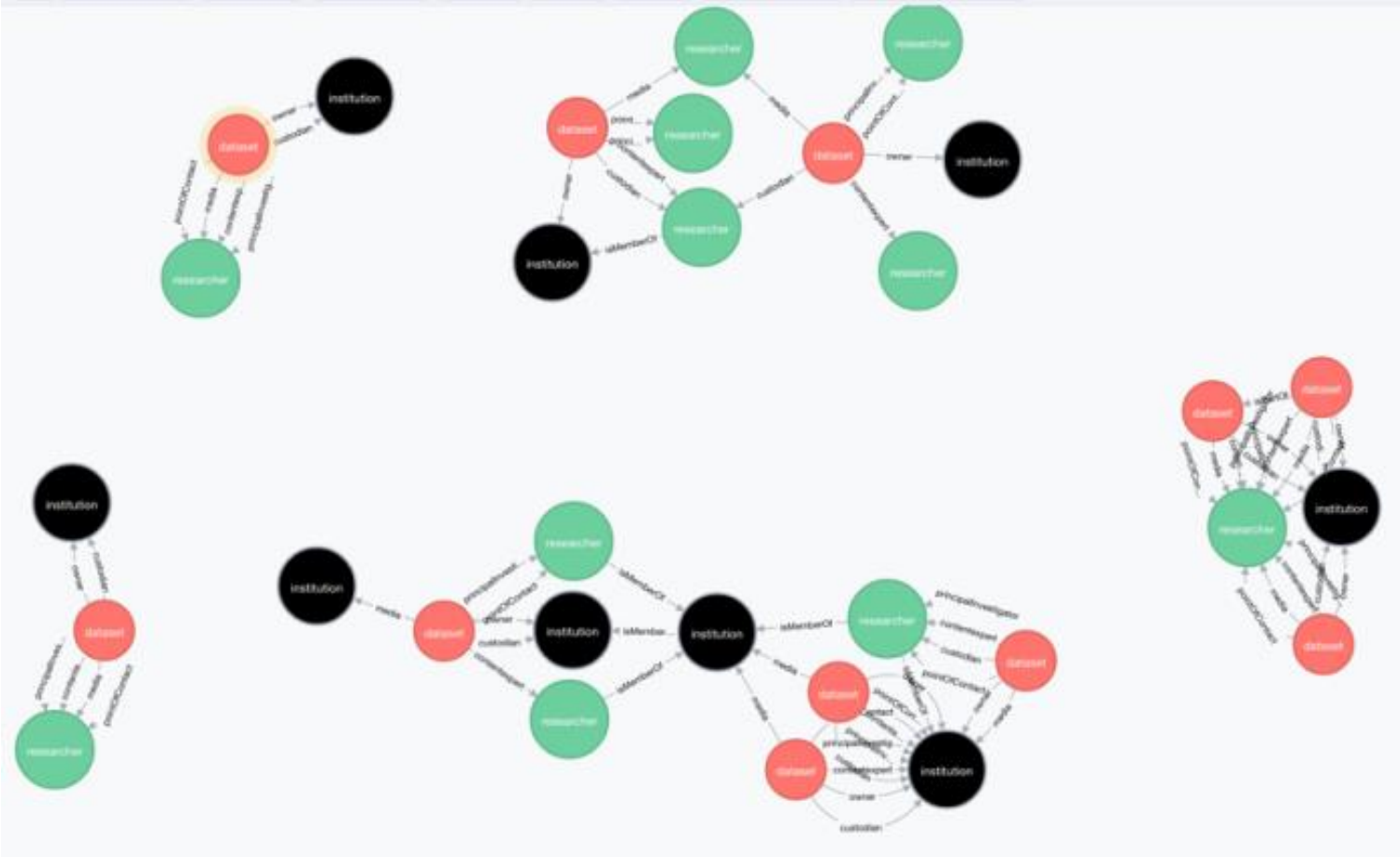
- Array Database Assessment WG
- Data Description Registry Interoperability (DDRI) WG
- Data Citation WG
- Data Versioning WG*
- Provenance Patterns WG

(*Indicates NCI person is Co Chair)

Research Graph is a Key Output NCI is interested in

User question:	RD-switchboard query:
What is the usage of NCI's datasets?	→ How many datasets published at NCI are being referenced in the research journal articles?
What is the awareness of the available datasets within the research community?	→ How many researchers/institutes are connected to the datasets?
If I would like to know more about this dataset, who should I contact? What is the previous research done using this dataset?	→ Finding connections between NCI datasets and international repositories such as CERN and ORCID -- data discovery.

Research Graph is a Key Output NCI is interested in



We added a new field into NCI's DMP:
Expected reuse

rational:

- compliant with FAIR* principle
- data reuse cases would be helpful
- data validation
- increase community trust
- increase data citation
- increase data value

(Source:*https://www.force11.org/group/fairgroup/fairprinciples)

