



## RESEARCH DATA ALLIANCE EUROPE

Project Acronym	RDA Europe
Project Title	Research Data Alliance Europe
Project Full Title	RDA Europe – the European plug-in to the global Research Data Alliance
Project Number	632756
Deliverable Title	Output Publishing Strategy
Deliverable No.	D3.2
Delivery Date	02/15
Authors	Fotis Karayannis (ATHENA), Timea Biro (TRUST-IT), Herman Stehouwer (MPG)

### ABSTRACT

D3.2 Output Publishing Strategy provides an overview of the activities performed to identify appropriate models and technologies to support the publishing and thus re-use of the RDA outputs across different channels including Open Access repositories.

# DOCUMENT INFORMATION

PROJECT		
<b>Project Acronym</b>	RDA Europe	
<b>Project Title</b>	Research Data Alliance Europe	
<b>Project Full Title</b>	RDA Europe – the European plug-in to the global Research Data Alliance (RDA)	
<b>Project Start</b>	1st October 2014	
<b>Project Duration</b>	18 months	
<b>Funding</b>	FP7-INFRASTRUCTURES-2013-2	
<b>Grant Agreement No.</b>	632756	
DOCUMENT		
<b>Deliverable No.</b>	D3.2	
<b>Deliverable Title</b>	Output publishing strategy	
<b>Contractual Delivery Date</b>	02/15	
<b>Actual Delivery Date</b>	02/15	
<b>Author(s)</b>	Fotis Karayannis (ATHENA), Timea Biro (TRUST-IT), Herman Stehouwer (MPG)	
<b>Editor(s)</b>	Fotis Karayannis (ATHENA), Timea Biro (TRUST-IT), Herman Stehouwer (MPG)	
<b>Reviewer(s)</b>	Patrice Ajai-Ajagbe (ACU), Juan Bicarregui (STFC)	
<b>Contributor(s)</b>	Fotis Karayannis (ATHENA), Timea Biro (TRUST-IT), Herman Stehouwer (MPG)	
<b>Work Package No. &amp; Title</b>	WP3 RDA Governance, Communications and Outputs	
<b>Work Package Leader</b>	Hilary Hanahoe, TRUST-IT	
<b>Work Package Participants</b>	ATHENA RC, CNR-ISTI, MPG, TRUST-IT	
<b>Estimated Person Months</b>	5	
<b>Distribution</b>	Public	
<b>Nature</b>	Report	
<b>Version / Revision</b>	2.0	
<b>Draft / Final</b>	Final	
<b>Total No. Pages (including cover)</b>	20	
<b>Keywords</b>	RDA Outputs, Repository, Open Access, Adoption	

# DISCLAIMER

---



Communications Networks, Content and Technology  
European Commission Directorate General

## DG CONNECT

---

RDA Europe (632756) is a Research Infrastructures Coordination and Support Action (CSA) co-funded by the European Commission under the Capacities Programme, Framework Programme Seven (FP7).

This document contains information on RDA Europe (*Research Data Alliance Europe*) core activities, findings and outcomes and it may also contain contributions from distinguished experts who contribute as RDA Europe Forum members. 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 its experts, 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 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 27 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.int/>).

Copyright © The RDA Europe Consortium 2014.

For more information on the project, its partners and contributors please see <https://europe.rd-alliance.org/>. You are permitted to copy and distribute verbatim copies of this document containing this copyright notice, but modifying this document is not allowed. You are permitted to copy this document in whole or in part into other documents if you attach the following reference to the copied elements: "Copyright © The RDA Europe Consortium 2014."

The information contained in this document represents the views of the RDA Europe Consortium as of the date they are published. The RDA Europe Consortium does not guarantee that any information contained herein is error-free, or up to date. THE RDA Europe CONSORTIUM MAKES NO WARRANTIES, EXPRESS, IMPLIED, OR STATUTORY, BY PUBLISHING THIS DOCUMENT.

# GLOSSARY

ABBREVIATION	DEFINITION
<b>AAI</b>	Authentication and Authorisation Infrastructure
<b>ATHENA (RC)</b>	Athena Research and Innovation Centre in Information Communication & Knowledge Technologies
<b>B2*</b>	EUDAT suite of research data management services
<b>B2SHARE</b>	EUDAT solution to store and share research data
<b>CERN</b>	European Organization for Nuclear Research
<b>CKAN</b>	An Open Source data portal platform
<b>CSC</b>	Finnish IT Centre for Science
<b>DSPACE</b>	An Open Source repository software package typically used for creating open access repositories for scholarly and/or published digital content.
<b>DOI</b>	Digital Object Identifier; is a character string (a "digital identifier") used to uniquely identify a digital object, such as an electronic document
<b>DoW</b>	Description of Work (the Technical Annex of the RDA Europe Grant Agreement)
<b>EUDAT</b>	An EC-funded project aiming at building a set of data management services (the B2* services) for data preservation, search, access, and processing
<b>FIGSHARE</b>	A repository where users can make their research outputs available in a citable, shareable and discoverable manner
<b>GIT</b>	A free and open source distributed version control system
<b>IETF</b>	Internet Engineering Task Force; the standards organisation for the Internet
<b>IG</b>	(RDA) Interest Group
<b>ISLANDORA</b>	An open-source software framework designed to help organizations and their audiences collaboratively manage, and discover digital assets
<b>HANDLE</b>	The Handle System <sup>®</sup> is a component of the <a href="#">Digital Object Architecture</a> and provides a resolution services for unique and persistent identifiers

