D3.2 INTERIM ICT TECHNICAL SPECIFICATIONS

Project Full Title	RDA Europe 4.0 – The European plug-in to the global Research Data Alliance
Project Acronym	RDA Europe 4.0
Project Number	777388
Start Date of Project	01.03.2018
Duration of Project	27 Months
Project Website	www.rd-alliance.org/rda-europe
Deliverable Title	Interim ICT Technical Specifications
Deliverable No.	D3.2
Author(s)	Sara Garavelli, Marieke Willems (Trust-IT Services)
Abstract	The first interim report tracks and reports the RDA technical specifications referenced in public procurement.

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

The work described in this document has been conducted within the project RDA Europe 4.0. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 777388. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

Document Information

Deliverable Title	Interim ICT Technical Specifications
Deliverable No.	3.2
Contractual Delivery Date	31 December 2018
Actual Delivery Date	
Author(s)	Sara Garavelli, Marieke Willems (Trust-IT Services)
Editor(s)	
Contributor(s)	
Reviewer(s)	Natalie Harrower (RIA), Daniel Bangert (UGOE)
Work Package Number & Title	WP3 Support to RDA involvement in European policies
Work Package Leader	UGOE
Work Package Participants	TRUST-IT, RIA, UEDIN

Versioning and Revision History

Version	Date	Author	Notes
0.1	10.12.18	Trust-IT	ToC
0.2	15.01.19	Trust-IT	First Draft Ready
0.3	22.01.19	Trust-IT	Liaison with RDA WGs to collect information on new candidate ICT Technical specifications and with EC to understand the status of the second wave of recommendations
0.4	04.02.19	Trust-IT	Deliverable ready to be reviewed
0.5	06.02.19	RIA	Comments by Natalie Harrower
0.6	12.02.19	UGOE	Comments by Daniel Bangert
0.7	18.02.19	Trust-IT	Final version for EB approval
1.0	22.02.18	Trust-IT	Deliverable submitted

Disclaimer

RDA Europe 4 (777388) is a Coordination & Support Action co-funded by the European Commission under the Research and Innovation Framework Programme, Horizon 2020 (H2020).

This document contains information on RDA Europe (Research Data Alliance Europe 4.0) core activities. Any reference to content in this document should clearly indicate the authors, source, organisation, and publication date.

The document has been produced with the funding of the European Commission. The content of this publication is the sole responsibility of the RDA Europe Consortium, and it cannot be considered to reflect the views of the European Commission. The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

The European Union (EU) was established in accordance with the Treaty on the European Union (Maastricht). There are currently 28 member states of the European Union. It is based on the European Communities and the member states' cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice, and the Court of Auditors (http://europa.eu/index_en.htm).

Glossary

Abbreviation	Definition
EC	European Commission
EU	European Union
ICT	Information & Communication Technology
IG	Interest Group
MSP	Multi-Stakeholders Platform
RDA	Research Data Alliance
SDO	Standards Development Organisation
RIA	Royal Irish Academy
TRUST-IT	Trust-IT Services
UEDIN	University of Edinburgh
UGOE	University of Göttingen
WG	Working Group

Table of Contents

List of Tables

List of Figures

Executive Summary

RDA Recommendations and Outputs are the technical and social infrastructure solutions developed by RDA Working Groups or Interest Groups that enable data sharing, exchange, and interoperability. These outputs have an important impact in two areas: i) solving problems, and ii) incorporation and/or adoption in infrastructure environments by individuals, projects, and organizations. RDA's goal is to expand awareness and adoption of these outputs, and hence their impact, within all regions of the world.

RDA is not a standards development organisation (SDO), but has been building synergies with SDOs and similar organisations so that RDA Recommendations can either get a fast track to becoming standards, be incorporated in or contribute to common goals and activities, and overall foster adoption and build trust and engagement in new user communities.

In particular, RDA Europe is liaising with the European Multi-Stakeholders Platform on ICT Standardisation (MSP - E02758), an advisory expert group on ICT standardisation.

At present, two waves of outputs from RDA Working Groups have been submitted to the European Multi-Stakeholders Platform on ICT Standardisation (MSP - E02758) to be recognised as ICT Technical Specifications. These were developed via the clear and transparent RDA process for producing and endorsing outputs as RDA Recommendations.

