RESEARCH DATA ALLIANCE

RDA 7th Plenary
Overview and outputs

Irina Kupiainen

31.3.2016

research data sharing without barriers

rd-alliance.org

Content

- General overview of Plenary 7
- Introducing the new outputs
- Adoption cases
- RDA outputs evaluation for ICT technical specifications by EC
- Questions to be addressed
- Links and contact information

GENERAL OVERVIEW

7 RDA Recommendations/outputs presented:

- Repository Audit and Certification DSA–WDS
- RDA/WDS Publishing Data Bibliometrics
- RDA/WDS Publishing Data Services
- RDA/WDS Publishing Data Workflows
- Wheat Data Interoperability Recommendations
- RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World Interim Recommendations
- Brokering Governance Interim Recommendations
- **11 adoption presentations**

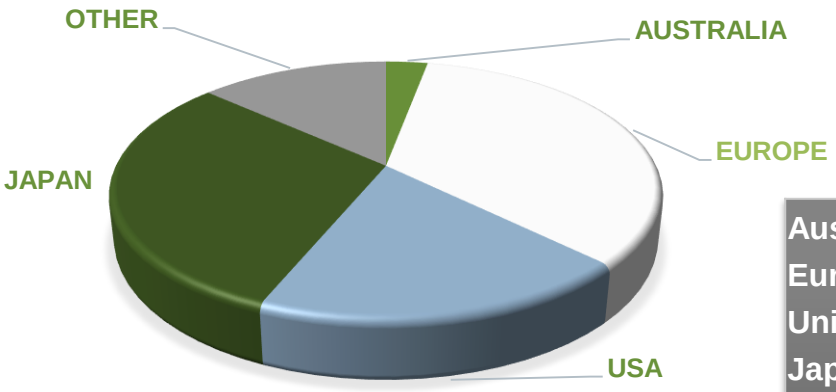
RDA Plenary 7 in Tokio 1-3 March 2016 Statistics

- 373 participants from 33 countries
- 30 international speakers over 5 plenary sessions, including EC high-level representatives
- 7 outputs and 11 adoption cases
- 8 Working Group meetings
- 25 Interest Group meetings
- 10 BoF sessions
- 9 Joint meetings
- 2 Organisational member meetings
- RDA for Newcomers meetings

RDA RESEARCH DATA ALLIANCE

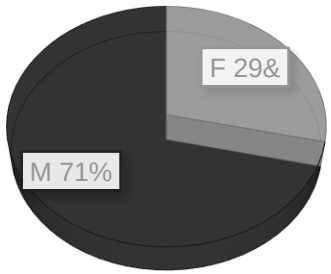
JST Japan Science and Technology Agency

REGIONAL BREAKDOWN



Australia	11	3%
Europe	128	34%
United States	71	19%
Japan	113	30%
Other	50	13%
Total Regional Breakdown	373	

GENDER BALANCE



Total Attendees: 373

of which 30% Japanese & 32% "first timer" attendees

rd-alliance.org

RESEARCH DATA ALLIANCE

Making data sharing work in the era of Open Science

First in Asia

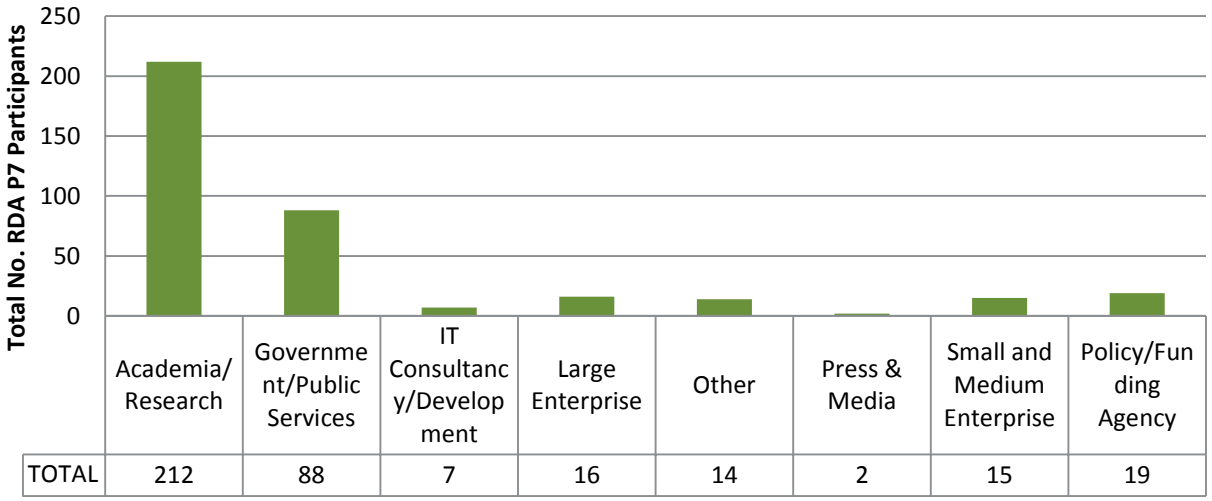
RDA 7th Plenary in Tokyo, Japan
March 1st to 3rd, 2016

