

# D4.1 – Description of Output Support Services and Maintenance Platform

Project Title	Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions
Project Acronym	RDA TIGER
Grant Agreement No	101094406
Instrument	HORIZON-INFRA-2022-EOSC-01
Торіс	HORIZON-INFRA-2022-EOSC-01-04
Start Date of Project	1 January 2023
Duration of Project	36 months
Project Website	https://www.rd-alliance.org/rda-tiger

**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)



Work Package	WP4 - Output Support
Lead Author (Org)	Wim Hugo (DANS)
Contributing Author(s) (Org)	Simon Saldner (DANS)
Reviewer	Ari Asmi (RDA)
Due Date	30.06.2023
Submitted Date	30.06.2023
Version	1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION
DOI	10.5281/zenodo.8096631







Version	Date	Authors	Notes
0.1	01.05.2023	Wim Hugo, Simon Saldner	Inception
0.2	02.05.2023	Wim Hugo, Simon Saldner	First draft
0.3	04.05.2023	Wim Hugo, Simon Saldner	Internal Review Copy
1.0	23.06.2023	Wim Hugo, Simon Saldner	Final draft

#### Versioning and contribution history

#### Disclaimer

RDA TIGER has received funding from the European Commission's HORIZON Coordination and Support programme under the Grant Agreement no. 101094406. The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.

#### **Abbreviations and Acronyms**

EC	European Commission
EOSC	European Open Science Cloud
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable, Accessible, Interoperable, Reusable
GORC	Global Open Research Commons
HEIS	Higher Education Institutions
KERs	Key Exploitable Results
IG	Interest Group
RDA	Research Data Alliance
SKG	Scientific Knowledge Graph
SRIA	The Strategic Research and Innovation Agenda
ТАВ	RDA Technical Advisory Board
TG	Task Group
TRL	Technology Readiness Level
WG	Working Group



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



### **Executive Summary**

Research Data Alliance's (RDA) Working Groups (WGs) aim to solve a specific research data related issue in a relatively short period of time. In practice, these groups are created by the RDA community, and rely on the voluntary contributions from the community to achieve its stated objectives. While this approach has been demonstrated to be successful in the last decade, certain areas have been identified where improvements can be made. The RDA TIGER project addresses some of these improvements [1].

RDA TIGER establishes a suite of concrete services to support the Working Groups (WG) of the Research Data Alliance (RDA). Work Package 4 in the TIGER project oversees the service for WG Output Support, particularly the RDA Maintenance Platform (RDA-MP), and various output guidance services. The objectives are to ensure high quality, usefulness, maintenance and sustainability of the WG outputs.

This report provides a work plan for the activities, outputs and, and services envisaged for the work package.

In section 1, we provide an overview of the issues, design considerations, requirements, and expectations that the work package needs to address. Section 2 explains how solutions to these can leverage existing efforts and opportunities in RDA and its stakeholder environment, and section 3 provides an inventory of the solution elements we have identified. Section 4 discusses some additional opportunities and the timelines associated with work package activities.

The solution elements cover aspects of policy and procedures, development of supporting standards and benchmarks, adopting or extending tools and systems to support the efforts of working groups, developing guidance and providing support, and enhancing the capacity of working groups.

We present additional information in appendices:

- Appendix A: On overview of the RDA Maintenance Platform;
- Appendix B: Thoughts on Digital Object Management in RDA;
- Appendix C: Detailed timelines and dependencies for Solution Elements;
- Appendix D: The Research Object Management Body of Knowledge ROM-BoK.

While it is not an explicit deliverable of the RDA TIGER project, we have documented an opportunity to create a Body of Knowledge for Research Object Management - involving a collaboration between RDA, ReSA, and its close stakeholder community. This opportunity is discussed in Appendix D, and its context is summarised in Section 3.6. It is relevant to RDA TIGER since it will build on infrastructure and services envisaged and created by the project.

As an unspoken but necessary outcome one could state that the WP aims to make RDA outputs and associated resources FAIR.