Currently eight RDA Recommendations have been recognised as ICT technical specifications:

- *TS1 Data Foundation & Terminology Model**
- *TS2 PID Information Types API- Persistent Identifier Type Registry**
- *TS3 Data Type Registries Model**
- *TS4 Practical Policies recommendations**
- *TS5 Dynamic-data Citation Methodology***
- *TS6 Data Description Registry Interoperability Model***
- *TS7 RDA/WDS Repository Audit and Certification Catalogues***
- *TS8 RDA/WDS Publishing Data Services***

**Approved and published in the Official Journal of the European Union "COMMISSION IMPLEMENTING DECISION (EU) 2017/1358" of 20 July 2017 on the identification of ICT Technical Specifications for referencing in public procurement https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017D1358;*

***Approved and awaiting publication.*

Within the RDA Europe 4.0 project a third wave of submission has started. In September 2018, RDA Europe started sounding the interest of WGs in submitting their endorsed recommendations to the MSP. Four expressions of interest have been received by the following Working Groups: Research Data Repository Interoperability WG, Research Data Collections WG, Publishing Data Workflows WG, Wheat Data Interoperability WG.

The RDA Europe 4.0 project is awaiting from the EC the publication of the second wave of ICT Technical Specifications and the green light to start the third wave of submissions. The full process has been delayed due to the internal reorganisation of DG GROW.

1 Introduction

RDA Recommendations and Outputs[1] are the technical and social infrastructure solutions developed by RDA Working Groups or Interest Groups that enable data sharing, exchange, and interoperability. These outputs have an important impact in two areas: i) solving problems, and ii) incorporation and/or adoption in infrastructure environments by individuals, projects, and organizations. RDA's goal is to expand awareness and adoption of these outputs, and hence their impact, within all regions of the world.

RDA is not a standards development organisation (SDO), but has been building synergies with SDOs and similar organisations so that RDA Recommendations can either get a fast track to becoming standards, be incorporated in or contribute to common goals and activities, and overall foster adoption and build trust and engagement in new user communities.

In particular, RDA Europe is liaising with the European Multi-Stakeholders Platform on ICT Standardisation (MSP - E02758), an advisory expert group on ICT standardisation. Set up at the end of 2011, the European MSP on ICT standardisation advises on matters related to the implementation of ICT standardisation policies such as:

- potential future ICT standardisation needs in support of European legislation, policies and public procurement;
- technical specifications for public procurements, developed by global ICT standards-developing organisations;
- cooperation between ICT standards-setting organisations.

The Multi Stakeholder Platform is composed of representatives of national authorities from EU Member States and EFTA countries, by the European and international ICT standardisation bodies, and by stakeholder organisations that represent industry, small and medium-sized enterprises and consumers (see Figure 1). It is co-chaired by the European Commission Directorates General Internal Market, Industry, Entrepreneurship and SME and CONNECT. 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.

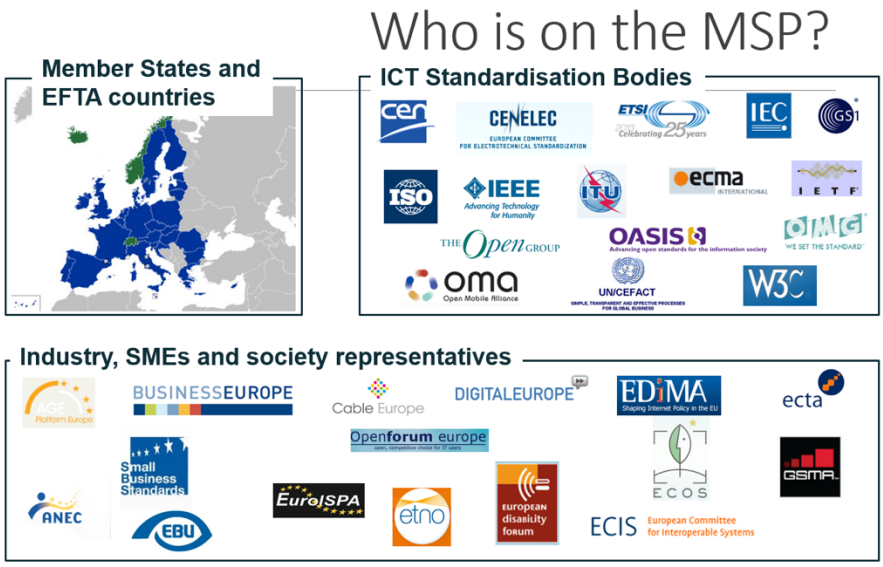

Figure 1: The MSP composition

At present, leveraging the clear and transparent RDA process for producing and endorsing its outputs

[1] https://www.rd-alliance.org/recommendations-outputs

as RDA Recommendations, two waves of outputs from RDA Working Groups have been submitted to the European Multi-Stakeholders Platform on ICT Standardisation (MSP - E02758) to be recognised as ICT Technical Specifications.