Data Sharing Symposium - pre-RDA Plenary
February 29th, 2016

RDA RESEARCH DATA ALLIANCE

JST Japan Science and Technology Agency

Participants breakdown by type of organisation



Type of Organisation	TOTAL	%
Academia/Research	212	57%
Government/Public Services	88	24%
IT Consultancy/Development	7	2%
Large Enterprise	16	4%
Other	14	4%
Press & Media	2	1%
Small and Medium Enterprise	15	4%
Policy/Funding Agency	19	5%

RESEARCH DATA ALLIANCE

RESEARCH DATA ALLIANCE

New outputs

research data sharing without barriers

rd-alliance.org

RESEARCH DATA ALLIANCE

RDA-WDS Publishing Data Bibliometrics Recommendations

research data sharing without barriers

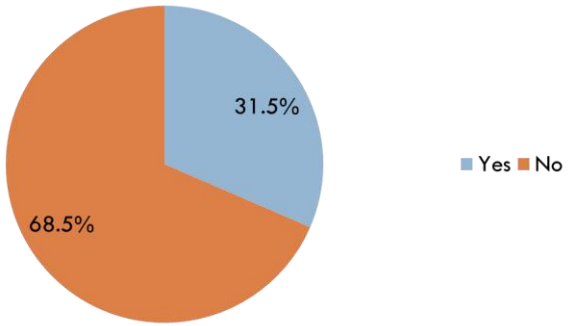
rd-alliance.org

RDA-WDS Publishing Data Bibliometrics Recommendations

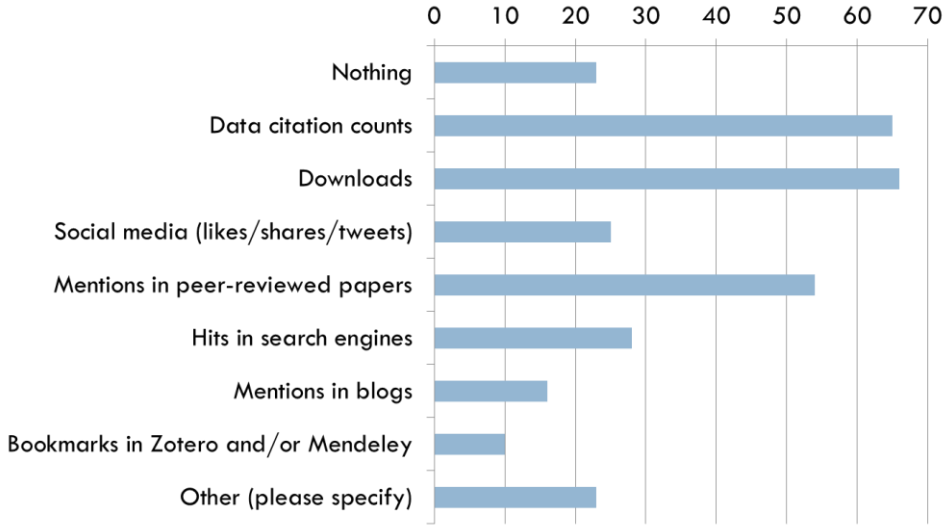
- To understand the impact and value of data being shared and distributed, lack of assessment framwork creates barriers to data sharing.
- Objective: to conceptualize data metrics and corresponding services
- 63 WG members from 20 countries
- Landscape survey, identification of focus areas
- Adopters have used this information to develop data metrics.
- Work within RDA continues, anyone is welcomed to join and contribute!

Summary of Survey of current status/opinions on data bibliometrics

Are the methods you use to evaluate impact adequate for your needs?



What do you currently use to evaluate the impact of data?



What is currently missing and/or needs to be created for bibliometrics for data to become widely used? (n-92)

1) Standards
2) Data Citation
3) Consistent use of PIDs/DOIs
4) Culture change/"A belief that they are valid"

Bibliometrics endorsements/adopters

- National Information Standards Organization (NISO)
- California Digital Library: Making Data Count project (NSF funded)
- JISC Giving Researchers Credit for their Data
- CASRAI Dataset Level Metrics Group
- Re3data.org schema

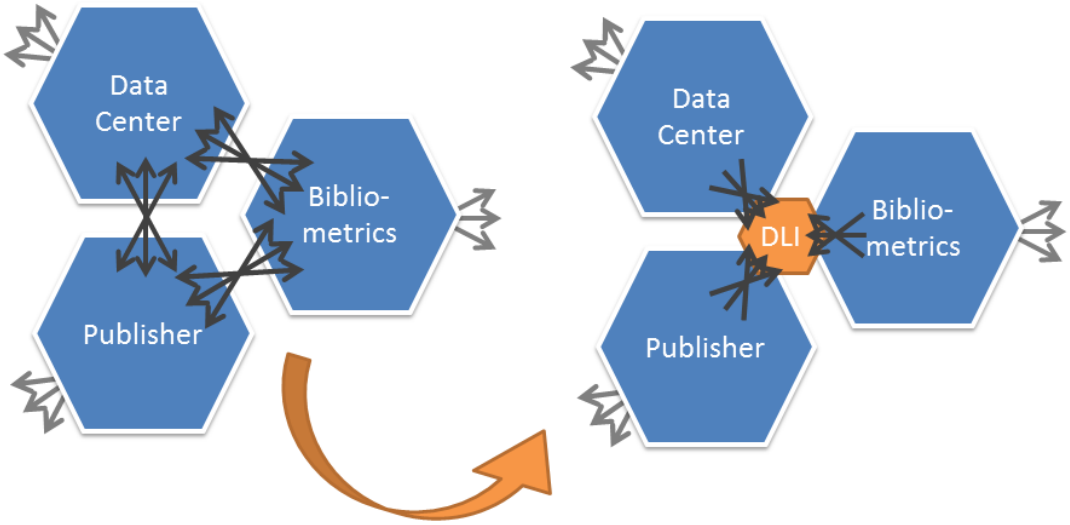
ICSU-WDS-RDA Data Publication Services

research data sharing without barriers
rd-alliance.org

- Finding new solutions for linking research data and literature, to increase visibility and discoverability, enable proper re-use, and support credit attribution.