### **Table of Contents**

Executive Summary	3
1. Introduction: aim(s) and scope of the task	6
1.1 Focus Areas Identified	7
1.1.1 Curation-Related	7
1.1.2 Process-Related	8
1.1.3 Standards-Related	8
1.1.4 Complexity and Granularity	9
1.1.5 Capacity	9
1.1.6 Objectives	10
1.2 Working Group Workflow and Processes	11
2. Alignment with Work In Progress	13
2.1 Making Use of Existing Resources and Ongoing Initiatives	13
2.1.1 Landscape Service (RDA TIGER WP3)	13
2.1.2 RDA Maintenance Platform (Included into RDA TIGER WP4)	13
2.2 Extension of Resources in the Future	14
3. Scope and Nature of Planned Solutions	15
3.1 Policy and Procedures	15
3.2 Standards and Benchmarks	16
3.3 Systems and Tools	16
3.4 Guidance and Support	17
3.5 Capacity Enhancement, Training, and Service Assistance	19
3.6 ROM-BoK: The Research Object Management Body of Knowledge	19
4. Implementation Planning	20
4.1 An Opportunity	20
4.2 Timelines and Dependencies	20
References	22
Appendix A: RDA Maintenance Platform Details	23
A.1 Core Components	23
A.1.1 Problem Statement	23
A.1.2 Concept	23
A.1.3 Additional Criteria	24
A.1.4 RDA Maintenance Platform Components	25
A.2 Extension to the 'RDA Graph' and ROM-BoK	26
A.2.1 Design Considerations	27
A.2.2 Options for Implementation	28
A.2.3 Hybrid Solution	28
A.2.4 Conceptual Data Model	29
Appendix B: Metadata Continuum and the 'RDA Graph'	30
B.1 The Metadata Continuum	30
B.2 The Need for Agreed Vocabularies	31





B.3 The 'RDA Graph'	33
Appendix C: Support Responses and Solutions - Timeline	36
C.1 Detailed Milestones and Plenary Alignment	36
C.2 Plenary Targets	39
Appendix D: ROM-BoK	41
D.1 Context	41
D.2 Concept	42
D.3 Outcome	43
D.4 Implementation	44





### 1. Introduction: aim(s) and scope of the task

Research Data Alliance's (RDA) Working Groups (WGs) aim to solve a specific research data related issue in a relatively short period of time. In practice, these groups are created by the RDA community, and rely on the voluntary contributions from the community to achieve its stated objectives. While this approach has been demonstrated to be successful in the last decade, certain areas have been identified where improvements can be made. The RDA TIGER project addresses some of these improvements [1].

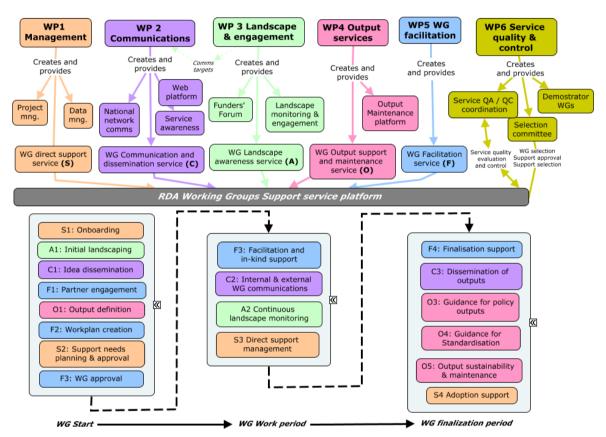


Figure 1. Work Package structuring and activities in RDA TIGER

Work Package 4 in the TIGER project oversees the service for WG Output Support, particularly the RDA Maintenance Platform (RDA-MP), and various output guidance services. The objectives are to ensure high quality, usefulness, maintenance and sustainability of the WG outputs. The WP will achieve this by contributing to and establishing the following:

- 01: Improvements to Output Definition;
- O3: Guidance for Policy Outputs;
- O4: Guidance for Standardisation;
- 05: Output Sustainability and Maintenance.

As an unspoken but necessary outcome one could state that the WP aims to make RDA outputs and associated resources FAIR.





### 1.1 Focus Areas Identified

In this section, we examine the broad issues, problems, desirable outcomes, and design considerations of RDA Working Groups that WP4 should focus on, grouped by theme.