2 From RDA Recommendations to ICT Technical Specifications

In September 2015, the European Commission invited RDA to submit its first set of recommendations to be considered as ICT Technical Specifications, following a specific procedure in the EU to allow ICT specifications issued by a non-standards body to be considered in the same way as standards and referenced in public procurement (particularly important for many of EU universities and research organisations).

2.1 The identification process

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.

2.2 Specifications' requirements

ICT technical specifications to be identified as possible "common technical specifications" must comply with the requirements set out in Annex II of Regulation 1025/2012 on European Standardisation[2] and this includes having market acceptance and being coherent with European Standardisation.

Moreover, the process by which the ICT technical specifications are developed must be open, transparent and based on consensus. The specifications must also meet defined attributes, specifically: maintenance, availability, intellectual property rights licensed on a FRAND basis, relevance, neutrality and stability and quality.

2.3 General Conditions for ICT technical specification identification

2.3.1 Market acceptance

RDA aims to make research data available to all, and specifications are necessary to achieve research data provision and sharing in a discoverable, interoperable and re-usable way. Commercial markets can benefit from RDA specifications, especially when applied to the data relevant to their business areas, however there is a higher percentage of adoption cases in academia or by publicly funded organisations. This is primarily due to the nature of the Recommendation development process, as most of the contributors are public sector researchers or academics and many of the challenges are closely related to the academic context.

2.3.2 Coherence principle

ICT specifications should not conflict with European Standards.

2.4 Identification procedure

The figure below describes the submission, evaluation and decision-making process for ICT technical specifications.

[2] https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:316:0012:0033:EN:PDF

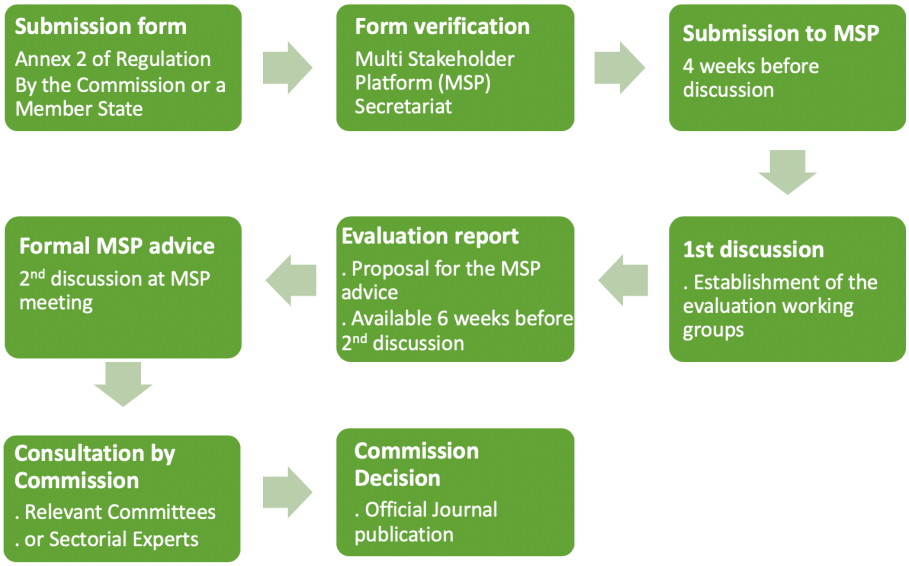

Figure 2: Identification procedure

3 The RDA ICT technical specifications

To date, the Research Data Alliance has presented eight RDA Recommendations to be evaluated and acknowledged as ICT Technical Specifications in two waves and it is currently preparing a third wave of submission within the mandate of the RDA Europe 4.0 project. The project has collected information related to the new candidate ICT technical specifications and is currently waiting the confirmation of the European Commission to start the new process.