What do we propose?

Objective: move from a plethora of (mostly) bilateral arrangements to a one-for-all service model infrastructure for the research data publication landscape



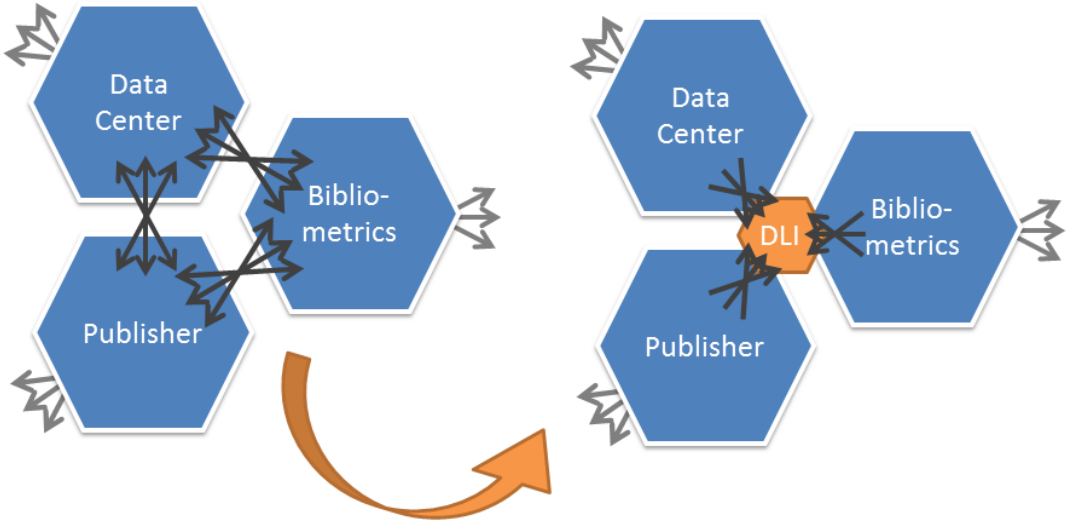
1. Increase interoperability
2. Decrease systemic inefficiencies
3. Power new tools and functionalities to the benefit of researchers

ICSU WORLD DATA SYSTEM

RDA RESEARCH DATA ALLIANCE

So.. what will this “orange blob” do?

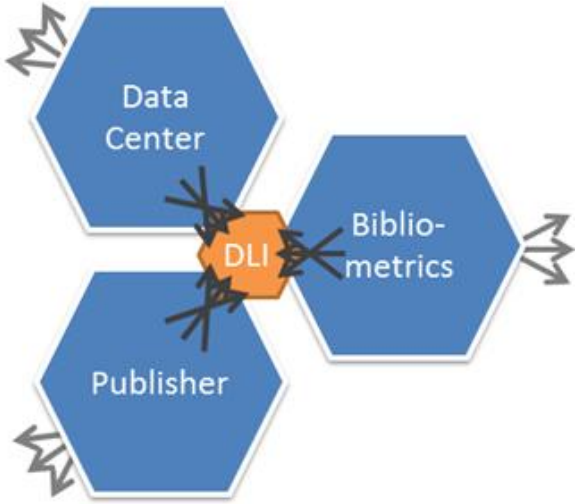
- Given article A, what relevant data D exists – and vice versa
- Additional metadata about the nature of the relationship, e.g. supplementary data, related data, formal citation.
- Additional metadata for article and/or data set



General Recommendations

The ideal data / literature interlinking service

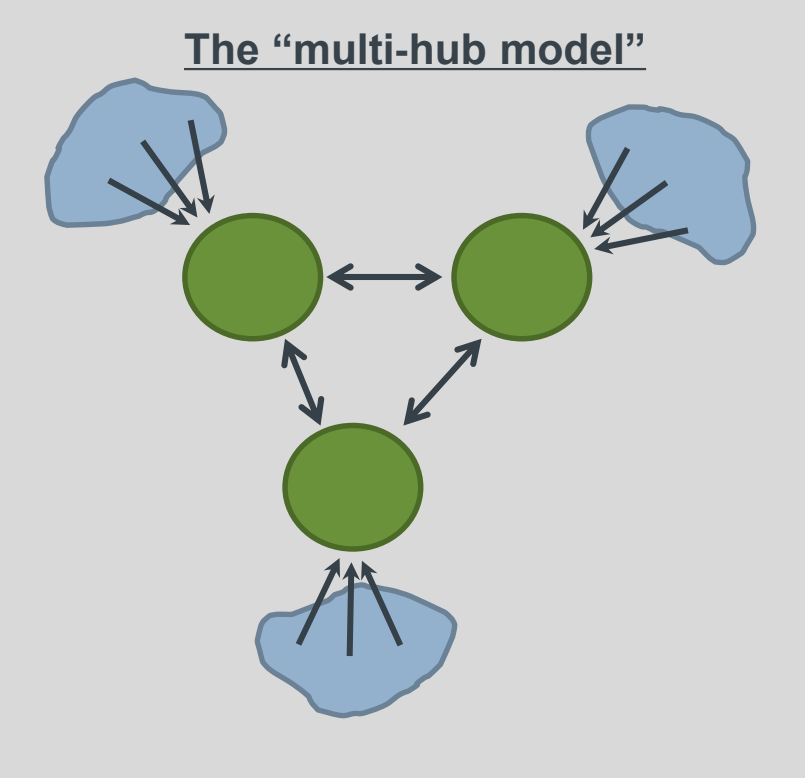
- Universal: cross-disciplinary, global
- Inclusive and participatory: supported by all stakeholder groups
- Open and non-discriminatory
- Quality through meticulous provenance and metadata (*not* "filtering at the gate")
- Standards-based