#### 1.1.1 Curation-Related

#	Focus	Description	References
1.1.1.1	Inconsistent Metadata	Over time, the publication of RDA outputs and supporting materials have been subject to change, and as a result, not all outputs and supplementary materials have similar metadata or a consistent curation process applied to them.	
1.1.1.2	Generalised Metadata	RDA outputs have been published to Zenodo for about 6 years now, and in such cases the minimum metadata associated with Zenodo/ DataCite publications are required. This metadata is generalised, and unless the depositor adds appropriate (and unmanaged) tags, RDA-specific elements are not accommodated in metadata.	[3]
1.1.1.3	Multiple repositories	RDA has used multiple repositories for both outputs and other supplementary materials over time. Zenodo contains a comprehensive collection of outputs, but it is not guaranteed to be complete. Zenodo also contains a partial collection of presentations, supplementary materials, reports, and other resources related to RDA, but so does the repository folders in the RDA website. Many supplementary materials, work in progress, and resources are stored in shared online repositories such as Google Drive, and are not archived under RDA control or curated by RDA.	
1.1.1.4	Collection Management	The RDA has established two collections in Zenodo, one for formal outputs, and one for supplementary materials. Both these collections appear somewhat variable in terms of content despite a moderation requirement for deposits to the collections. The formal outputs collection, for example, contain many non-outputs in the narrow sense of the work (i.e., RDA-endorsed recommendations).	[2], [4]
1.1.1.5	Long-Term Preservation	Outputs that are maintained and made available from the RDA website have very limited preservation certainty, and those published in Zenodo are likely to be preserved but Zenodo does not explicitly provide a long-term preservation service aligned with trustworthy repository criteria.	







#### 1.1.2 Process-Related

#### Table 2. Process-Related Focus areas

#	Focus	Description	References
1.1.2.1	Research Methodology	Recent improvements in annotation of web-based resources have implications for research methodology and management of references that can be included into working group practices.	
1.1.2.2	Landscape Analysis	Similarly, most Working Groups need to do some form of landscape analysis, and this is repeated without benefiting from similar analysis done within RDA or in associated initiatives in the landscape. (This aspect is addressed in detail in the work of RDA TIGER WP3).	[5]
1.1.2.3	Mobilisation and Re-use of Web Resources	During the course of the work of a Working Group, it is common to review a number of web-based resources and publications, and then analyse and synthesise the information, with working group and community input, into a set of recommendations, a model, or similar output. The effort expended in reviewing these input resources is often not re-usable except with some difficulty, given that at most the resources will be made available as references.	
1.1.2.4	Structuring of Recommendations	Recommendations, best practices, and guidelines are commonly reported in monolithic documents, making it difficult to find and re-use them across different documents or to aggregate a set of recommendations for a given topic.	
1.1.2.5	Publication Process	The publication process for RDA outputs and supplementary materials is not standardised and has changed over time.	
1.1.2.6	Dissemination Channels	RDA outputs are disseminated primarily via the RDA website, and there is limited possibility for third-party applications to harvest and use the corpus of RDA work in a different environment (currently only possible for material published in Zenodo).	
1.1.2.7	Re-Use and Impact	Dissemination channel options, and partial publication of RDA outputs with persistent identifiers leads to limitations on re-use and gathering of metrics in respect of re-use.	

#### 1.1.3 Standards-Related

#### Table 3. Standards-Related Focus areas

#	Focus	Description	References
---	-------	-------------	------------





1.1.3.1	Precision	The precision that can be expected of annotation increases significantly with standardisation of the vocabularies used for tagging.	
1.1.3.2	Metadata Standards	Standards for metadata applicable to different types of RDA outputs will improve interoperability and reusability	
1.1.3.3	Templates, Branding, and Version Management	Standards for presentations, reports, recommendations, and the minimum content, branding, and versioning information for each of these will improve the image, reusability, and quality of the outputs.	