3.1 RDA ICT technical specifications: 1st wave

The first wave of RDA recommendations was submitted in 2016 and the four Recommendations were approved as ICT Technical Specifications. On 20 July 2017, the first wave of RDA ICT Technical Specifications was published in the Official EC Journal[3].

The table below reports the first four RDA Technical Specifications.

RDA Recommendation	Description	Status
TS1 Data Foundation & Terminology Model[4]	Produced by the Data Foundation & Terminology WG which ensures researchers use a common terminology when referring to data.	Approved and published in the Official Journal of the European Union

[3] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017D1358
[4] https://rd-alliance.org/group/data-foundation-and-terminology-wg/outcomes/data-foundation-and-terminology

TS2 PID Information Types API - Persistent Identifier Type Registry[5]	Produced by the PID Information Types WG, a conceptual model for structuring typed information to better identify PIDs, common interface for access to this information.	"COMMISSION IMPLEMENTING DECISION (EU) 2017/1358" of 20 July 2017 on the identification of ICT Technical Specifications for referencing in public procurement https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017D1358
TS3 The Data Type Registries Model[6]	Published by the Data Type Registries WG providing machine-readable and researcher-accessible registries of data types that support the accurate use of data.	
TS4 Practical Policies Recommendations[7]	Produced by the Practical Policy WG designed to support data sharing and interchange between communities.	

Table 1: RDA ICT Technical specifications – 1st wave

3.2 RDA ICT technical specifications: 2nd wave

The second wave of RDA Recommendations was submitted in 2017. Four out of five recommendations have been approved during the MSP meeting on 5 December 2017 and RDA is currently awaiting publication in the EC Official Journal.

The table below reports the four approved RDA recommendations under the second wave.

RDA Recommendation	Description	Status
TS5 Data Citation of Evolving Data[8]	Supports efficient processing of data and linking from publications referencing precise subsets of changing data.	Approved and awaiting publication
TS6 Data Description Registry Interoperability Model[9]	An Interoperability model addressing the problem of cross platform discovery by connecting datasets together on the basis of co-authorship or other collaboration models such as joint funding and grants.	

[5] https://rd-alliance.org/group/pid-information-types-wg/outcomes/pid-information-types
[6] https://rd-alliance.org/group/data-type-registries-wg/outcomes/data-type-registries
[7] https://rd-alliance.org/group/practical-policy-wg/outcomes/practical-policy
[8] https://www.rd-alliance.org/group/data-citation-wg/outcomes/data- citation-recommendation.html; Technical specification: http://dx.doi.org/10.15497/RDA00016
[9] https://www.rd-alliance.org/group/data-description-registry- interoperability-ddri-wg/outcomes/data-description-registry-interoperability; Technical specification: http://dx.doi.org/10.15497/RDA00003

TS7 Core Trustworthy Data Repositories Requirements[10]	Establishes the criteria and provides technical guidance for assessment and core level certification of repositories as trustworthy. Quality performance and information management along with sustainability and trustworthiness are evaluated following 16 criteria organised along three focus areas: organisational infrastructure, digital object management and technology. Note: the CTS has subsequently become the recommended certification for FAIR data in the EOSC.[11]	
TS8 RDA/WDS Publishing Data Services[12]	An open, universal literature-data cross-linking service to improve data visibility, discoverability, re-use and reproducibility.	

Table 2: RDA ICT Technical specifications – 2nd wave

A full description of these upcoming RDA ICT Technical specifications is reported in Chapter 4.

3.3 RDA ICT technical specifications: 3rd wave

Within the RDA Europe 4.0 project a third wave of submission has started. In September 2018, RDA Europe started sounding the interest of WGs in submitting their endorsed recommendations to the MSP.

Four expressions of interest have been received by the following Working Groups:

- Research Data Repository Interoperability WG[13];
- Research Data Collections WG[14];
- Publishing Data Workflows WG[15];
- Wheat Data Interoperability WG[16].

The following table describes the four new RDA recommendations that are ready to be submitted:

RDA Recommendation	Reason for the submission, the need and intended use for the specification	Has the specification been used for different implementations by different vendors/suppliers?