General Recommendations – long-term view

The ideal data / literature interlinking service

- Infrastructure & service layer
- Create sustainable infrastructure as extensions of existing systems
- "Follow the content": use established processes as natural aggregation points ("hubs") for different constituencies
- Interoperability between the hubs through common standards
- Inclusive – new hubs welcome

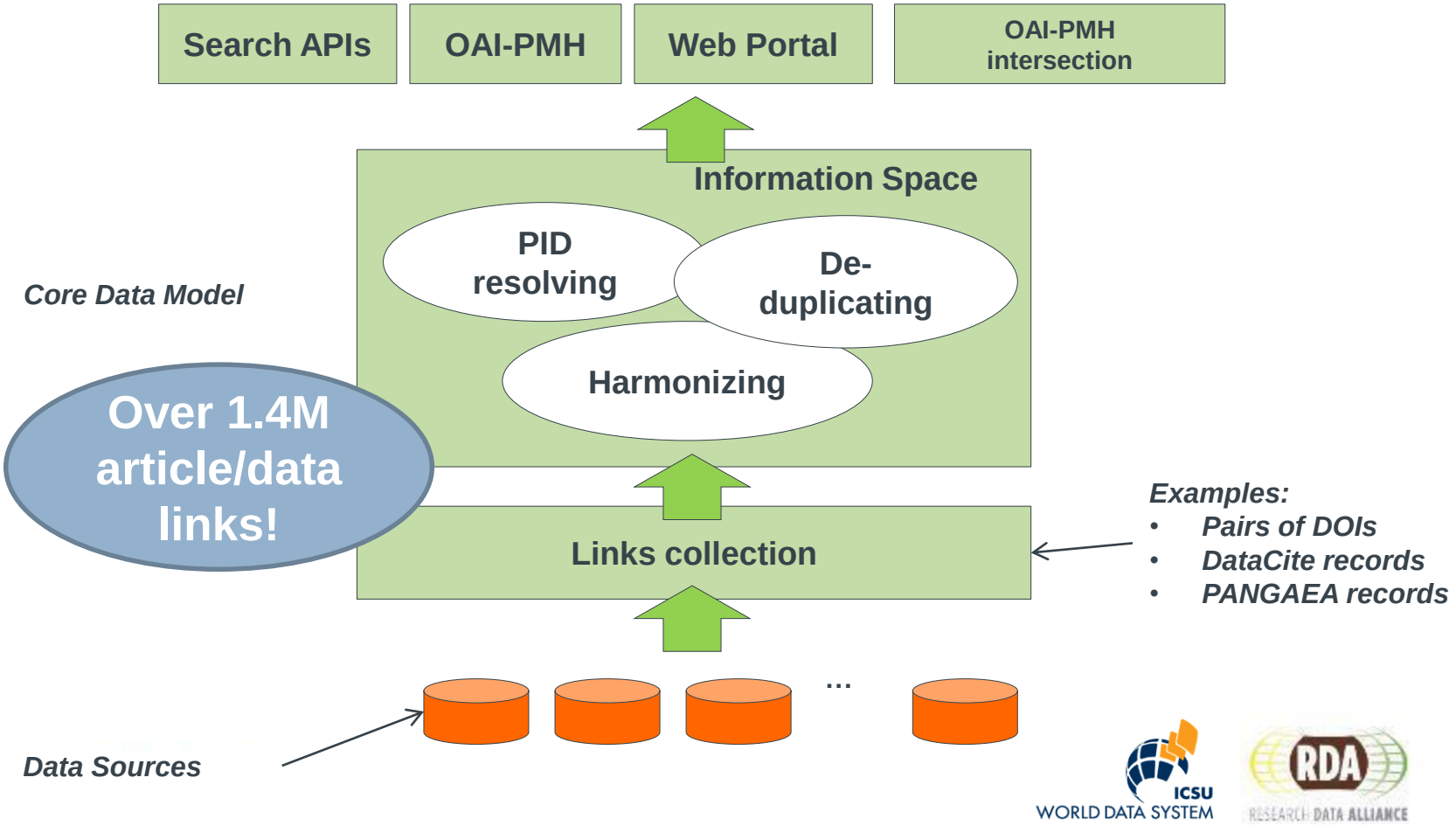


ICSU WORLD DATA SYSTEM

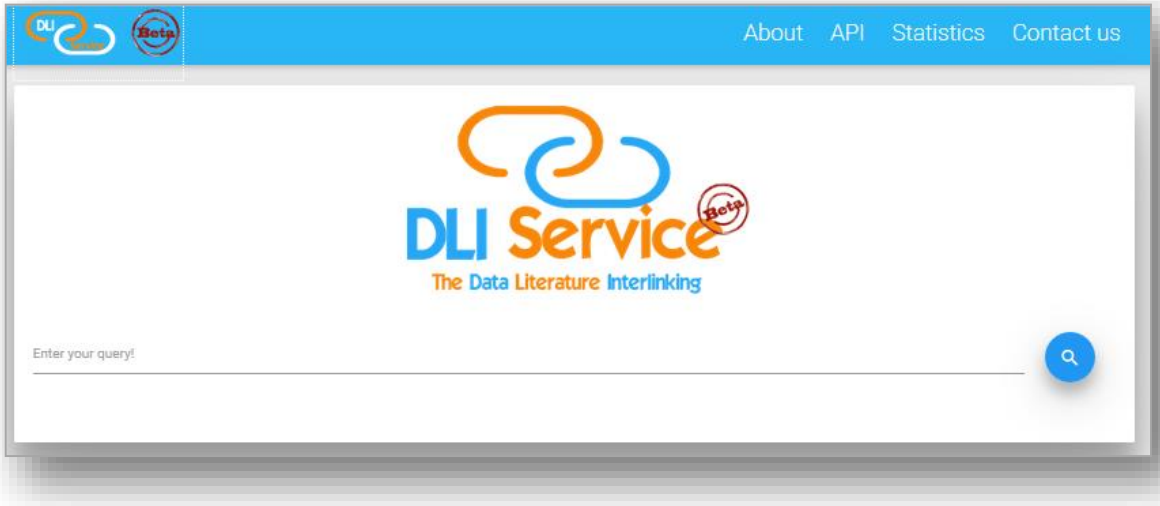
RDA RESEARCH DATA ALLIANCE

WG Output: the "Data-Literature Interlinking (DLI)" Service

(Prototype) interlinking service developed with OpenAIRE and PANGAEA



WG Output: the “Data-Literature Interlinking (DLI)”



Powered by OpenAIRE D-NET software and PANGAEA search engine

Give it a spin: http://dliservice.research-infrastructures.eu

Examples of adoption & implementation

- Long-term view is adopted by CrossRef, DataCite and OpenAIRE – supporting the “multi-hub system” infrastructure
- Europe-PMC has adopted the DLI metadata standards
- Connected with RD-switchboard (output of RDA WG “Data Description Registry Interoperability”)
- Ad-hoc information requests on linked data