#### 1.1.4 Complexity and Granularity

#### Table 4. Complexity and Granularity -Related Focus areas

#	Focus	Description	References
1.1.4.1	Complex Landscape	<ul> <li>RDA working groups need to become aware of and understand their positioning in a landscape that is becoming ever more complex, with multiple initiatives and thrusts focused on the same issues and problems in the research data management ecosystem. It is important to identify: <ul> <li>what has been previously done on a given subject and what should be considered in the Case Statements;</li> <li>who is working on this subject area in the global Open Science (OS) landscape, including e.g., European Open Science Cloud projects, and;</li> <li>who has the necessary expertise and interest to participate in these actions.</li> </ul> </li> </ul>	RDA TIGER landscape analysis [5]
1.1.4.2	Proliferation of Guidance and Outputs	There is a significant investment in the developed world in the creation and formulation of best practices, guidance, and recommendations in respect of a wide variety of topics of interest to the community. As yet, there is no real attempt to consolidate, standardise, and collate these resources to avoid duplication, divergence, and maximise return from these investments, or to make them machine- readable and actionable.	
1.1.4.3	Buried Recommendations	Moreover, many of the recommendations, best practices, and guidance are reported within deliverables and outputs that are linked to project websites and structures, and as such are often not published in mainstream channels. Such reports seldom result in widely adopted recommendations or specifications governed by organisations such as ISO <sup>1</sup> ,	

<sup>&</sup>lt;sup>1</sup> ISO as a standards body may present other challenges: for example standards are not openly accessible by default.



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



W3C, or IETF.	
---------------	--

#### 1.1.5 Capacity

#### Table 5. Capacity-Related Focus areas

#	Focus	Description	References
1.1.5.1	Voluntary Contributions	Working Groups are largely resourced via voluntary contributions from members, and in most cases only a few workgroup members contribute to the main body of work in the working group.	
1.1.5.2	Bias and Skewed Resources	Resources are sometimes available due to overlap and synergy with funded tasks and outcomes in which workgroup members participate. On the one hand, this contributes funded resources to the working group, but on the other hand, it results in disproportionate contributions from these sources - not always therefore a balanced view or broad consensus.	Support services description provides potential solution (D1.1)
1.1.5.3	Content Focus	Working Groups tend to focus on and have intentions to create content (a set of recommendations, a specification) and do not have capacity, experience, or incentives to look at other aspects of product readiness: communication, adoption, dissemination, standards creation, and so on.	
1.1.5.4	Capital and Operating Expenditure	RDA receives most of its 'capital expenditure' for free, by way of voluntary contributions of members (many of whom are funded indirectly via grants or their institutions to participate). It is not feasible to rely on voluntary maintenance of RDA assets (outputs, especially) via the same mechanism: grants often do not focus on these activities, and maintenance cannot be left to ad-hoc arrangements.	

#### 1.1.6 Objectives

#	Objective	Description	References
1.1.6.1	Improved Quality	There is a need to make sure that all WG outputs meet minimum quality standards, and that these improve over time.	[1], [14]
1.1.6.2	Improved Usefulness	Usefulness of outputs are dependent on the perspective of the end-user, and working groups are expected to align with	[1]





		identified needs and expectations of the community. At least some usability is a direct result of standardisation and presentation (e.g. machine-readability, findability).	
1.1.6.3	Proper Maintenance	Proper maintenance of RDA outputs after publication - in the short term and in respect of long-term preservation - is not addressed adequately in current working group obligations - essentially an optional aspect of RDA member activity.	[1]
1.1.6.4	Improved Sustainability	Sustainability of RDA outputs may have to be evaluated against emerging principles for open scholarly infrastructure, for example as exemplified by POSI.	[1], [6], [7]

#### 1.2 Working Group Workflow and Processes

The Landscape Analysis report [5] produced by RDA TIGER identifies the following high-level steps in the life cycle of a working group:

- **Preparation**: The project support actions include careful preparation of the Working Groups, such as planning their timeline, wide dissemination of their existence for wider participation, resource allocation, and landscape analysis of similar activities. Significant efforts are made to gain necessary international and European participation in these WGs.
- Facilitation: "During the active WG period, the [RDA TIGER project] provides facilitation services, including dedicated facilitator(s) to ensure that the work progresses appropriately, communication and analysis services (e.g. landscape monitoring), and several optional direct support actions, such as hiring consultancy for support tasks (e.g. surveys), providing travel grants for meeting participants, and even supporting minor but necessary development or research tasks related to WG work with small 3rd party grants. The WGs are closely monitored by the facilitators and the project management to ensure that WG progress is appropriate and corrective actions are taken, e.g. via offering additional support actions.
- Finalisation: Towards the end of each Working Group's life cycle, and after, the project supports different actions towards concretisation, exploitation, maintenance, and sustainability of the WG results. These actions include last-mile facilitation, advisory support actions towards standardisation or on international policy harmonisation, and can include concrete actions towards result maintenance in RDA, as well as extensive dissemination actions. Additional 3rd party grants can be awarded for adoption test cases to demonstrate the feasibility of the results.





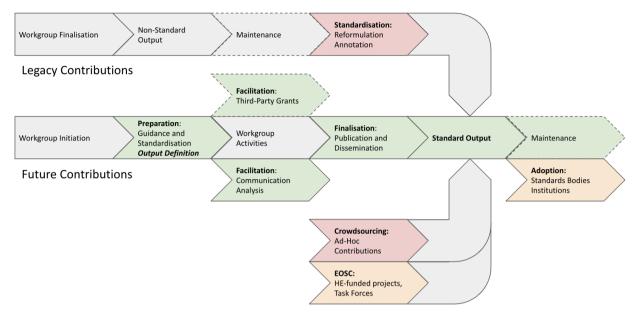


Figure 2. Amended and Proposed Workflows for Working Group Operations The figure illustrates additional workflow steps and stages that WP4 (RDA TIGER Service contribution) adds - shown in green with potential collaborations (requiring additional engagement) in amber, and funding needs in red.

The RDA TIGER WP4 will add to these broad outlines with concrete actions during the life cycle of current and future RDA Working Groups and has some specific recommendations in respect of legacy Working Group outputs.

Historically, RDA Working Groups have generated outputs in a non-standardised way, and the scope, granularity, and composition of the outputs were largely left to the working group to decide. This has advantages, since being too prescriptive can lead to limitations on working group creativity, and standardisation places a burden on voluntary effort to ensure compliance. As a result, working group outputs have been diverse. Moreover, maintenance is currently also voluntary, and there is no certainty that outputs will be maintained, or progress to wider adoption vis. standards authorities such as ISO, IETF, or W3C.

Work Package 4 in RDA TIGER provides an opportunity to address these deficiencies, at least in part, by creating solution elements that contribute to the following steps in the workflow:

- **Preparation**: Standardisation of the way in which workgroups do their research to maximise its reusability, and upfront guidance in respect of structuring and standardisation of outputs and supplementary materials.
- Service support for WGs (Facilitation, communication, landscape analysis and direct support services): These aspects will not be addressed by WP4, but by other activities in RDA TIGER. Optional 3rd-party grants allocated to selected working groups may require guidance and support from WP4.





- **Finalisation**: This step will require direct assistance and support from WP4, ensuring proper publication, structuring, annotation, and dissemination of outputs, web-based resources referenced or created by the working group, and relevant supplementary materials.
- **Maintenance**: The RDA TIGER project also has a finite existence, but structures and procedures for maintenance of RDA outputs need to be considered and established during the lifetime of the project.
- Adoption: Likewise, the RDA TIGER project does not have in scope to facilitate the adoption of outputs formally in standards organisations such as ISO or W3C, but the process to do so can be established, agreed with relevant standards organisations, and included into WG guidance and procedures.

In addition, it may be possible to make progress, at least in part, with the following collaborative opportunities:

- **Standardisation of Legacy Outputs**: RDA has historically produced outputs that are not standardised, and funding or collaboration opportunities need to be found to address this.
- Adoption in Standards Bodies: TIGER WP4 will endeavour to document the processes for use by RDA working groups, but it will not be possible within the existing scope to assist directly.
- **Crowdsourcing**: It will be possible for any interested party to contribute to the annotation effort, and initially, RDA members and participating institutions provide a starting point.
- **EOSC Focus**: This will require engagement with current EOSC projects and initiatives to maximise the standardisation of their future outputs.