[10] https://www.rd-alliance.org/group/repository-audit-and-certification- dsa%E2%80%93wds-partnership-wg/outcomes/dsa-wds-partnership; Technical specification: Catalogue of Common Requirements. https://doi.org/10.17026/dans-22n-gk35
[11] Turning FAIR into Reality. DOI: 10.2777/1524
[12] https://www.rd-alliance.org/group/rdawds-publishing-data-services- wg/outcomes/rdawds-publishing-data-services-wg-recommendations; Technical specification: http://dx.doi.org/10.15497/RDA00002
[13] https://www.rd-alliance.org/group/research-data-repository-interoperability-wg/outcomes/research-data-repository-0
[14] https://rd-alliance.org/group/research-data-collections-wg/outcomes/rda-research-data-collections-wg-recommendations
[15] https://rd-alliance.org/group/rdawds-publishing-data-workflows-wg/outcomes/rdawds-publishing-data-workflows-wg
[16] https://rd-alliance.org/group/working-and-interest-group-chairs-wheat-data-interoperability-wg/outcomes/wheat-data

Research Data Repository Interoperability Recommendations	The motivation for submitting this RDA output is to increase the visibility of an approach facilitating research data repository platform interoperability to a broader audience. With its simplicity and very low adoption effort, the presented recommendation has the potential to lower the barrier for considering research data repository platform interoperability and promises a broad adoption, either on platform level or as community effort.	The specification has been implemented in four different repository platforms as part of the early adoption process.
Data Collections API	The specification defines a generic way to interact with collections of research or other data as part of a client-server model. The specification is particularly concerned with automating access and management of data collections, independent from storage location or method, which is a need identified in multiple data communities. The specification can be used to build client applications and server systems that ease management of data collections. At the client side, it can facilitate development and adoption of management tools independent of a particular user community or provider, establishing a market for such tools. At the server side, adhering to the specification can provide unified interfaces and establish seamless workflows, which can lead to generic data collection service providers.	The specification has been successfully implemented in multiple ways and with different approaches by several early adopters from the research services community: The Helmholtz Centre Potsdam (GFZ) has prototyped it for use in seismological data management. The Perseids Project at Tufts University has used it to improve data sharing and annotation management. The Max Planck Computing and Data Facility (MPCDF) has included it in a generic data repository software. The German Climate Computing Center (DKRZ) has prototyped its use for large-scale climate data management and processing. While these are service providers for specific scientific research communities, the early adoption prototypes demonstrate that, in principle, service provisioning based on the specification is feasible and not bound to particular usage domains.

RDA/WDS Workflows for Research Data Publishing Model	Data publishing is a major cornerstone of open science, reliable research, and modern scholarly communication. It enables researchers to share their materials via dedicated workflows, services and infrastructures and ultimately is intended to ensure that data – and, in particular, datasets underlying published results — are well documented, curated, persistent, interoperable, reusable, citable, attributable, quality assured and discoverable. Data publishing workflows potentially have an enormous impact on researchers, research practices and publishing paradigms, as well as on funding strategies, and career and research evaluations. It is crucial for all stakeholders to understand the options for data publishing workflows and to be aware of emerging standards and best practices.	Data organisations engaged in data publication have adopted and are employing elements of the reference model. Adopting organizations include: GigaScience, ResearchSpace, The University of Edinburgh, DataShare, Elsevier Research Data Management Solutions, Scientific Data, the Jisc project "Giving Researchers Credit for their Data" and the Digital Curation Center (DCC).
Wheat Data Interoperability Recommendation	The International Wheat Initiative[17] (http://www.wheatinitiative.org) has identified the easy access and interoperability of all wheat related data as a top priority for the wheat research community. An important goal is to make the best use of existing genetic, genomic, and phenotypic data in fundamental and applied wheat science. Hence, data interoperability has become a priority in this community, given the ever-growing data deluge coming from improvements in technologies and numeric methods for DNA (Deoxyribonucleic acid) and RNA (Ribonucleic acid) sequencing, high throughput genotyping and phenotyping, high throughput imaging and satellite monitoring. The motivation for submitting this RDA output is to increase the visibility of an approach facilitating easy access and interoperability of all wheat related data to a broader audience.	The initial members of the RDA WG have implemented and widely promote the adoption of the data framework: the French National Institute for Agricultural Research (INRA), the Food and Agriculture Organisation of United Nations (FAO) and the International Maize and Wheat Improvement Center (CIMMYT).