- Several data repositories are exploring connection with current API interface.

RESEARCH DATA ALLIANCE

WDS/RDA Publishing Data Workflows: Working groups outputs

research data sharing without barriers

rd-alliance.org

WDS/RDA Publishing Data Workflows

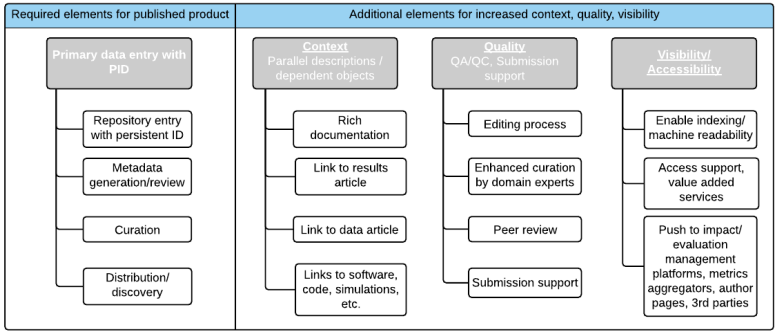
- Aiming for an analysis of a representative range of existing and emerging workflows and standards for data publishing, incl. deposit and citation, and providing reference models - a “classification”
- Testing implementations of key components for application in new workflows
- Illustrating benefits of reference models for researchers and organisations
- Stakeholders: researchers and research projects, data publishers (repositories and journal publishers), research workflow developers

Highlights of the Deliverable (1/2)

Recommendations

1. Start small, building modular, open source and shareable components
2. Implement core components of the reference model according to the needs of the stakeholder
3. Follow standards that facilitate interoperability and permit extensions
4. Facilitate data citation, e.g. through use of digital object PIDs, data/article linkages, researcher PIDs
5. Document roles, workflows and services