### 2. Alignment with Work in Progress

### 2.1 Making Use of Existing Resources and Ongoing Initiatives

#### 2.1.1 Landscape Service (RDA TIGER WP3)

The Landscape Service consists of two major components: the development and maintenance of the database containing landscape analyses, and the delivery of the actual landscape analyses to the WGs. It is important to understand that RDA TIGER is part of the larger Open Science environment and the landscape analysis provided focuses on what the project perceives as the immediate needs of the RDA WGs, a perception that will evolve over the life of the project.

The Landscape Analysis Database will provide an extension of the database created for the RDA Maintenance Platform (RDA-MP) and enables linking of existing and future RDA outputs (resources, activities, structures, and stakeholder engagement), with EOSC resources, and with a wider Open Science context.



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



#### 2.1.2 RDA Maintenance Platform (Included into RDA TIGER WP4)

RDA TIGER will make use of and further refine the RDA Maintenance Platform, currently under development (see Appendix A). The platform consists of 4 subservices in addition to the infrastructure:

- **Provision of a Helpdesk and Service Level Support**: Assisting users with queries, problems, issues, and suggestions for improvement, measured against a service-level agreement in terms of response and resolution time;
- **Issue and Request Resolution**: providing functional controller and developer support for resolution of technical system issues, resolving non-technical issues with assistance and contribution from RDA Secretariat staff;
- Limited New Functionality Development: A prioritised list of new feature requests to be maintained from the ticket system, together with estimates of time and effort, system impact. If maintenance time is not taken up by issue resolution, the team can implement new features;
- **Technical Workshops at RDA Events**: Provision of one or more workshops peripheral to RDA events (plenaries) to coach and support system users.

This platform directly supports the 'O' service collection in RDA TIGER [1], supplemented by the tools and solution components defined in Section 3:

- **O1 Output Definition (WG Start)**. This activity includes:
  - initial definition of the output scope,
  - o the user communities (inc. potential adopter communities),
  - o required support actions (e.g., for standardisation support),
  - o and maintenance and sustainability requirements.
  - The output definition document is agreed with the WG partners, describes the intended output in detail for the project management and outreach, and is used to plan the activities.
- **O2: Guidance for policy outputs (WG Finalisation)**. This helpdesk service provides guidance to WGs that develop policies in the area of research data in the broad sense, including Open and FAIR data policies. On domain level, national level and in the EOSC context this kind of policy is being formulated and revised, e.g., by EOSC Association Task Forces. The helpdesk helps RDA WGs to check how their planned or draft policy document fits into the wider policy landscape.
- O3: Guidance for Standardisation (WG Finalisation): RDA working groups face a significant challenge in internalising and parsing the EOSC landscape, standards, and requirements as an input to working group activities, and likewise, ensuring that their recommendations and outputs are aligned with and implementable in the context of EOSC. An expert service is provided to bridge the gap and maintain guidelines, design considerations, and specifications applicable. The service integrates with the RDA Maintenance Platform.
  - Maintain an inventory of EOSC regulations, policies, requirements, specifications, guidance, and recommendations for technical services and components.



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



- Structure, classify, and annotate the resources on the basis of alignment and overlap with RDA groups.
- Enable to IG and WG members to contribute to annotation.
- Assist working groups with publication of standards and make use of EU-funded structures designed for the purpose<sup>2</sup>.

#### 2.2 Extension of Resources in the Future

There are several ways in which the work already underway can be leveraged and multiplied in terms of impact and usefulness to the community:

- The RDA Maintenance Platform provides a basic set of tools and infrastructure (Appendix A.1).
- This can and will be extended to accommodate the needs of RDA TIGER for landscape-related links (Appendix A.2, Appendix B)
- It can also be extended within RDA to accommodate other digital objects that need to be managed (Appendix B)
- The net can be cast wider: standardising the annotation and availability of domain knowledge about research data and code infrastructure and its management in a formal 'Body of Knowledge' - Appendix D.

The work to establish vocabularies and graph infrastructure for use by WP4, based on 2.1 above, is well advanced and a first release can be expected at the end of June 2023.