	of digital objects.
<b>MPG</b>	Max Planck Gesellschaft
<b>ResearchGate</b>	A system connecting researchers and make it easy for them to share and access scientific output, knowledge, and expertise.
<b>OPENAIRE</b>	An EC-funded project aiming at promoting open scholarship and improve the discoverability and reusability of research publications and data.
<b>OPENAIRE/Zenodo</b>	Builds and operates a service that enables researchers, scientists, EU projects and institutions to share and showcase multidisciplinary research results (data and publications).
<b>RDA Europe</b>	Research Data Alliance Europe
<b>RDA OAB</b>	Organisational Advisory Board; one of the governing bodies of RDA composed of the RDA Organisational Members with a particular role in the outputs adoption.
<b>RDA Secretariat</b>	The RDA Secretariat, led by the RDA Secretary General, is responsible for the supporting the day to day business of all the RDA processes.
<b>RDA TAB</b>	Technical Advisory Board; one of the governing bodies of RDA providing technical advice
<b>URL</b>	A Uniform Resource Locator; is a reference to a resource that specifies the location of the resource on a computer network and a mechanism to retrieve it
<b>W3C</b>	World Wide Web Consortium; a standards organisation for the Web
<b>WG</b>	(RDA) Working Group

# TABLE OF CONTENTS

---

Executive Summary .....	7
1. Introduction .....	9
1.1 Target audiences .....	9
1.2 Structure of the document .....	9
2. Technologies and tools for publishing outputs – Selection Process .....	11
2.1. Introduction .....	11
2.2. Technological requirements for the Outputs system .....	11
2.3. Identified solutions.....	13
2.4. Group decision.....	16
2.5. Current status and first feedback .....	16
3. Outputs publishing strategy .....	16
3.1 Introduction .....	16
3.2. Outputs publishing strategy .....	17
4. Conclusions and actions .....	20

# Executive Summary

---

The Research Data Alliance (RDA) builds the social and technical bridges that enable open sharing of data across technologies, disciplines, and countries. RDA Europe (the European branch of RDA), leads discussions on global solutions and ensures that the scientific excellence generated in Europe around research data, remains prominent. The RDA Europe project catalyses global and European policy, scientific and industrial engagement, showcases scientific excellence and innovation, and facilitates the publication of RDA outputs.

This document summarises the work of the RDA Europe project and the RDA Outputs Repository Task Force in developing an overall RDA Outputs publishing strategy. This includes the process for selecting the Outputs Repository where all RDA outputs will ideally be deposited and the corresponding requirements and selection criteria.

## The Outputs Publishing Strategy

RDA outputs are expected to be a variety of different things in a variety of different formats, ranging from specifications, policies, best practices and recommendations, to frameworks, pieces of software, or prototype solutions or tools. Thus any outputs publishing strategy needs to take into account this variety of outputs, and adopt different approaches for different audiences. The main points of the Outputs publishing strategy were developed based on this principle, and are summarised as follows:

- The Outputs publishing strategy will be **tailored to different audiences as required** i.e. the outputs will be made available to each audience in way that is relevant to that particular audience.
- Considering the RDA's own organisational future is still being developed, the Outputs publishing strategy will be **responsive and adapt to changes** as necessary.
- The Outputs publishing strategy will make clear that it will take some time to develop a comprehensive set of outputs and that some outputs may be in a primitive state. Thus the strategy **will have to be realistic and modest, not overselling the outputs**, to avoid disillusioning potential early adopters.
- The Outputs Publishing Strategy will be **aligned with the overall RDA Strategic Planning exercise** that was initiated in RDA Plenary 5 in San Diego.
- The solution(s) and in particular the Outputs repository will be **Open Access following the RDA Outputs Policy**, approved by the RDA Council.

## The Outputs Repository

RDA Outputs Repository Task Force (ORTF) – established and managed by members of the RDA secretariat - focuses on the selection of relevant repositories for RDA Outputs. In selecting a definitive repository solution, the ORTF considered the following 9 options: CKAN, DSpace, EUDAT/B2SHARE, Figshare, GIT, Islandora, OpenAIRE/Zenodo, the RDA File Repository Content Type, and ResearchGate. These options were categorised into the following 4 different types:

- 'Developed in house'
- 'Requires open source software' (that would need to be installed, ran and maintained in the RDA servers)
- 'External/commercial/social services'
- 'Consolidated solutions' (that would be externally ran and managed).