Key components



Definitions

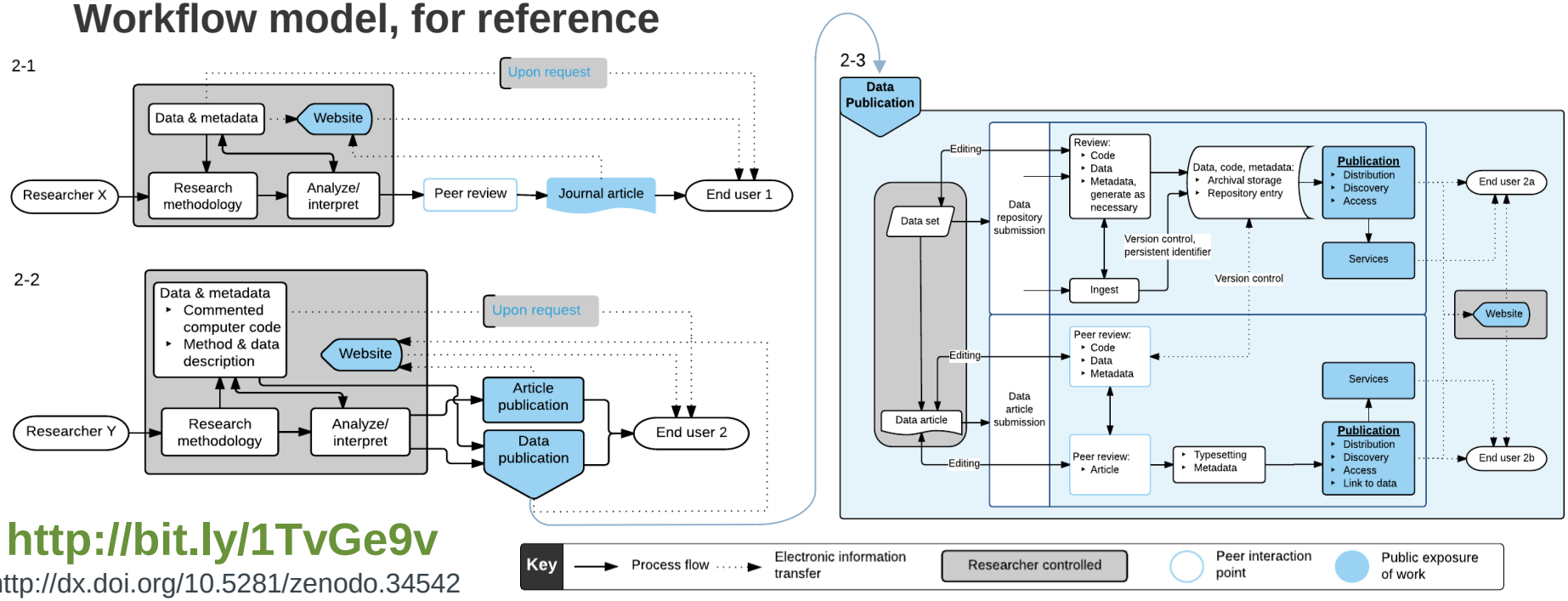
- Entered into RDA Term Definition Tool: TeD-T **http://bit.ly/TeDT-RDA**
- Included in Research Data Canada / CASRAI Glossary: Research Data Domain **http://bit.ly/1KY3XzP**

Austin, Claire C et al.. (2015). Key components of data publishing: Using current best practices to develop a reference model for data publishing. Zenodo. http://dx.doi.org/10.5281/zenodo.34542

http://bit.ly/1TvGe9v

Highlights of the Deliverable (2/2)

Workflow model, for reference



http://bit.ly/1TvGe9v

http://dx.doi.org/10.5281/zenodo.34542

Preliminary report:

- Connecting data publication to the research workflow: a preliminary analysis

Endorsements/Adopters

- Giving Researchers Credit for their Data – app to support researchers in submitting data papers directly to journals, developed by the WG
- Academic Commons at Columbia University: WDS/RDS reference model in use and as benchmarking tool
- Elsevier Research Data Management Solutions
- ISPS Data Archive, Yale
- Digital Curation Centre (DCC)
- Research Space
- Edinburg University Data Library
- GigaScience
- Scientific Data

Who can use this deliverable?

Do you: Can you us this?

- Generate publishable research data?.....**YES**
- Publish research data?.....**YES**
- Fund publishable research data?.....**YES**
- Have interest in any of the above?.....**YES**

How?

- Follow recommendations: bit.ly/24cNfmG http://dx.doi.org/10.5281/zenodo.34542
- Sign on as an adopter

OECD Global Science Forum Project on Sustainable Business Models for Data Repositories

research data sharing without barriers

rd-alliance.org

OECD Global Science Forum Project on Sustainable Business Models for Data Repositories

- Sustaining digital data infrastructure is a major issue for science policy
- Need for developing business models.
- OECD work will build partly RDA IG on Cost Recovery: landscape survey on 25 data repositories: Final Report: Income Streams for Data Repositories https://rd-alliance.org/system/files/documents/Income_Streams_for_Data_Repositories-FINAL-160210.pdf

Preliminary Conclusions from Survey and Analysis

- Structural funding supports c.50% of repositories surveyed.
- Structural funding often supplemented and some concerns expressed about flexibility and adaptability.
- Many data repositories value participation in R&D projects, and many are highly dependent on this grant income but overheads need to be considered.
- Concern about administrative overheads and that encourage cheaper, lower levels of curation.
- Many repositories interested in charging for value-added services.
- Data deposit fees are being explored by a small number of repositories.

Impact and Adoption: OECD GSF Project on Business Models

- Questions to address:
  1. How are data repositories currently funded?
  2. What innovative income streams are available?
  3. How do income streams match willingness/ability to pay of various stakeholders?
  4. How do income streams/willingness to pay fit together into a sustainable business model?