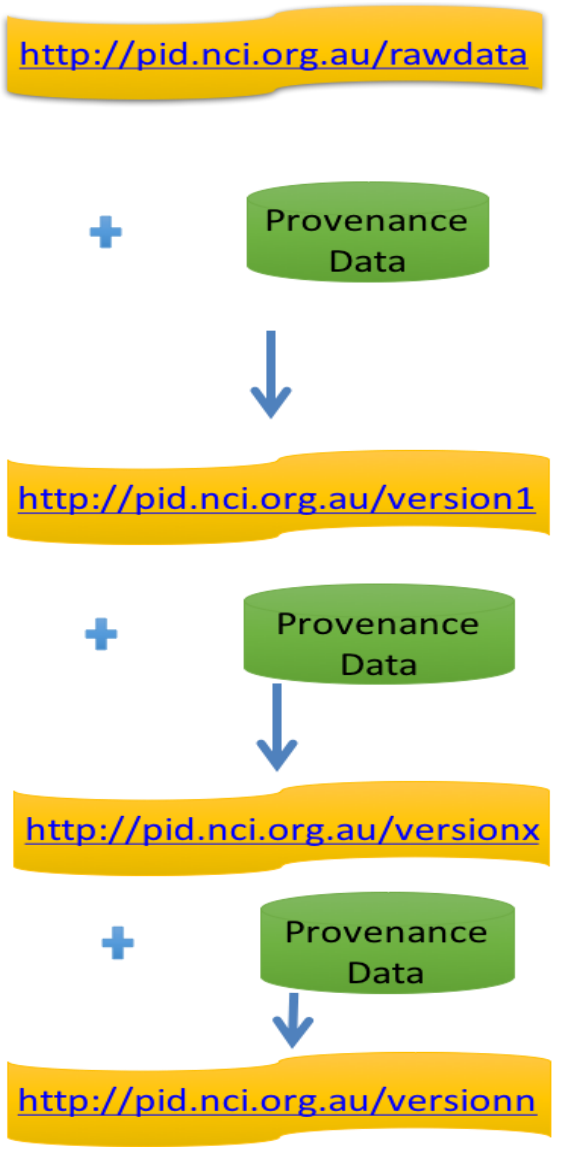
23 DMP Theme at DCC	NCI's DMP	comments
ID	File Identifier	Equivalent
PROJECT DESCRIPTION	Abstract/Activities	Equivalent
RELATED POLICIES	ISO19115?	No perfect match field
EXISTING DATA	Lineage	Close field
RELATIONSHIP TO EXISTING DATA	Lineage	Close field
DATA DESCRIPTION	Abstract	Equivalent
DATA FORMAT	DATA FORMAT DATA VERSION	Equivalent
DATA VOLUMES	Transfer size	Equivalent
DATA TYPE		n/a
DATA CAPTURE METHODS	Tools, software, algorithm, library	Multiple fields
DATA QUALITY	QA/QC process	Equivalent
DOCUMENTATION	Readme and additional Documentation, Paper Reference	Multiple fields
METADATA	ISO19115	Close field
DISCOVERY BY USERS	Dataset URI, DATA Services	Multiple fields
ETHICAL ISSUES	Limitation on using data	Close field
IPR OWNERSHIP AND LICENCING	Data Ownership/Custodianship, Access/use constraints	Multiple fields
STORAGE AND BACKUP	Local file path, backup and recovery plan	Multiple fields
DATA SECURITY	Limitations on using dataset	Close field, n/a for most of our open data
DATA SELECTION	Maintenance process	Close field
PRESERVATION PLAN	See our Data backup form	n/a in DMP, separated doc
PERIOD OF PRESERVATION	Retention of Data	Equivalent
DATA REPOSITORY	Local file path, data services	Equivalent
EXPECTED REUSE	Readme and additional document, How-to Guide	Multiple fields, but would be nice to explicitly have such a field.
METHOD FOR DATA SHARING	Data Services	Equivalent
TIMEFRAME FOR DATA SHARING	Retention of data	Equivalent
RESTRICTION ON SHARING	Limitations on using dataset	Close field
MANAGED ACCESS PROCEDURES	License	Equivalent
RESPONSIBLITIES	roles	Multiple fields
RESOURCING	roles	Multiple fields

Data Citation and Data Versioning WGs

Recommendation is a good for managing the dynamic nature of a database programmatically, however, it does not apply to large scale data collections because storing multiple time stamped snapshots of these is not feasible, fundamentally due to cost of the infrastructure.

Instead, we proposed a preservie the recipe that created the data - i.e. DOI/PID+provenance model to handle large scale data citation issue and presented at AGU, 2015 (https://agu.confex.com/agu/fm15/webprogram/Paper78116.html)

This is now likely to see both WGs combine and put the suggestions through to W3C Data on the Web Best Practices Recommendations (https://www.w3.org/TR/dwbp/)