The ORTF thought that the “Consolidated solutions” (“EUDAT/B2SHARE” and “OpenAIRE-Zenodo” respectively) were overall the best options. EUDAT/B2SHARE was better marked in terms of versioning and Drupal integration. OpenAIRE/Zenodo was more advanced in terms of deployment time (including the Invenio version) and features such as the community support and better data linking capabilities. The involved team debated on opposing opinions and perspectives between the two. Further considerations were made, including non-technical ones, in order to be able to make a decision. EUDAT is an RDA Organisational member and there is a strong participation of EUDAT members to the RDA groups, as well as a firm commitment of EUDAT to be an RDA output adopter. EUDAT also offers a bundle of services, the repository solution can become part of an integrated package of services that includes a metadata portal B2FIND, data staging B2STAGE, safe replication through B2SAFE, easy data synchronisation and sharing through B2DROP & upcoming workflow related services and integration of AAI capabilities. At the end these points prevailed and it was decided to adopt the EUDAT B2SHARE offer.

Testing of the EUDAT B2SHARE solution has already begun (started in early 2015) and the initial feedback is positive, although some technical issues are still being worked on. The next evaluation of the **EUDAT B2SHARE** solution will be conducted after the 6<sup>th</sup> RDA plenary (in September 2015) - where more outputs, from even more audiences should be available. (Further details on the selection process can be found in section 2 of this document)

A single repository or platform, may not serve optimally the wide variety of RDA outputs and audiences. However, the RDA-E project group and the ORTF recommends that **EUDAT B2SHARE** (the provisional recommended outputs repository solution) is tested by a variety of audiences (to help inform the strategic process), and that these audiences are kept engaged.

In the second period of the project, it will be investigated whether the owners of outputs are interested in publishing them in open journals. In parallel RDA Europe undertakes to publish attractive printed material (such as the Outputs booklet<sup>1</sup>) to disseminate its first outputs to interested audiences.

This deliverable (developing the overall RDA Outputs publishing strategy) has a European flavour, as the European RDA community played a key role in developing the strategy and repository solution, but it is global in scope. It is aimed at any individuals and organisations who may be interested in engaging and working with the RDA - in particular, data practitioners and individuals with an interest in the outputs publishing strategy and tools.

---

<sup>1</sup> [https://rd-alliance.org/sites/default/files/attachment/RDA\\_Outputs\\_May2015\\_web.pdf](https://rd-alliance.org/sites/default/files/attachment/RDA_Outputs_May2015_web.pdf)



# 1 Introduction

## 1.1 Scope of the document

RDA Europe acts as the European branch of RDA, ensuring that the scientific excellence generated in Europe around research data is prominent and leading discussions on global solutions. Thus RDA Europe contributes to the overall RDA aim, which is to build social and technical bridges that enable open sharing of data across technologies, disciplines, and countries. In particular, RDA Europe project catalyses global and European policy, scientific and industrial engagement, showcases scientific excellence and innovation and facilitates publishing its outputs. The latter objective is also the main focus of this deliverable, i.e. to create a formal framework and identify suitable channels so as to publish RDA Working group outputs. The deliverable reflects in part the work of Task T3.1 of the RDA Europe project, however as this topic is an RDA “global” one of particular importance, there have been related efforts undertaken mainly by a corresponding Task Force, entitled the “RDA Outputs Repository Task Force” set up and managed by members of the RDA secretariat. The Task Force focused on the selection of the corresponding repository for the RDA Outputs. The Task Force initially involved the RDA Secretary General, Mark Parsons, along with Herman Stehouwer (MPG) and Timea Biro (Trust-IT), whose work started just before summer 2014 as part of the RDA secretariat and after the start of RDA Europe project in October 2014, other members joined further on in the process, namely Fotis Karayannis from ATHENA and Yolanda Meleco from RDA US. This document summarises the overall publishing strategy, along with the process for selecting the Outputs Repository including the corresponding requirements and criteria for selection.

## 1.2 Target audiences

This deliverable has mainly a global scope, identifying the repository and related models and technologies that support the publishing and thus the re-use of the RDA WG/IG outputs. However, it also has a European flavour, as the European RDA community has played a key role in the related strategy along with the selection of the repository solution. The target audience are both internal to RDA, i.e. RDA members, and in particular the European ones for an RDA Europe deliverable, but also external ones, who may be interested in engaging and working with RDA and its WG/IGs. In this way, a better understanding of its outputs strategy and related technologies can be achieved. Yet, as it is a public document, the deliverable has a much broader scope, especially for all possible data practitioners and users with an interest in the outputs publishing strategy and tools.