- Builds on existing work of RDA-WDS Working Group.
- Broader landscape study of current funding models.
- Focus group on innovative income streams.
- In-depth economic analysis of business models.
- Test business models with stakeholder groups.
- ➢ **Policy recommendations based on concrete business model options.**

RESEARCH DATA ALLIANCE

Core Common Certification Requirements and Procedures

research data sharing without barriers

rd-alliance.org

Core Common Certification Requirements and Procedures

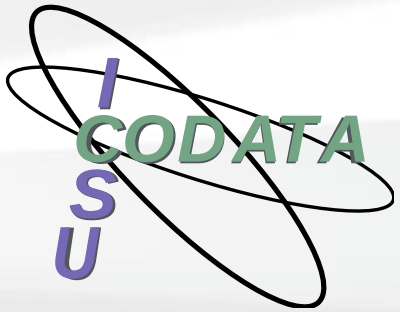
- Harmonizing core certification requirements and procedures, ultimately setting a stage for a global shared framework including other standards
- Implementation plan for Common Procedures
- Testbed – “real world” evaluation of common requirements and procedures
- Aims for more coherent, increasingly stringent and compatible standards for repository certification – ultimately a critical mass of certified repositories across a range of domains and disciplines
- Increasing trust among data collectors, funders, publishers and users
- IG on Repository Audit and Certification ongoing

Adopters

- ICSU World Data System
- Data Seal of Approval
- CLARIN
- IOC International Oceanographic Data and Information Exchange (IODE) Programme
- Other repositories

How You Can Endorse

- Who? Data repositories (data centres, data analysis services) – and also Funders, Collectors, Publishers and Users
- How?
- Common Core certification requirements/criteria (http://tinyurl.com/pm9sflp)
  - Implementation plan for Common Procedures (http://tinyurl.com/os6vb94)
  - Testbed – "Real-world" valuation of Common Requirements and Procedures
  - Also through DSA and WDS websites
  - Testbed results through WG webpage: https://rd-alliance.org/group/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg/outcomes/dsa-wds-partership



CODATA-RDA Research Data Science Short Courses

research data sharing without barriers

rd-alliance.org

Deliverables from the Working Group

- A curriculum for a broad introductory course in ‘research data science’.
- Reusable materials available for online delivery (as trialled for the champions and used in Trieste.
- School in Trieste and a first cohort of ‘champions’ to run schools internationally.
- A model for expansion and creating a sustainable network of schools for a broad research data science curriculum.
- A prototype business model and plans for sustainability.

Introductory Research Data Science Curriculum

- **Open Science** – reflection on ethos and requirements of sharing/openness
- **Open Research Data** – Data Publishing, Life-Cycle, Metadata and annotation
- **Data Carpentry** – Introduction to SQL databases
- **Software Carpentry** – Introduction to programming in R, the Unix shell and Git (sharing software and data)
- **Visualisation** – Tools, Critical Analysis of Visualisation
- **Analysis** – Statistics and Machine Learning (Clustering, supervised and unsupervised learning)
- **Computational Infrastructures** – Introduction to cloud computing, launching a Virtual Machine on an IaaS cloud

Impact of the Deliverable

- First Introductory Course, 1-12 August, ICTP, Trieste.
- Subsidised accommodation and meals for up to 120 students.
- So far: 55K euros funding for students and tutors committed by ICTP, TWAS, CODATA, GEO, ACU and RDA Europe and GEO.
- Strong emphasis on training new teachers for courses in 2017 (online preparation and ‘champion’ role).
- Deadline for applications is 18 April 2016: http://indico.ictp.it/event/7658/

The Abdus Salam International Centre for Theoretical Physics

research data sharing without barriers
rd-alliance.org

CODATA

RDA RESEARCH DATA ALLIANCE

Impact of the Deliverable

- Beijing Introductory Research Data Science School, 4-15 July, CNIC, Beijing, China.
- CODATA China in collaboration with the WG, CNIC and RADI.
- Scholarships available for c.20 students from LMICs.
- Introductory course will follow the basic curriculum designed by the CODATA-RDA Working Group.
- **Application process to be announced very shortly on the CODATA website: http://www.codata.org**

CHINESE ACADEMY OF SCIENCES

CODATA

Impact of the Deliverable: Next Steps

- Online course for ‘champions’ from May in preparation for the first introductory ICTP Trieste School in August 2016.
- Materials from the first school will be openly available for reuse and online study.
- Repeat of introductory and a new advanced school planned for summer 2017.
- Plans for regional introductory schools in Brazil, India and South Africa + *Indonesia, Kenya?*
- Plans for regional specialised schools in South Africa (Social Sciences and Bio-Informatics)…
- **Create a sustainable network of schools, training teachers, reusing materials, hub and node funding…**

Research Data Science Champions
Teach-New-Teachers

Wheat Data Interoperability WG outputs

research data sharing without barriers
rd-alliance.org

The deliverables

- Guidelines (http://wheatis.org/DataStandards.php)
  - Data exchange formats
    - Example: VCF (Variant Call Format) for sequence variation data, GFF3 for genome annotation data, etc.
  - Data description best practices
    - Consistent use of ontologies, consistent use of external database cross references
  - Data sharing best practices
    - Share data matrices along with relevant metadata (example: trait along with method, units and scales or environmental ones)
  - Useful tools and use cases that highlight data formats and vocabularies issues
- A portal of wheat related ontologies and vocabularies (http://wheat.agroportal.lirmm.fr/ontologies)
  - Allows the access to the ontologies and vocabularies through APIs.
- A prototype
  - Implementation of use cases of wheat data integration within the AgroLD (Agronomic Linked Data) tool: http://volvestre.cirad.fr:8080/agrold/

Benefits for many target users

As a data producer or manager

- Easily conform to the well-recognized data repositories and facilitate the deposit of your data within these repositories;
- Share common meanings of the words you utilize to describe your data and make your data more machine-readable and computable
- Contribute to foster the development of smarter search tools and make your data more visible and discoverable

As a wheat related information system or tool developer

- Basing your tool or information system on the recommended data formats and vocabularies will make it easier to integrate data from various data sources, deliver smarter outputs for a wider audience

As a wheat related ontology developer

- Share your ontologies through the WDI wheat ontologies portal and make them more visible to the community
- Reuse or link your ontologies to existing concepts and terms in wheat related ontologies to enrich them, make them more visible and in some cases save you time.