A researcher's perspective on the value and uptake of the Biosharing registry

Jeff Christiansen

research data sharing without barriers

rd-alliance.org

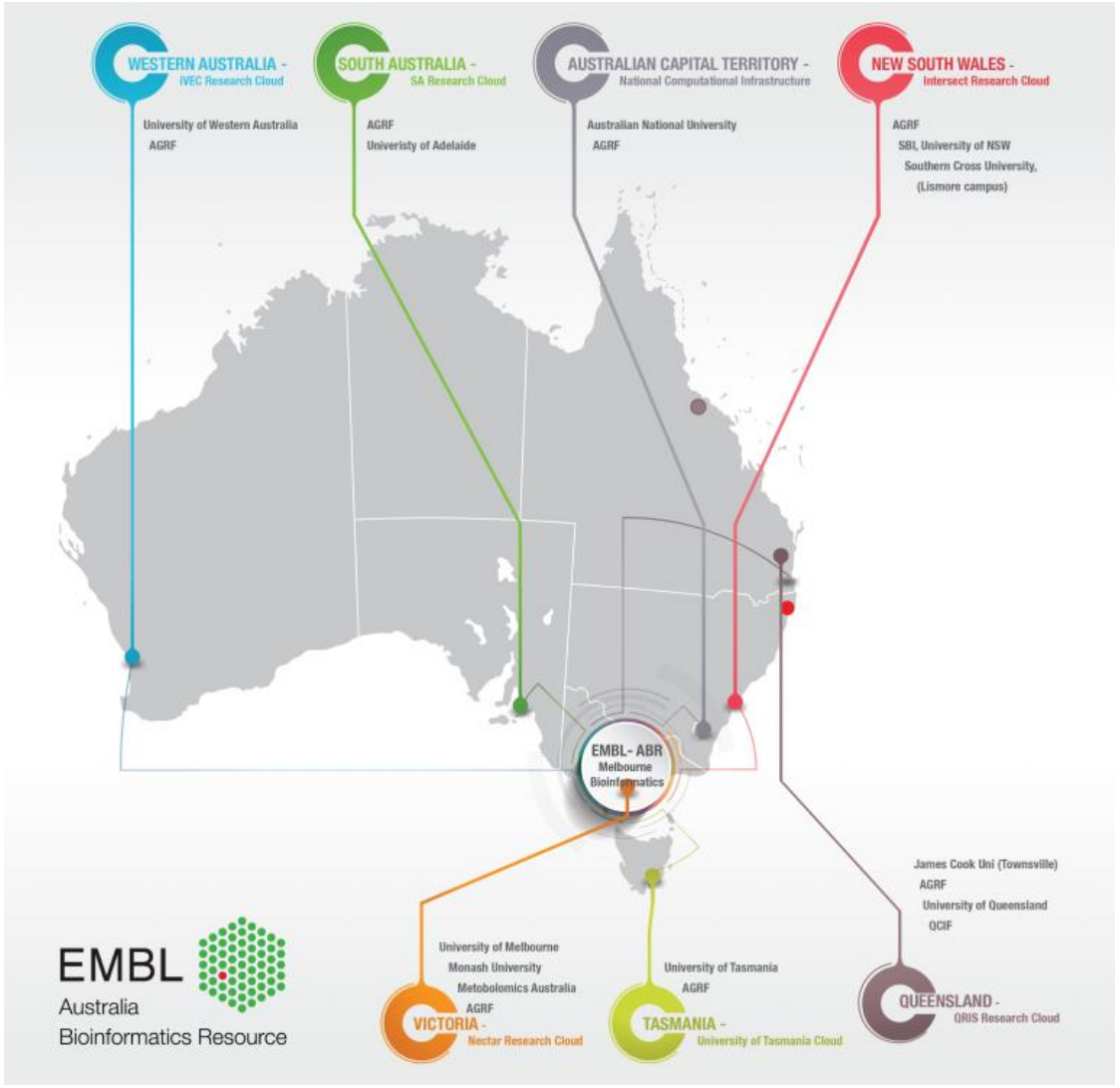
biosharing.org

Researchers' perspectives

Jeff Christiansen
EMBL-ABR Key Area Coordinator

This work is licensed under a Creative Commons Attribution 3.0 Australia License.