### 3. Scope and Nature of Planned Solutions

The following tables lists the proposed solutions and tools that are planned for use in WP4 to support RDA Working Groups. In the table, the solutions are matched with the problem(s) that are addressed, and an indication is given whether the solution benefits the Working Groups identified for support by RDA TIGER, all RDA Working Groups, or both. There are two identified problems that do not have at least one solution element that mitigates it: these are 1.1.6.3- Proper Maintenance, and 1.1.6.4- Improved Sustainability. These require wider consultation and remedial action within the RDA community.

### 3.1 Policy and Procedures

#	Solution	Description	Benefits	Problem(s) Solved
3.1.1	Policy - Outputs	RDA will benefit from a policy in respect of output production and publication - dealing with aspects	<ul><li>No ambiguity</li><li>Alignment</li></ul>	1.1.2.1, 1.1.2.4 1.1.2.5, 1.1.2.6 1.1.3.2, 1.1.3.3

Table 7. Policy and Procedures Solutions

<sup>&</sup>lt;sup>2</sup> <u>https://www.hsbooster.eu/</u>



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



		such as licensing and access, branding, ownership, attribution, and maintenance.		1.1.4.2, 1.1.5.3 1.1.6.1, 1.1.6.2
3.1.2	Procedure: Workgroup Creation and Life Cycle	This exists to some degree, but must be reviewed in the light of additional support materials, guidelines, mechanisms, and tools required for the ' <b>Output Definition</b> <b>Report</b> ', and must align with policy <sup>3</sup> .	<ul> <li>Clarity</li> <li>Defined process</li> <li>Improved Planning</li> </ul>	1.1.2.1, 1.1.2.4 1.1.5.2, 1.1.5.3 1.1.6.1
3.1.3	Procedure: Research Processes	Procedures for annotation and referencing of resources and publications in the course of the workgroup's activities.	<ul> <li>Reusability</li> <li>Cross-workgroup benefits</li> <li>Landscape</li> </ul>	1.1.2.1, 1.1.2.2 1.1.2.3, 1.1.2.7 1.1.3.1, 1.1.4.1 1.1.4.2, 1.1.5.2 1.1.6.1, 1.1.6.2
3.1.4	Procedure: Publication of Outputs	RDA Working Groups need to adopt a common procedure for publication of outputs and supplementary materials, based on the RDA Maintenance Platform.	<ul> <li>Reusability</li> <li>Consistency</li> <li>Discoverability</li> <li>Machine- readability</li> </ul>	1.1.1.1, 1.1.1.2 1.1.1.3, 1.1.2.4 1.1.2.5, 1.1.2.6 1.1.3.1, 1.1.3.2 1.1.3.3, 1.1.4.2 1.1.6.1, 1.1.6.2
3.1.5	Procedure: Curation of RDA Collections	RDA collections in Zenodo need accession procedures that will ensure consistency of materials and resources in the collections.	<ul> <li>Discoverability</li> </ul>	1.1.1.1, 1.1.1.2 1.1.1.3, 1.1.1.4 1.1.2.4, 1.1.2.5 1.1.2.6, 1.1.2.7 1.1.6.1, 1.1.6.2
3.1.6	Procedure: Selection and Maintenance of Vocabulary	The process whereby RDA identifies vocabulary and how it is governed and maintained must be documented.	<ul> <li>Reusability</li> <li>Consistency</li> <li>Discoverability</li> </ul>	1.1.1.1, 1.1.3.1 1.1.4.1, 1.1.6.1 1.1.6.2

### 3.2 Standards and Benchmarks

Table 8. Standards and I	Benchmarks Solutions
--------------------------	----------------------

#	Solution	Description	Benefits	Problem(s) Solved
3.2.1	Tag Lists, Registries, and Vocabularies	RDA needs to use standard APIs and encodings for its own vocabularies. These need to extend to and allow the linking of external resources as intended in	<ul> <li>Machine- readability</li> <li>Landscape integration</li> </ul>	1.1.1.1, 1.1.2.1 1.1.2.2, 1.1.2.3 1.1.2.4, 1.1.2.7 1.1.3.1, 1.1.4.1 1.1.4.2, 1.1.5.2

<sup>&</sup>lt;sup>3</sup> Clarity is needed from RDA on how this differs from Case Statements required from new groups. Needs to be differentiated clearly.