## 1.3 Structure of the document

The document has five (5) main sections and two (2) appendices:

- Section 1 is the Executive Summary
- Section 2 provides an introduction with overall scope of the document, its target audience and its structure
- Section 3 analyses technologies and tools for storing and publishing outputs, along with the overall process for selecting the corresponding tools and in particular the repository.
- Section 4 proposes a simple and adaptive outputs publishing and dissemination strategy, taking into account the analysis of the technologies and models of section 3
- Section 5 provides some conclusions and next steps.



info@europe.rd-alliance.org | [europe.rd-alliance.org](http://europe.rd-alliance.org)



## 2 Technologies and tools for publishing outputs – Selection Process

### 2.1. Introduction

RDA outputs are expected to not only be specifications, policies, best practices and recommendations, but also frameworks, implemented code, tools or prototypes. However, we should note that RDA itself does not maintain a codebase. Thus, publishing tools are needed based on new technologies and models that allow linking to data or software. Modern publishing allows us to link the specification output of RDA WGs with the textual descriptions and other metadata, the implementations of tools, services, policies, and thus, enable the interested expert to immediately take action to increase data sharing. Efficiency of the procedures is nowadays essential for take-up, since in all domains there is only little time to adopt new ways: if it works immediately people will use it, if not people will ignore it. This section summarises the technologies and models reviewed aiming to support the preservation, the publishing and thus the re-use of the RDA outcomes across different channels including Open Access repositories. The overall process for selecting the corresponding repository, along with appropriate integrated technologies and tools is summarised.

### 2.2. Technological requirements for the Outputs system

As explained in the introduction, the work for the Outputs system specification started already before summer 2014, as part of the RDA secretariat work. The agreed starting point was that the website custom repository present in the group working areas will be used for “discussion documents”, yet another, more formal and controlled system will be used for the Outputs. Integration with the Drupal-based website system would be desirable but not mandatory. Although there were frequent discussions about the RDA outputs system in the secretariat calls, the related Task Force composed of Timea Biro (Trust-IT), Mark Parsons (RDA Secretary General) and Herman Stehouwer (MPG) was formed during summer and the first conference call on the Outputs system repository was organised on the 19<sup>th</sup> of August 2014. After the start of the RDA Europe project in October 2014, other members joined the Task Force namely Fotis Karayannis from ATHENA/RDA Europe and Yolanda Meleco, RPI/ RDA US.

#### 2.2.1. Requirements

The main identified requirements for the Output system included:

- *Open access, but controlled uploading and deleting* of documents
- *Rigorous version control* of the uploaded documents.
- *A simple directory listing* of outputs
- *Descriptive standard metadata*
- *DOIs* assigned to final outputs that are out for review
- *A URL* for the DOI. This is a landing page that points to a collection of files. The landing page could be done in the Drupal-based RDA website.
- *A documented versioning scheme* that describes how documents move from a draft (no DOI) to final version. Each new final version gets a DOI. Drafts do not.
  - Note: if it is out for public comment it is a final version deserving a DOI. The process also needs a way to deprecate old versions. Perhaps the versioning history could be described on a landing page.
- *Basic navigation of earlier versions* (possibly through the landing page)

- A documented process for registering and updating a document. This can be manual initially and automated later on. This relates closely to the versioning process.
- A defined list of acceptable file types. This will need to evolve. In the starting phase, no proprietary formats may be allowed and machine readable formats (or at least UTF-based as opposed to a PDF) are preferred.

## 2.2.2. Evaluation criteria table

The evaluation table used, building on top of the previously selected requirements is summarised below.

**Table 1: Evaluation Criteria Table**

Requirements	Notes/ details on requirement
Open access but controlled upload/delete	Open access but controlled uploading and deleting of documents. With simple file upload.
Simple directory listing	Search & Browse
Discoverability	Allowing content to be indexed, searched and retrieved - both human & machine readable interface
File format flexibility	All file types should be accepted. The outputs policy should eventually include a defined list of acceptable file types. This will need to evolve. In the starting phase, no proprietary formats may be allowed and machine readable formats (or at least UTF-based as opposed to a PDF) are preferred.
Metadata	Descriptive standard metadata - Evaluate the input & view functions - title, creator, subject, date etc.
Versioning	Rigorous version control of the uploaded documents.
Versioning history	Documented versioning with basic navigation of earlier versions .The process also needs a way to deprecate old versions
DOIs assignment	DOIs assigned to final outputs. It can be done either through an integrated repository service or through a partnership with a DOI assigner (considering DataCite). Data citation analytics also to be considered.
Archiving/Long term preservation	How is the long term preservation addressed?
Drupal / Drupal integration	Drupal or Drupal plug-in for quick access & user permission control
Open Source/Cost	Is the solution an open source one if not what are the costs involved
Solution maturity & user/ support community	Project should be a mature one, with a consistent user database and community of developers
Implementation	Evaluation of the ease of implementation and maintenance. Storage & hosting should also be taken under consideration.