EMBL-Australia Bioinformatics Resource
EMBL-ABR



Data

Tools

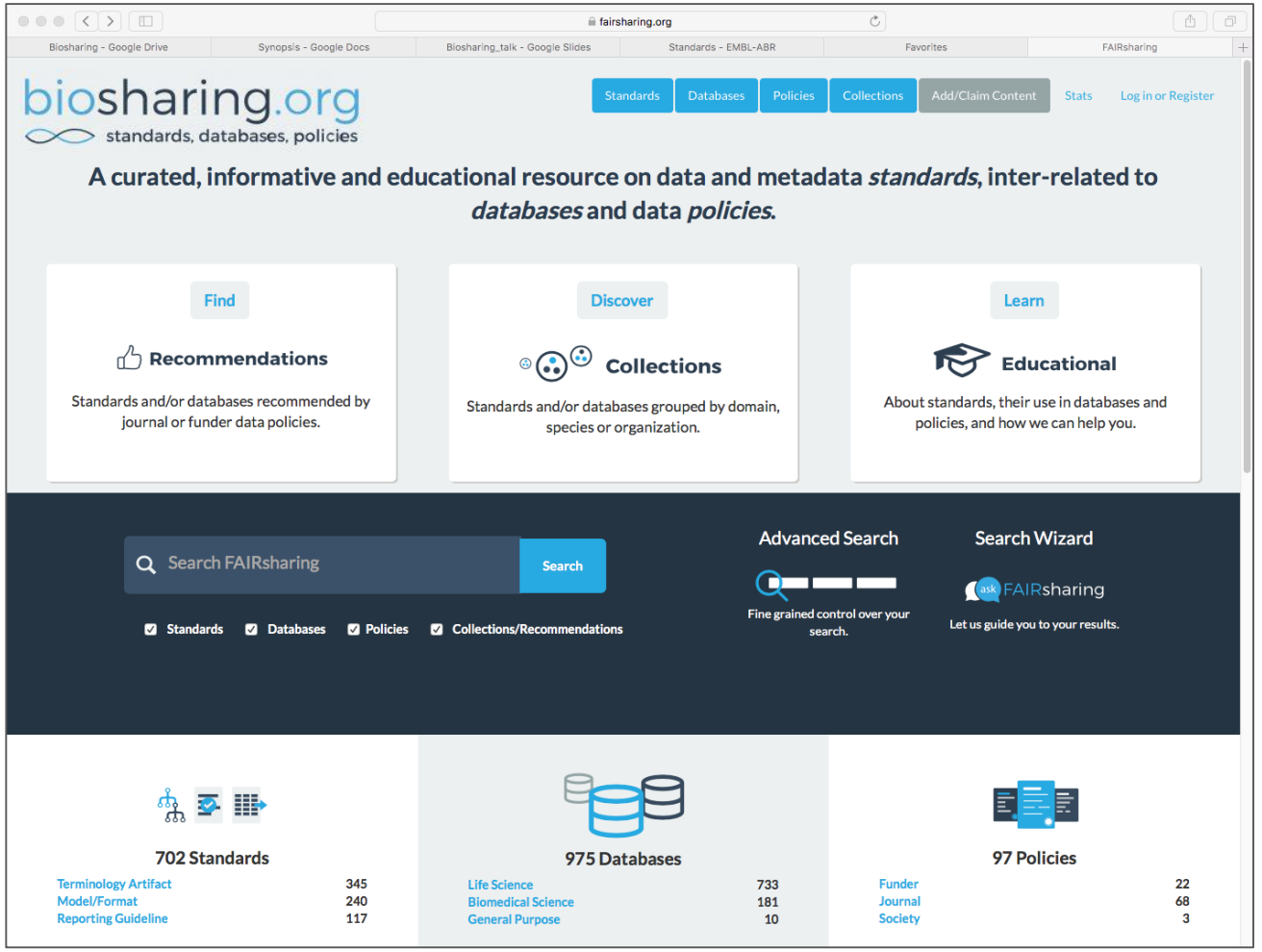
Platforms

Compute

Training

Standards

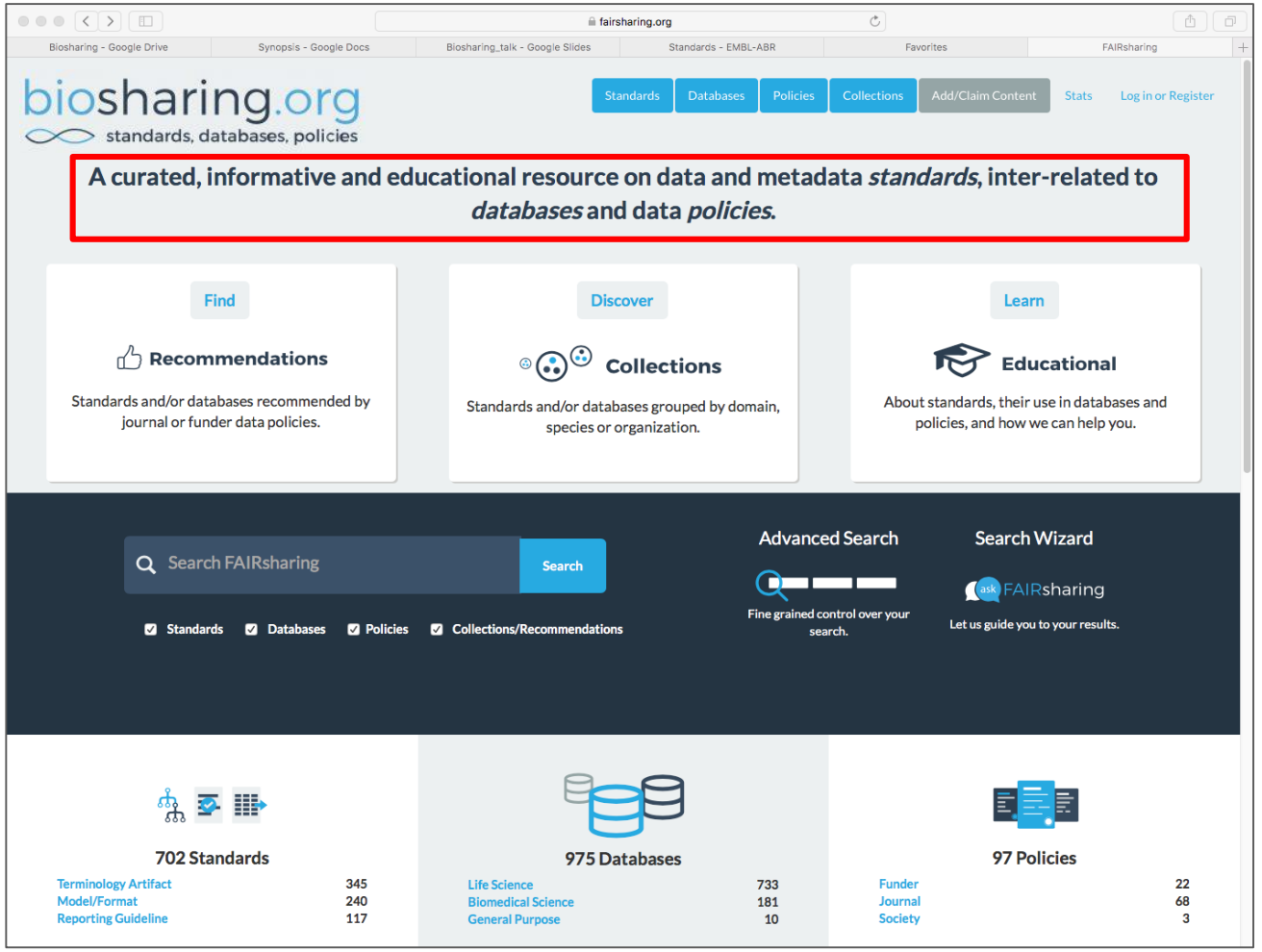
biosharing.org



CC BY

biosharing.org

What is it?



CC BY

biosharing.org

Why was it developed?

For researchers/curators looking for guidance

- To find the appropriate standard and database for your data;
- To find journal requirements or funding agency data policies

CC BY

biosharing.org

Why was it developed?

For researchers/curators looking for guidance

- To find the appropriate standard and database for your data;
- To find journal requirements or funding agency data policies

For developers seeking to make their resource more findable

- To register or update an standard and/or database description to make them more discoverable to others and get credit for it.

CC BY

biosharing.org

Why was it developed?

For researchers/curators looking for guidance

- To find the appropriate standard and database for your data;
- To find journal requirements or funding agency data policies

For developers seeking to make their resource more findable

- To register or update an standard and/or database description to make them more discoverable to others and get credit for it.