Table 3: RDA ICT Technical specifications – 3rd wave

[17] http://www.wheatinitiative.org

4 Conclusions

The RDA Europe 4.0 project is awaiting from the EC the publication of the second wave of ICT Technical Specifications and the green light to start the third wave of submission. The full process has been delayed due to the internal reorganisation of DG GROW.

5 Annex 1: RDA-ready to publish ICT Technical Specifications (2nd wave)

5.1 TS5 Dynamic-data Citation Methodology

- Public working area: https://www.rd-alliance.org/group/data-citation-wg/outcomes/data-citation-recommendation.html
- Technical specification: http://dx.doi.org/10.15497/RDA00016
- Associated publication: Andreas Rauber, Ari Asmi, Dieter van Uytvanck, Stefan Pröll. Identification of Reproducible Subsets for Data Citation, Sharing and Re-Use. In: Bulletin of the IEEE Technical Committee on Digital Libraries, 12:1, 2016. http://www.ieee-tcdl.org/Bulletin/v12n1/papers/IEEE-TCDL-DC-2016_paper_1.pdf

Background and Rationale

Digitally driven research is dependent on quickly evolving technology. As a result, many existing tools and collections of data were not developed with a focus on long-term sustainability. Researchers and analysts strive for fast results and promotion of those results, but without a consistent and long-term record of the validation of their data, evaluation and verification of research experiments and business processes is not possible.

Long-term and scalable reproducible research is a requirement for interoperable research, where researchers can easily share and re-use results, including the data that they used.

There is a strong need for data identification and citation mechanisms that identify arbitrary subsets of large and possibly dynamic data sets with precision in a machine-actionable way. These mechanisms need to be user-friendly, transparent, machine-actionable, scalable and applicable to various static and dynamic data types.

Solution description

Data citation services complying to this specification enable researchers and data centres to identify and cite data used in experiments and studies. Instead of providing static data exports or textual descriptions of data subsets, the solution supports a dynamic, query centric view of data sets. The solution comprises the following core requirements:

- 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 timestamped, i.e. any additions, deletions are marked with a timestamp, ensuring a continuous data versioning service (as opposed to batch releases at arbitrary intervals).
- Data Identification: The data used shall be identified via a persistent identifier (PID) pointing to a time-stamped query, resolving to a landing page. This specification thus excludes semantic versioning such as V.X.Y.Z, that is prone to ambiguities, and requires a machine-actionable data subsetting services based on a query store.

Impact and Benefits

The overall benefit of this specification is to ensure research procurement to fund reproducible, interoperable research, fostering sustainable innovation.

Data citation services complying to this specification have several benefits compared to those using current approaches relying on individual data exports for each data set or ambiguous natural language descriptions of data set characteristics.

- It allows identifying, retrieving and citing the precise data set with minimal storage overhead by only storing the versioned data and the queries used for creating the data set. In many environments data versioning is already considered a best practice. Data subsets can be recreated on demand without the need to understand and negotiate with local versioning / subsetting systems. This is of particular interest now that cross-disciplinary and cross-border research studies require to extract data from distributed data centers.
- It allows retrieving the data both as it existed at a given point in time as well as the current view on it, by re-executing the same query with the stored or current timestamp, thus benefiting from all corrections and updates made since the query was originally issued. This allows tracing changes of data sets over the time and comparing the effects on the result set, paving the way for continuous monitoring of analyses outcomes.
- The query stored as a basis for identifying the data set provides valuable and precise provenance information on the way the specific data set was constructed, thus being semantically more explicit than a mere data export.
- The query store offers a valuable, central basis for analysing data usage.
- Metadata such as checksums support the verification of correctness and authenticity of data sets retrieved.
- The specifications are applicable across different types of data representation (e.g. relational data (RDBMS/SQL), XML, Linked Data / LOD, Column Stores, CSV, files/filesystems) and data characteristics (big or small data; static or highly dynamic; identifying single values or the entire data set).
- If data is migrated to new representations, the queries can also be migrated, ensuring stability across changing technologies.
- Distributed data sources can rely on local timestamps at each node, avoiding the need for expensive synchronization in loosely coupled systems.
- The recommendations are neutral to the specific persistent identifier system used (DOI, ARK, URI ...)
- The recommendations are in-line with the Data Citation Principles[18], specifically Principle 7 (Specificity and Verifyability), but also 2 (Credit and Attribution), 4 (Unique Identification), 5 (Access) and 6 (Persistence).