Table 1 – Evaluation Criteria Table

### 2.2.3. Evaluation scoring

A scale was used to evaluate how the various solutions answered the requirements:

Evaluation

0-3 scale:

0= requirement not met at all

1= requirement met only to some extent

2= requirement met to satisfactory extent

3= requirement met to full extent

### 2.3. Identified solutions

A total of 9 solutions were considered – these were: GIT; File Repository Content Type; EUDAT/B2SHARE; Islandora; DSpace; CKAN; OpenAIRE/Zenodo; ResearchGate; and Figshare. These options, (also shown in table 2) were grouped into the following 4 types:

- Developed in-house (using Drupal content items)
- Open source software (that needs to be installed, ran & maintained in the RDA server)
- External/commercial/social service
- Consolidated solutions<sup>2</sup> (externally ran and managed)

**Table 2: The Solutions Considered - Grouped by Type**

Solutions Considered	Developed in-house using Drupal content items	Open source software (that would need to be installed, run & maintained in the RDA server)	External / commercial/social service	Consolidated solutions (externally ran and managed)
<ul style="list-style-type: none"> <li>• CKAN <a href="http://ckan.org/">http://ckan.org/</a></li> </ul>		✓		
<ul style="list-style-type: none"> <li>• DSpace <a href="http://www.dspace.org/">http://www.dspace.org/</a></li> </ul>		✓		
<ul style="list-style-type: none"> <li>• EUDAT/B2SHARE <a href="http://www.eudat.eu/b2share">http://www.eudat.eu/b2share</a></li> </ul>				✓

<sup>2</sup> A consolidated solution is a solution that has been tested and is in use, combining a number of features that respond to user requirements previously analysed, features that have been developed in answer to specific requests from communities similar to ours. A consolidated solution would save installation, configuration and testing time and allow to leverage on the previous both technical & non-technical experience, as well as a set of successful or failed tests and developments.

<ul style="list-style-type: none"> <li>• Figshare <a href="http://figshare.com/">http://figshare.com/</a></li> </ul>			✓	
<ul style="list-style-type: none"> <li>• File Repository Content Type <a href="https://rd-alliance.org/node">https://rd-alliance.org/node</a></li> </ul>	✓			
<ul style="list-style-type: none"> <li>• GIT <a href="http://git-scm.com/">http://git-scm.com/</a></li> </ul>				
<ul style="list-style-type: none"> <li>• Islandora <a href="http://islandora.ca/">http://islandora.ca/</a></li> </ul>		✓		
<ul style="list-style-type: none"> <li>• OpenAIRE/Zenodo <a href="http://zenodo.org/">http://zenodo.org/</a></li> </ul>				✓
<ul style="list-style-type: none"> <li>• ResearchGate <a href="http://www.researchgate.net/">http://www.researchgate.net/</a></li> </ul>			✓	

### 2.3.1. Analysis

- 1) The in-house solution **File Repository Content Type** is relatively easy to implement as an area to store documents, offers versioning, DOIs can be assigned, it is open access and there is flexibility in management. However, it is not a consolidated solution; there is no long term sustainability plan in place and it requires long term maintenance effort.
- 2) Although they answer more **specifically to requirements of scientific repositories**, the open source software solutions like **CKAN, DSpace, Islandora** are very similar to the in-house solution, as it will still be a first-time experience, requiring extended testing, bug fixing and updates, as well as long term support from the technical team managing the rd-alliance.org website. The need for a long term maintenance and preservation plan would also have to be considered in this case. In the case of the RDA website and its content, the technical team has extended experience in the set-up, development and maintenance of such web platforms; however, in the case of a repository the situation is different as it would be a first time experience. A consolidated solution ran externally is thus advisable, instead of an out-of-the-box solution like CKAN, DSpace or Islandora, that will require installing, configuring and testing effort and time. Furthermore a different (from the website content) management procedure is required in the case of the outputs for a number of reasons: outputs will not be updated once published and DOIs assigned (an update will equal a new version, new iteration) and the requirements related to the DOI assignment are binding. In the case of the website content, on the other hand, it is important that content managers are able to access & make frequent changes in order for it to be constantly up-to-date. The priorities in this case are different: long term preservation & DOI contractual obligations vs. full autonomy & control.
- 3) In the case of these 2 types of solutions (in-house & open source repository software) **the benefits of choosing a consolidated solution prevail on the benefits of autonomy & customizing flexibility.**

The Figshare/ ResearchGate solutions come with a consolidated community **but complete external dependency is not advisable**. Solutions like EUDAT and Zenodo provide a middle-way as they allow customisation of the upload interfaces and of the display layout & rendering. An external solution like Figshare / ResearchGate will not provide an API and the uploads will be presented like in pre-set views that may or may not be a perfect fit for the outputs rendering.

- 4) Choosing solutions like **EUDAT and Zenodo** would allow to **build on their technical as well as non-technical experience coupled with the benefit of opening access to a consolidated community that RDA targets**. Solutions like EUDAT or Zenodo would allow us to take advantage of previous developments and tests done by the technical teams behind these solutions, experience that can include technical requirements (defining metadata, APIs) as well as non-technical issues (related to the behaviour patterns of the potential adopters). Furthermore both EUDAT and Zenodo already have a user base and communities that they cater to, that can become adopters of RDA, if not already (a number of the members of the first RDA WGs providing outputs are also working on initiatives like EUDAT).