For journal editors/librarians creating guidelines

- To create a view on an interrelated set of standards and/or databases, a simple way to complement data guidelines and assist users.

CC BY

biosharing.org

Why was it developed?

For researchers/curators looking for guidance

- To find the appropriate standard and database for your data;
- To find journal requirements or funding agency data policies

For developers seeking to make their resource more findable

- To register or update an standard and/or database description to make them more discoverable to others and get credit for it.

For journal editors/librarians creating guidelines

- To create a view on an interrelated set of standards and/or databases, a simple way to complement data guidelines and assist users.

For funders developing data policies

- To help refine policy by discovering which standards and databases are inter-related, more used and mature, or are funded by the organisation

CC BY

biosharing.org

What is in it?

STANDARDS

The standards in biosharing are manually curated from a variety of sources, including:

- BioPortal (repository of biomedical ontologies),
- MIBBI (repository of Minimum Information specifications for Biological and Biomedical Investigations)
- Equator Network (database of health research reporting guidelines)

Current content - 701 standards:

- Terminologies (ontology, controlled vocabulary, staging systems, etc) 344 (7 Aus)
- Model/Format (exchange formats, markup languages, RIF-CS, etc) 240 (11 Aus)
- Reporting Guideline (diagnostic reporting guidelines, minimum info specs, etc) 117 (10 Aus)

CC BY

biosharing.org

What is in it?

DATABASES

A catalogue of databases, described according to the BioDBcore guidelines (a community-defined, uniform, generic description of the core attributes of biological databases), along with the standards used within them; partly compiled with the support of Oxford University Press (NAR Database Issue and *DATABASE* Journal).

Current content - 975 databases:

- Life Science (non-human) 733 (20 Aus)
- Biomedical Science (human) 181 (5 Aus)
- General Purpose (incl. figshare, OSF, zenodo, wikidata, RDA, etc) 10 (1 Aus)

CC BY

biosharing.org

What is in it?

POLICIES

A catalogue of data preservation, management and sharing policies from international funding agencies, regulators and journals.

Current content - 97 Policies

- Funder 22 (1 Aus)
- Journal 68 (0 Aus)
- Society 3 (1 Aus)

CC BY

biosharing.org

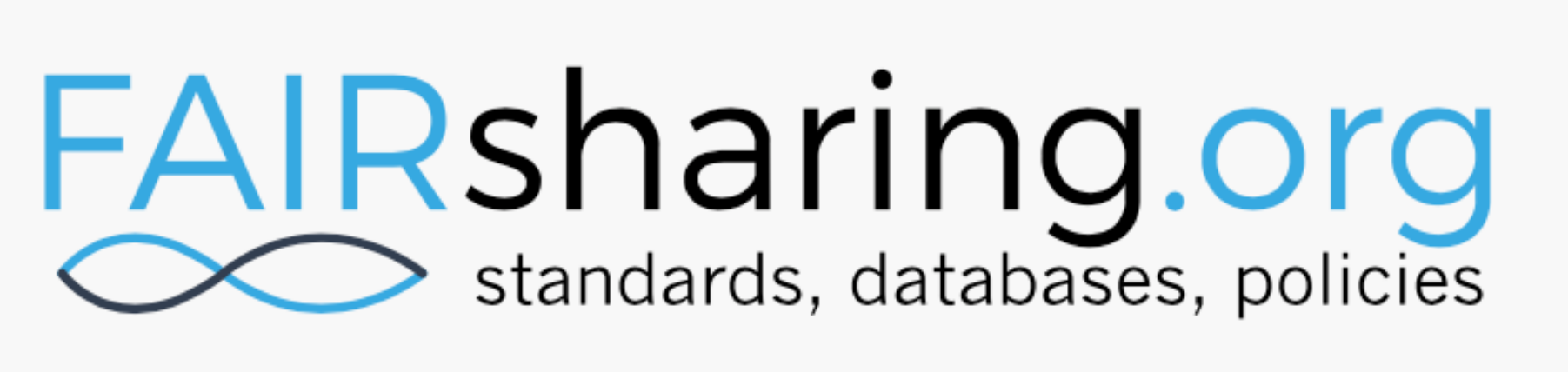
Anyone can register to contribute content, or claim ownership of records created by a 3rd party

CC BY

biosharing.org

Anyone can register to contribute content, or claim ownership of records created by a 3rd party

Recently (2017) biosharing has become discipline agnostic:



CC BY

biosharing.org

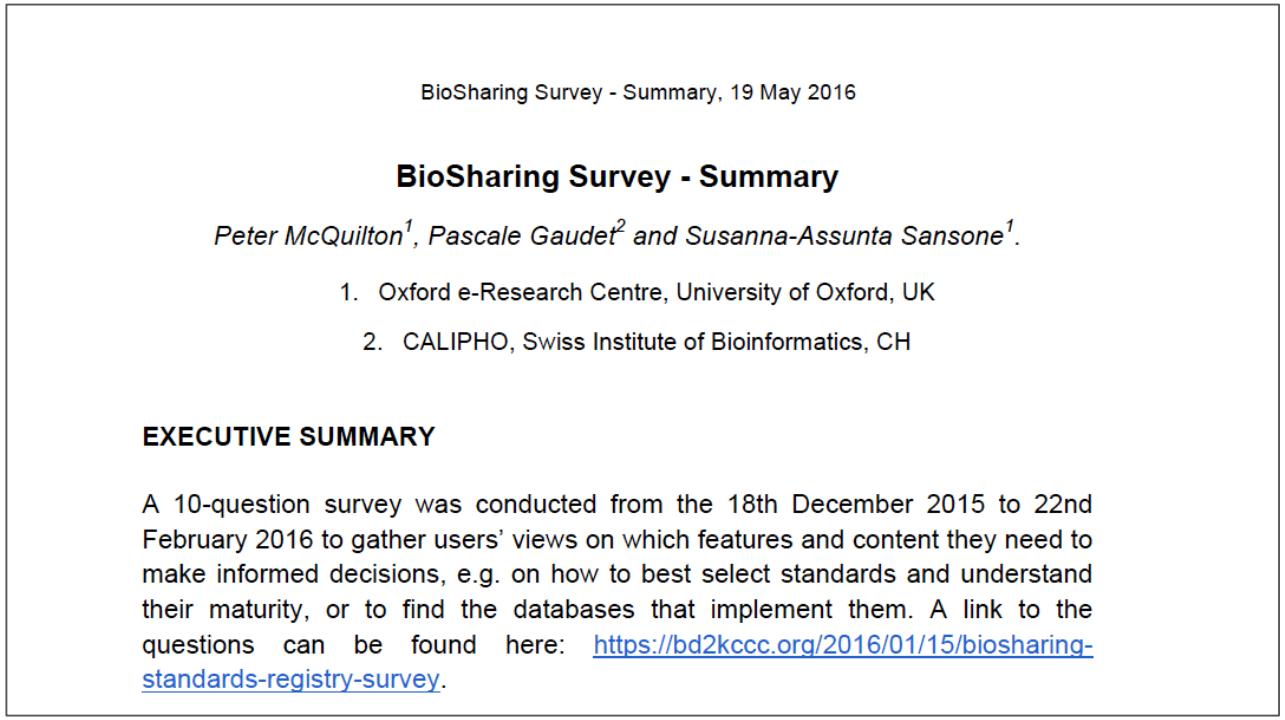
Researcher attitudes

CC BY

biosharing.org

Researcher attitudes

2015/16 survey



BioSharing Survey - Summary, 19 May 2016

BioSharing Survey - Summary

Peter McQuilton[1], Pascale Gaudet[2] and Susanna-Assunta Sansone[1].

1. Oxford e-Research Centre, University of Oxford, UK
2. CALIPHO, Swiss Institute of Bioinformatics, CH

EXECUTIVE SUMMARY

A 10-question survey was conducted from the 18th December 2015 to 22nd February 2016 to gather users' views on which features and content they need to make informed decisions, e.g. on how to best select standards and understand their maturity, or to find the databases that implement them. A link to the questions can be found here: https://bd2kccc.org/2016/01/15/biosharing-standards-registry-survey.

https://figshare.com/articles/New_draft_item/3795810

CC BY

biosharing.org

Researcher attitudes

2015/16 survey - 3 months, 533 respondents

Researcher	323
Tool/database developer	274
Standard Developer/maintainer	206
Data Curator	151
Data Manager	150
Journal Publisher/Editor	31
Librarian	20
Funding Agency	20

- 25% assoc. with Elixir (EU)
- 21% assoc. with NIH BD2K (US)