5.2 TS6 RDA Data Description Registry Interoperability Model

- Public working area: https://www.rd-alliance.org/group/data-description-registry-interoperability-ddri-wg/outcomes/data-description-registry-interoperability
- Technical specification: http://dx.doi.org/10.15497/RDA00003

Background and Rationale

Driven by the rapid development of data storage technology, the number of research data repositories is growing fast and researchers more than ever have access to a range of data repositories including university data storage, discipline specific repositories and national (regional) level data infrastructures. The problem is that these infrastructures are often operating in silos; that is, they cannot connect their datasets to the related research or datasets in other platforms.

[18] https://www.force11.org/datacitationprinciples

Solution description

The model provides researchers with a mechanism to connect datasets in various data repositories based on collaboration network, co-authorship, joint funding, and grants.

Impact and Benefits

- Researchers have difficulties finding related works. They will be able to find similar datasets connected by co-authorship, and collaboration in research projects and grants.
- Registries/repositories need to disambiguate and connect datasets. They will be able to find connected datasets across multiple platforms.
- Universities will be able to find datasets produced by their researchers.

5.3 TS7 RDA CoreTrustSeal (CTS) (2016)

- Public working area: https://www.rd-alliance.org/group/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg/outcomes/dsa-wds-partnership
- Technical specification: Catalogue of Common Requirements. https://doi.org/10.17026/dans-22n-gk35

Background and Rationale

To be sharable, data needs to be stored in a trustworthy digital repository. Data that is collected, created and used should be managed, curated, and archived in such a way to preserve the initial investment. Organisations depend on data repositories to ensure their data assets remain usable, useful and meaningful into the future. Certification against the CoreTrustSeal requirements as a trustworthy digital repository provides assurance that data are managed, curated and archived in such a way as to preserve and enhance the initial investment.

CoreTrustSeal represents the harmonization of the requirements and procedures, previously governed by the Data Seal of Approval (DSA) and the ICSU World Data System (ICSU-WDS), accomplished by the Research Data Alliance (RDA) Working Group on 'Repository Audit and certification'. The CoreTrustSeal board administers the process and certification of repositories, including the transition of 120 existing certified DSA and/or WDS repositories to the aligned requirements.

CoreTrustSeal website: https://www.coretrustseal.org/

Solution description

The CoreTrustSeal Data Repositories Requirements are used to certify a repository in the OAIS11 sense (Reference model for an Open archival Information System ISO14721) as a Trustworthy Digital Repository (TDR). TDR status encompasses strict expectations acrossinfrastructure, digital object management, policies and technology.

The CoreTrustSeal Infrastructure requirements deal with a repository's mission/scope, licences and continuity of access, but also handle confidentiality/ethics, the organisational structure itself and expert guidance. The requirements on digital object management cover data integrity and authenticity as well as appraisal, documented storage procedures, preservation planning, data quality and workflows. Next to that data discovery and identification are included as well as data reuse. The requirements for technology involve the technical infrastructure and security of a digital repository.

The detailed nature of supporting evidence will vary depending on the mission and scope of the applicant. But aside from this context, the criteria for defining quality and performance are consistently approached across all organisations seeking to be recognised as trustworthy stewards of data with a remit for ensuring long term access.

By becoming certified, repositories can demonstrate to their depositors, users and funders that an independent authority has evaluated them and endorsed their trustworthiness.

The CoreTrustSeal Board maintains the Requirements and Procedures and offers a service based on self-assessment and peer-review followed by board approval and publicly available certification and evidence.

Impact and Benefits

The impact and benefits of CoreTrustSeal certification apply across all domains where data are collected, preserved and made available for re-use, whether scientific, commercial or governmental. These specifications will provide a step towards having more coherent, increasingly stringent and compatible standards for repository certification. The global community of data depositors, repositories and users benefit from internationally recognized requirements to demonstrate Trustworthy Data Repository status.