Having a consolidated option prevails over requirements like full Drupal integration, complete customizing flexibility and the risks related to outsourcing the technical aspects of the service.

EUDAT and Zenodo are also the advised choice as publicly funded solutions – showcasing for the re-use of outputs of public funds & sustainability. They build on previously funded initiatives and strengthening synergies with such initiatives is important for organisations like RDA that is itself publicly funded. A further motivation is the low or lack of cost for RDA in comparison to commercial solutions. EUDAT & Zenodo are both based on Invenio open source software.

### **Two critical issues were pointed out:**

- **long term sustainability beyond project funding**
- **versioning – no native versioning system in place for either**

Both projects, EUDAT and OpenAIRE have had a similar evolution and the 2 services have been launched at close interval from one another, with Zenodo launched first in May 2013 and B2SHARE at the end of 2013.

Sustainability beyond the supporting project's lifetime (OpenAIRE or EUDAT) is in both cases related to the commitment of large research institutions like CERN (with similar if not the same storage conditions as data from LHC) for Zenodo<sup>3</sup> and CSC, MPG, Juelich for EUDAT<sup>4</sup>.

At this point in the assessment, the 2 solutions have similar technical evaluation. EUDAT/B2SHARE is better marked in terms of versioning and Drupal integration. OpenAIRE/Zenodo is more advanced in terms of deployment time and features (including the Invenio version) such as the community feature and supports better data linking capabilities (that were not present in the table).

### **Further considerations were taken into account:**

---

<sup>3</sup> <http://zenodo.org/faq> - In the highly unlikely event that ZENODO will have to close operations, we guarantee that we will migrate all content to other suitable repositories, and since all uploads have DOIs all citations and links to the dataset will not be affected.

<sup>4</sup> <http://www.eudat.eu/b2share-faq-generic> - EUDAT: In the unlikely event that the B2SHARE service would draw to a close in the future, we guarantee to keep your data archived and accessible for at least 2 years and will assist in the migration of your data to other suitable repositories.

Further considerations were made, including non-technical ones in order to be able to make a decision. EUDAT offers a bundle of services, the repository solution can become part of an integrated package of services that includes a metadata portal B2FIND, data staging B2STAGE, safe replication through B2SAFE, easy data synchronisation and sharing through B2DROP & upcoming workflow related services and integration of AAI capabilities. EUDAT is also an RDA Organisational member and there is a strong participation of EUDAT members to the RDA groups, as well as a firm commitment of EUDAT to be an RDA output adopter.

## 2.4. Group decision

**In November 2014, it was decided to follow up on the EUDAT B2SHARE offer**, provided the issues of long term sustainability beyond the project funding and the versioning are addressed in a satisfactory manner. It was then confirmed with EUDAT service managers that sustainability and versioning were addressed in a satisfactory manner, providing further details on the EUDAT offer.

## 2.5. Current status and first feedback

Given the fact that a third party solution was chosen, there would be no actual development phase, rather a “testing phase” and then a “use phase”. The initial testing of the EUDAT solution started in 2015. However, given the pressing needs of the March RDA plenary in the US that was given full priority, it was only April 2015 where the use phase started. The initial feedback is positive, with some issues faced being discussed (e.g. the metadata of the outputs are not editable after being submitted), which are expected to be fixed soon. By May 2015, all important RDA policy documents along with the first WG outputs have been deposited to the EUDAT B2SHARE repository.

A next evaluation will be performed after the RDA plenary 6, so as to have more outputs from different audiences available. By that time the first fully endorsed outputs will also be uploaded into the repository.

# 3 Outputs publishing strategy

## 3.1 Introduction

According to the RDA Europe DoW, the Outputs publishing strategy will be devised based on the analysis of existing technologies, models and tools suitable for publishing and disseminating the outputs (section 3). One may claim however that a strategy should be independent of existing technologies and tools and in case specific strategic directions cannot be implemented with existing technologies and tools, these may have to be extended or new ones to be developed. In other words, the existing solutions may funnel and thus constrain the strategy by the existing technologies. Having reviewed and analysed existing technologies and tools in section 3, it may not be the case that a single solution (even the so-called “consolidation solutions” as defined in section 3) can serve all the outputs. After summarising the publishing strategy, we review this point.

A follow-up point that needs to be reminded in this introduction is that the RDA outputs will be a mix of different sorts of things, ranging from specifications, policies, best practices and recommendations, to frameworks, pieces of software, or prototype solutions or tools. And the strategy needs to take into account this variety of outputs that may target different audiences and thus required different approaches.



A final point in this introduction is the reviewers' recommendations on the final report of the first RDA Europe project and the direction that RDA should take.