- 65% aware of biosharing
- 30% had used biosharing

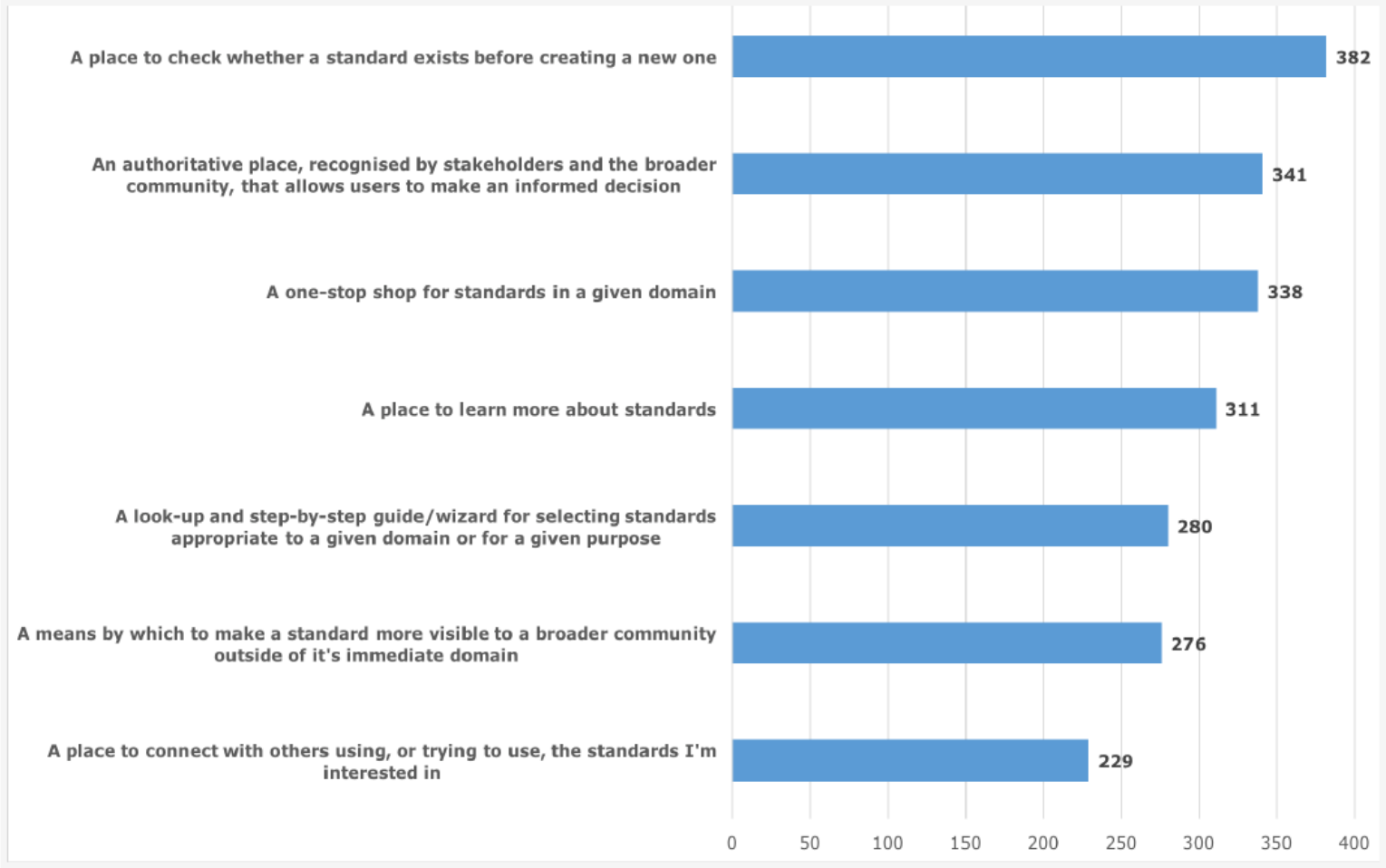
CC BY

biosharing.org

Researcher attitudes

2015/16 survey

What do you need from a standards registry?



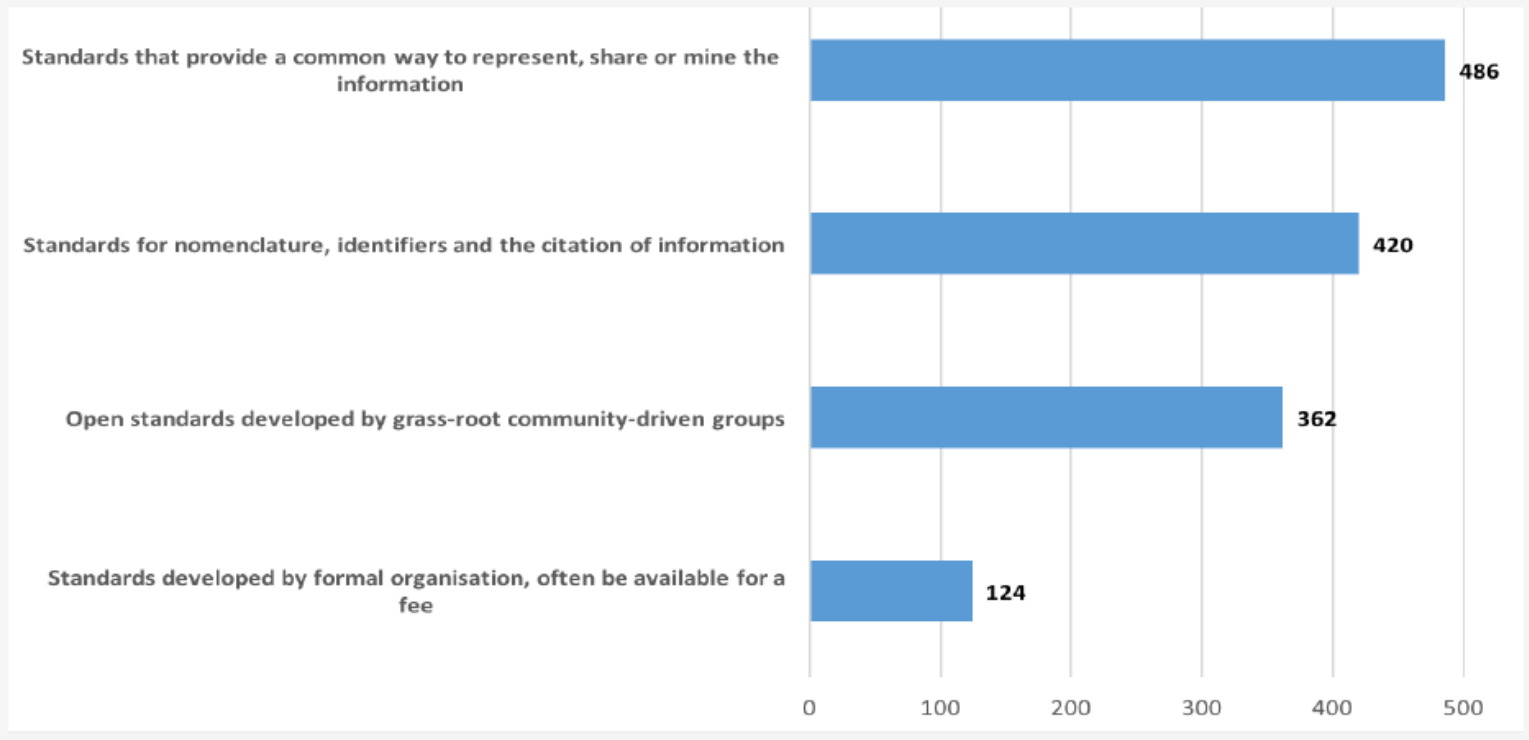
CC BY

biosharing.org

Researcher attitudes

2015/16 survey

What types of standards would you like to see in the registry?