RESEARCH DATA ALLIANCE

RDA Europe funding

Collaboration Projects (testing/adopting RDA Outputs)

Peter Wittenburg

research data sharing without barriers

rd-alliance.org

1st Call for Collaboration Projects

- testing & adopting RDA WG outputs

- small 6 months projects building on existing infrastructures
- max. amount of funding 15.000 €

- 25 applications
- 7 passed criteria (certainly not enough!)
- currently working on administrative details

- in April 2nd call

Adoption projects

- Dynamic Data Citation – adopted by ARGO Team together with CrossRef and DataCite
- Metadata Standards Directory – adoption by DMPOnline
- Data Foundation and Terminology – adoption by CLARIN
- Dynamic Data Citation – adoption by VAMDC (Virtual Atomic and Molecular Data Centre)
- Publishing Data Service – adoption by OpenAIRE
- PID Information Types and Data Type Registry – adoption by National Academy of Sciences in Armenia

RESEARCH DATA ALLIANCE

The RDA Technical Specifications: EC evaluation

research data sharing without barriers

rd-alliance.org

Identification of ICT specifications in Europe

- **The European Commission has a flexible approach to standardisation when identifying new ICT technical**

WHY?

The European Commission can identify ICT technical specifications that are not national, European, or international standards, provided they meet precise requirements. Once identified and approved, these specifications can then be referenced in European public procurement. This flexible approach allows the EU to respond to the fast evolution of technology in ICT. It also helps encourage competition, promote interoperability and innovation, and facilitate the provision of cross-border services.

The Research Data Alliance was invited to present the first 4 RDA Outputs under this scheme.

Who is involved in this process?

The European Multi Stakeholder Platform (MSP) is an expert advisory group on ICT standardisation. It sets up evaluation groups to examine the compliance of technical specifications in the field of ICT that are not national, European or international standards based on a set of requirements.

The Multistakeholder platform (MSP) is chaired and coordinated by the European Commission.

Who is on the MSP?

Member States and EFTA countries

ICT Standardisation Bodies

CEN
CENELEC EUROPEAN COMMITTEE FOR ELECTROTECHNICAL STANDARDIZATION
ETSI 2013 Celebrating 25 years
IEC
GS1
ISO
IEEE Advancing Technology for Humanity
ITU
ecma INTERNATIONAL
IETF
THE Open GROUP
OASIS Advancing open standards for the information society
OMG WE SET THE STANDARD
oma Open Mobile Alliance
UN/CEFACT SIMPLE, TRANSPARENT AND EFFECTIVE PROCESSES FOR GLOBAL BUSINESS
W3C

Industry, SMEs and society representatives

BUSINESSEUROPE
Cable Europe
DIGITALEUROPE
EDiMA Shaping Internet Policy in the EU
ecta
Openforum europe open, competitive choice for IT users
Small Business Standards
ECOS
GSMA
ANEC
EBU
EuroISPA
etno
european disability forum
ECIS European Committee for Interoperable Systems

RDA Compliance with Requirements for ICT Technical Specifications

ANNEX II: Requirements

REQUIREMENTS

- ✓(a) maintenance
- ✓(b) availability
- ✓(c) intellectual property rights
- ✓(d) relevance
- ✓(e) neutrality and stability
- ✓(f) quality

- Market acceptance & promote interoperability
- No conflict with EU standards
- Developed by a non-profit making organisation which is a professional society, industry or trade association or any other membership organisation

Organisational Processes of the organisations must fulfil the following criteria:

- Openness: the technical specifications were developed on the basis of open decision-making
- Consensus: decision-making process was collaborative and consensus based
- Transparency

How are new ICT specifications identified?

1. **Application by RDA to assess technical specifications – submission of details on organisational processes & RDA outputs:**
   - **TS1: Data Foundation & Terminology Model**
   - **TS2: PID Information Types API**
   - **TS3: Data Type Registries Model**
   - **TS4: Practical Policy**
2. **Favourable assessment and invitation to present to the Multistakeholder platform (26 November 2015), as a result of which …**
3. **Evaluation Committee set up and Evaluation process started in December 2015.**

The Research Data Alliance was favourably evaluated and is now in step 3.

4. **Feedback – identification of specifications or reapplication**

RESEARCH DATA ALLIANCE

How to get involved?

research data sharing without barriers

rd-alliance.org

Adopt and participate!

Recommendation and outputs:

https://rd-alliance.org/recommendations-and-outcomes/all-recommendations-and-outcomes

Working and Interest Groups:

https://rd-alliance.org/groups

Thank you!

research data sharing without barriers

rd-alliance.org