*...Indeed, does RDA intend to present itself as a platform or does it prefer to be a standards/rules-based organization. This is an inherent tension that haunts the present evolution of RDA Europe: should it view itself more like organizations such as the IETF or W3c in the Internet world, or should it become a kind of Github of prototypes where various kinds of solutions applied to different kinds of data are presented in such a way as to facilitate transposition, applications, transformations, etc. The latter comparison would draw RDA-Europe closer to a platform.*

*At the present stage, the IETF/W3c metaphor may look more appropriate than a kind of Github transposed to the world of data, and it helps justify the changes mentioned later in the analysis of WP3. However, having a platform to initiate and test new solutions before moving to any attempt at standardizing and regulating would allow incorporating this bit of Internet wisdom so well captured in the phrase: "release early, release often so as to generate rough consensus and working code". Standardizing too soon and too fast without disposing of tested technical solutions could lead to solutions that cannot be easily implemented despite their apparent logical structure. Keeping close to the actual development communities is essential in this kind of development.*

*Once RDA starts to be recognized as equivalent to IETF and/or W3c, other structures may emerge later to take advantage of its standard-setting capacity and to help manage the concrete handling of specific categories of data. Here, the free-software universe, where similar issues have been encountered, offers another source of institutional and management wisdom.*

In other words, RDA should decide whether it should present itself in the future as "platform" or as a "standards or rules-based organisation". I.e. something like a Github of prototypes or snippets of code or something like IETF or W3C for Research Data. The reviewers hint towards the second, but highlight also the benefits of something close to a platform, as this will keep RDA close to the actual development communities and generate consensus.

### 3.2. Outputs publishing strategy

Taking into account the points raised in section 3.1, the following directions emerge for the outputs publishing and dissemination strategy:

- The publication and dissemination mechanism may need to be different for the different target audiences, given the wide variety of outputs (from policies to software). So, it will have to be **made available to each audience in way that will be relevant to them**. For the software snippets software repositories with versioning systems such as Github are more appropriate, while for policies and recommendations appropriate forums may be more suitable.
- In addition to the previous point, and given the fact that RDA is a new organisation, its future direction being currently developed<sup>5</sup>, the strategy needs to be **adaptive to changes**.
- As the initial outputs may be considered as **primitive, limited or still evolving**, it has to be made clear that it will take time to come up with a comprehensive set of outputs and thus a **modest approach** to the publication and dissemination of outputs is recommended, **not overselling the outputs**. The opposite may bring disillusionment and drive away potential adopters. If RDA was to be positioning in the so-called "hype cycle" (see figure on next page), it will still be positioned in the

<sup>5</sup> In fact an effort to devise the RDA future is on-going with a strategic consultation process.

“innovation trigger” area (where only first generation outputs are available and still lots of more effort is needed) and thus before the “peak of inflated expectations” area. So it has to be made clear that there may be a considerable and lengthy path towards second and third generation outputs. And in order to achieve these next generation results appropriate communities engagement is necessary acting as early adopters and working towards the next outputs.

Other points that need to be taken into account are the following:

- The Outputs Publishing Strategy needs to be **aligned with the overall RDA Strategic Planning exercise**<sup>6</sup> (as presented in RDA Plenary 5 in San Diego). The three main themes emerged from this exercise were Coordination, Communication and Engagement (see also figure 2). This will include the development and dissemination of stories around adopted outputs, the development of RDA domain or regional champions/ambassadors for better engagement and adoption, as well as the engagement of all appropriate stakeholders. These encompass individual data scientists, research and business organisations including SMEs, along with new regions and countries.
- The solution(s) and in particular the Outputs repository need to be **Open Access** and **following the RDA Outputs Policy**<sup>7</sup> (approved by the RDA Council).
- The dynamic role of the different RDA governing bodies should be taken into account, especially of the RDA Secretariat, TAB and OAB, as currently being discussed.