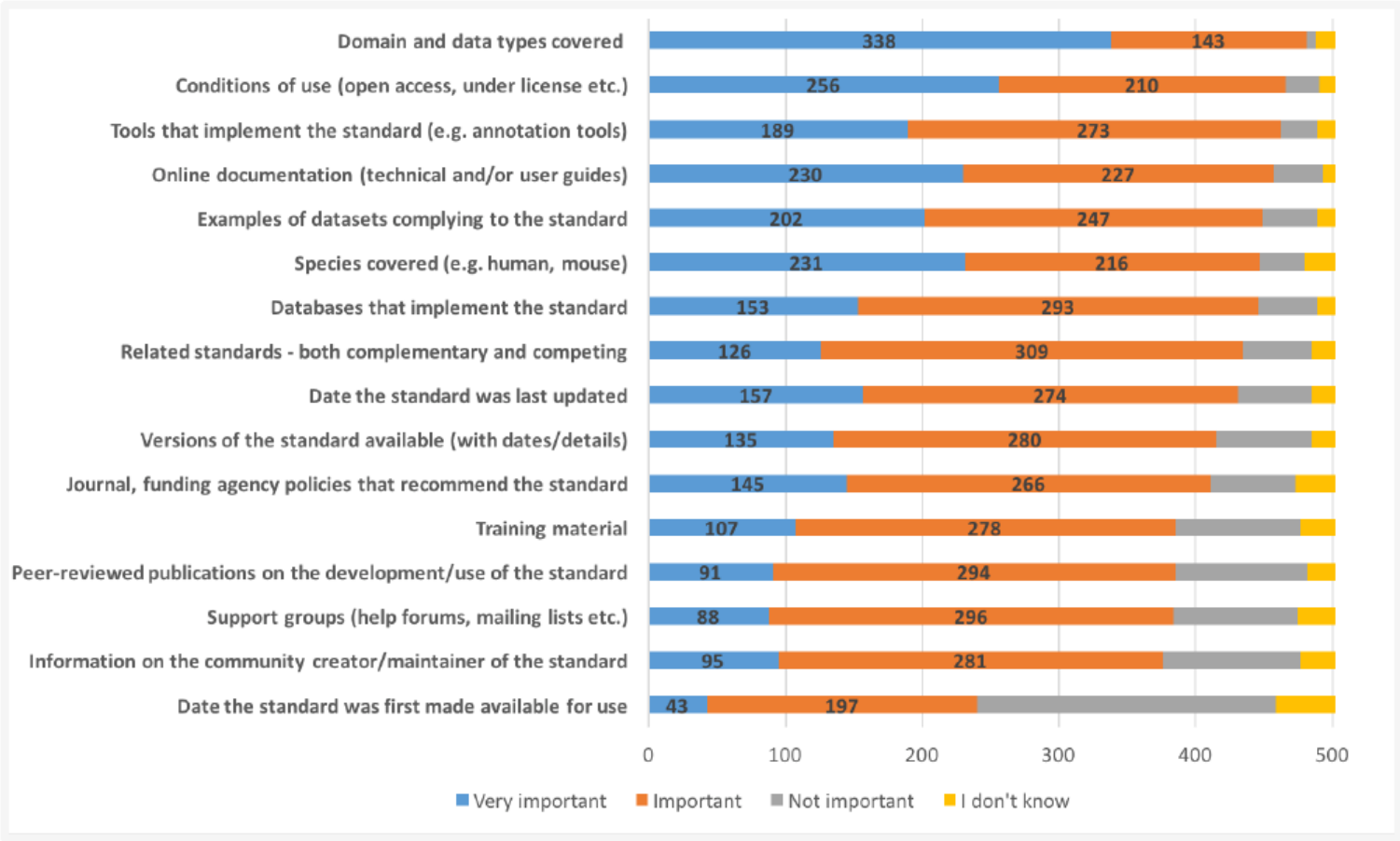


CC BY

biosharing.org

Researcher attitudes

2015/16 survey
What information should a standards registry capture about a standard?



CC BY

biosharing.org

How aware is Australia's bioscience community of Standards?

CC BY

biosharing.org

How aware is Australia's bioscience community of Standards?

EMBL-ABR Standards Key Area: 2017 Survey into Standards

- Aim: to collect information to direct EMBL-ABR node efforts and resources for the maximum impact for the needs of the Australian research community when it comes to standards across key areas in bioinformatics including data, tools, workflows and training

- Promotion: via each node (email lists, social media, etc.)

- Response rate: extremely low (*n*=15) *

CC BY

* previous data and tool surveys have previously attracted ~100 responses

biosharing.org

How aware is Australia's bioscience community of Standards?

EMBL-ABR Standards Key Area: 2017 Survey into Standards

- Is interest in and/or understanding of standards by biology and bioinformatics researchers in Australia therefore low?

- The general opinion is "yes"

CC BY

biosharing.org

How aware is Australia's bioscience community of Standards?

EMBL-ABR Standards Key Area: 2017 Survey into Standards

- Education is required
- For researchers: EMBL-ABR Standards webpages updated to provide better context about:
  - what standards are
  - why they are important
  - where to find out about them (i.e. biosharing.org)

CC BY

biosharing.org

How aware is Australia's bioscience community of Standards?

EMBL-ABR Standards Key Area: 2017 Survey into Standards

- Education is required
- For eResearch practitioners and funders: The EMBL-ABR QCIF Node is championing the building of national eResearch item level metadata infrastructure to support FAIRsharing.org (or other) standards (underlying schemas / information templates reflecting minimum information standards, controlled vocabs, etc) so that researchers can just use them

CC BY

thanks

Jeff Christiansen
EMBL-ABR Key Area Coordinator
j.christiansen@uq.edu.au

Thank You!


RDA
11 PLENARY
MEETING
21-23 MARCH 2018
Berlin, Germany
RDA
RESEARCH DATA ALLIANCE
RESEARCH DATA SHARING WITHOUT BARRIERS


INTERNATIONAL DATA WEEK
IDW 2018
Gaborone, Botswana: 22–26 October 2018

RESEARCH DATA ALLIANCE

Questions

research data sharing without barriers

rd-alliance.org