CoreTrustSeal Repository certification:

- Creates a critical mass of certified repositories across a range of domains and disciplines.
- Increases stakeholder confidence, as it signifies to funders that the data they have invested in will continue to be available for reuse. Data producers can be confident that the data they have worked hard to create will be protected, and data consumers can be sure that the data they are using have been managed optimally.
- Provides a mechanism to communicate and raise awareness around digital preservation. In an era of mass data creation and instant communication, people often focus on access to digital resources but not the importance or the challenges of preserving data for future reuse.
- Forms the core level to the European Framework for Digital Certification of trustworthy digital repositories in the OAIS sense (Reference model for an Open archival Information System ISO14721). Additional tiers are provided by nestor's DIN3164412: Information and Documentation – Criteria for Trustworthy Digital Archives and ISO1636313: Audit and Certification of Trustworthy Digital Repositories.
- Is now the recommended certification for FAIR data in the EOSC.[19]
- Facilitates the provision of cross-border services. CoreTrustSeal is already adopted by several infrastructures and international bodies e.g.:
    - Within the Common LAnguage Resources and Technology Infrastructure (CLARIN) it is used as one of the requirements to become a CLARIN-B centre.
    - CESSDA Consortium of European Social Science Data Archives have incorporated the CoreTrustSeal in their organisational structure to warrant the quality of member the archives.
    - Within the ICSU World Data System (ICSU-WDS) it is used as part of its membership application procedure.

In addition to demonstrating a commitment to ensuring data remain useful for future generations, compliance with the CoreTrustSeal provides the following benefits to repositories:

- Recognition as a trustworthy source for data.
- Improvements in communication: Preparing for the self-assessment prompts a repository to communicate internally about their overall mission and goals in ways not always present in day-to-day interactions.
- Improvement in processes: Conducting the self-assessment stimulates a repository to improve its processes and procedures and move to a higher level of professionalism, with an incentive to improve its operations over time.
- Transparency: The core certification process is designed to provide a public statement of repository evidence enabling stakeholders to evaluate the repository's operations and policies.
- Differentiation from others: There are a growing number of options for depositing data. Certification enhances the repositories reputation, demonstrating in an easily recognized way that the repository is following best practice.

[19] Turning FAIR into Reality. DOI: 10.2777/1524

- Integrates rather than conflicts with other standards. CoreTrustSeal can be successfully applied alongside related or complementary standards including DIN31644, ISO16363 or ISO2700114 for information security.

5.4 TS8 RDA RDA/WDS Publishing Data Services

- Public working area: https://www.rd-alliance.org/group/rdawds-publishing-data-services-wg/outcomes/rdawds-publishing-data-services-wg-recommendations
- Technical specification: http://dx.doi.org/10.15497/RDA00002

Building on pre-existing components and international initiatives, the RDA/WDS Publishing Data Service focused on building a one-for-all cross-referencing service for the links between data and publications. A follow-on RDA working group was then created to implement and expand the specification further: The Scholarly Link Exchange Working Group focused on the enabling infrastructure for a global view of data - literature links and the supporting systems for that.

Background and Rationale

Linking Research Data with the Literature is of great value, yet current solutions are not realizing the potential. Problems are:

- Many disconnected sources (publishers, data centers, repositories, infrastructure providers, ...)
- Heterogeneity of practices, for example:
    - Different PID systems (DOI, accession numbers)
    - Different ways of referencing data (formal citations, in-text references, ...)
    - Different moments of citing data (at publication, post publication, ...)
    - There is a need for a shift from a plethora of (mostly) bilateral arrangements to a one-for-all service model infrastructure for the research data publication landscape

Solution description

A one-for-all cross-referencing service for the links between data and publications comprised of:

- (i) a recommendation for a long-term approach to sharing information about the links between the literature and research data (Scholix Framework), and
- (ii) a prototype implementation of a data literature interlinking system that is expected to develop into a key element of the proposed long-term model (DLI system).

Impact and Benefits

For data repositories and journal publishers: linking data and the literature will increase their visibility and usage, and can support additional services to improve the user experience on online platforms (for example, offering links to relevant data sets with articles, or offering links to the literature that will help place data in context). In contrast to the bilateral arrangements that often exist between data centers and journal publishers, the proposed service will make the process of linking data sets and research literature a more robust, comprehensive, and scalable enterprise.

For research institutes, bibliographic service providers, and funding bodies: the service will enable advanced bibliographic services and productivity assessment tools that track datasets and journal publications within a common and comprehensive framework.

For researchers: firstly, the service will make the processes of finding and accessing relevant articles and data sets easier and more effective. Secondly it will make it possible for researchers to track long-term impact of their data (and publications), thereby providing additional incentives to share data.