## The Hype Cycle of Innovation

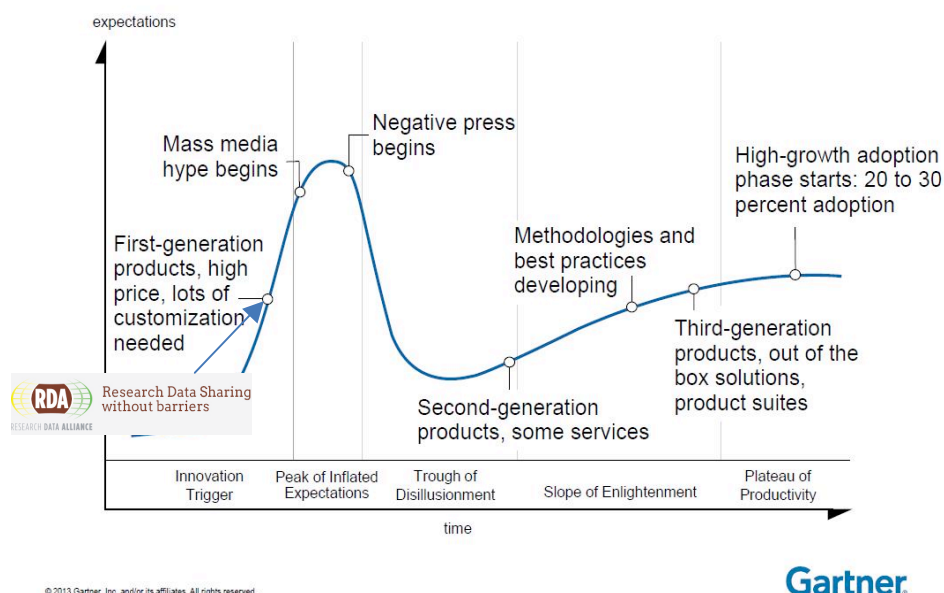


Figure 1 – RDA in the Gartner Hype Cycle

<sup>6</sup> <https://rd-alliance.org/groups/future-directions-planning.html> and <https://www.rd-alliance.org/future-directions-7-march-2015-workshop-report.html>

<sup>7</sup> <https://www.rd-alliance.org/filedepot/folder/132?fid=502>

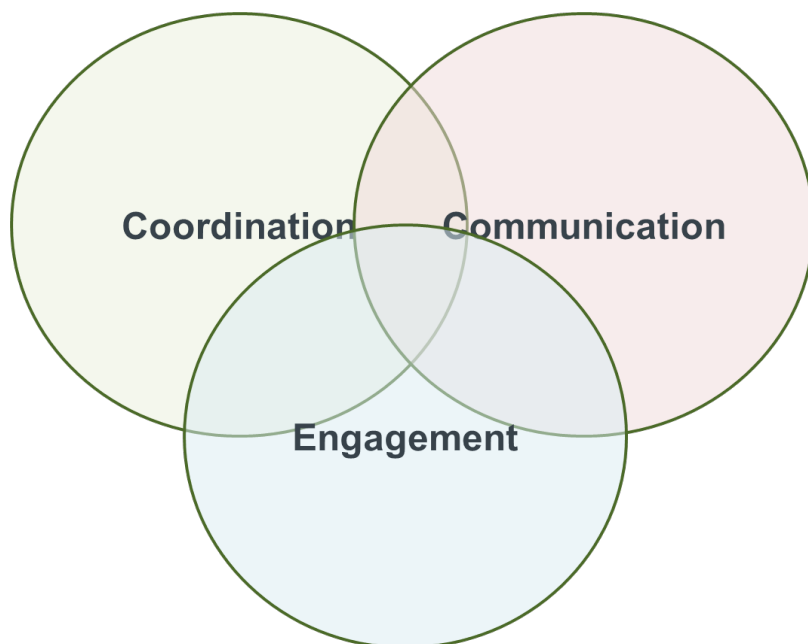


Figure 2 – RDA Strategic Consultation main themes

## 4 Conclusions and actions

Based on the above analysis, and the various (and sometimes conflicting) requirements, it is unlikely that a single repository or platform, even from the identified “consolidated solutions”, will fully serve the variety of RDA outputs and audiences in the long run. However -

- The selection of EUDAT B2SHARE, which appears to be an overall good solution, is a step towards this goal and will be used initially. It will continue to be re-evaluated after the next plenary by RDA Global, as more outputs with different characteristics become available.
- The EUDAT B2SHARE solution should be tested by a variety of audiences - this will help inform the strategic process, and keep audiences engaged.
- The ‘efficiency of procedures’ is essential for take-up – individuals are more likely to adopt tools and systems that work immediately. Where systems and tools don’t appear to work immediately (and /or provide immediate benefits) individuals are more likely to ignore them. Furthermore, outputs are still primitive and it will take some time to develop a comprehensive set of outputs. Thus the dissemination approach will have to be realistic and modest, not overselling the outputs, to avoid disillusioning potential early adopters.
- RDA outputs are expected to be not only specifications but also implemented code, adopted policy or practice, implemented tools, etc. Modern publishing allows us to link the specification outputs from RDA WGs with the textual descriptions and implementations of tools, services, policies, etc. - this enables interested individuals to take action immediately, to increase data sharing. Such resources will be exploited to facilitate fast and efficient exchanges.
- Outputs publishing solutions policies and strategies will be open, and in-line with the wider RDA Outputs policy.
- Related actions and recommendations also include the development and dissemination of stories around adopted outputs, the development of RDA domain champions for better engagement and adoption, as well as the engagement of all appropriate stakeholders. This encompasses individual data scientists, research and business organisations including SMEs, along with new regions and countries.
- The RDA must be clear on how it presents itself in the future. The recommendations from the reviewers report on RDA Europe 1 project<sup>8</sup> hint towards a standards/rules-based organisation, but also highlight the benefits of the RDA being a platform, as this helps keep the RDA close to the actual development communities. However, RDA is not intended to act as a platform, but nor as a standards organisation like IETF. RDA is its own thing, acting more as a vehicle for facilitating data sharing and creating data bridges among communities and regions that can be adopted. The social aspects of the RDA work prevent it from being as straightforward as a technical organisation like IETF.

---

<sup>8</sup> The first phase of the RDA Europe project