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



		O1 and O3.		1.1.6.1, 1.1.6.2
3.2.2	Preferred Formats for Dissemination and Preservation	These are needed to maximise utility for the community and to ensure long-term preservation	<ul> <li>Reusability</li> </ul>	1.1.1.4, 1.1.1.5 1.1.2.1, 1.1.2.5 1.1.2.6, 1.1.2.7 1.1.6.1
3.2.3	Annotation Standards and Schema	RDA will define a number of standard schemas: metadata, rating, and landscape context are examples.	<ul> <li>Discoverability</li> <li>Reusability</li> <li>Machine- readability</li> </ul>	1.1.1.1, 1.1.1.4 1.1.1.5, 1.1.2.1 1.1.3.1, 1.1.4.1 1.1.6.1, 1.1.6.2
3.2.4	Minimum Metadata Schema	These have to be differentiated for different output and resource types, while satisfying repository and RDA requirements.	See above	1.1.1.1, 1.1.1.2 1.1.2.5, 1.1.2.6 1.1.3.1, 1.1.3.2 1.1.6.1, 1.1.6.2
3.2.5	Templates for Output Reports, Recommendations, Standards and Specifications	A number of standardised templates are needed for typical RDA outputs and supporting materials. One objective is to simplify annotation and machine- readability.	<ul> <li>Consistency</li> <li>Machine- readability</li> </ul>	1.1.1.3, 1.1.1.4 1.1.1.5, 1.1.2.1 1.1.2.4, 1.1.2.5 1.1.2.6, 1.1.2.7 1.1.3.3, 1.1.4.1 1.1.4.2, 1.1.4.3 1.1.6.1, 1.1.6.2

### 3.3 Systems and Tools

Table 9. S	vstems (	and Tool	Solutions
10010 010	,		001010110

#	Solution	Description	Benefits	Problem(s) Solved
3.3.1	RDA Maintenance Platform	The RDA Maintenance Platform plays a central role in support of WGs and associated RDA activities: publication, annotation, discovery, and preservation. See Appendix A.	<ul> <li>Consistency</li> <li>Machine- readability</li> <li>Landscape integration</li> <li>Discoverability</li> </ul>	1.1.1.1, 1.1.1.2 1.1.2.1, 1.1.2.2 1.1.2.3, 1.1.2.4 1.1.2.5, 1.1.2.6 1.1.3.1, 1.1.4.3 1.1.5.2, 1.1.6.2
3.3.2	Reference Managers	Integration between mainstream Reference Managers and the RDA MP will benefit end-users.	<ul><li>Usability</li><li>Efficiency</li></ul>	1.1.1.1, 1.1.1.4 1.1.2.1, 1.1.6.2
3.3.3	Annotation Platforms	Integration with annotation platforms such as OKRG and hypothes.is will broaden the scope of end users and extend the RDA graph/ knowledge base.	<ul> <li>Consistency</li> <li>Landscape integration</li> <li>Discoverability</li> <li>Usability</li> <li>Cross-workgroup benefits</li> </ul>	1.1.1.1, 1.1.2.1 1.1.2.2, 1.1.2.3 1.1.2.6, 1.1.2.7 1.1.3.1, 1.1.4.1 1.1.5.2, 1.1.6.2





3.3.4	Helpdesk	The Helpdesk assists working groups and other RDA members to request assistance with tools, software, templates, or any other solution element	<ul> <li>Usability</li> <li>Efficiency</li> <li>Clarity</li> <li>Defined process</li> </ul>	See below
3.3.5	Helpdesk Knowledge Base	The guidance and support materials (below) serve as an entry point to the policies, procedures, standards, tools, and best practices available to working groups. These are made available as HelpDesk knowledge base articles, in addition to being published and annotated.	<ul> <li>Consistency</li> <li>Usability</li> <li>Efficiency</li> <li>Clarity</li> <li>Defined process</li> <li>Improved Planning</li> </ul>	1.1.1.1, 1.1.2.1 1.1.2.2, 1.1.2.3 1.1.2.4, 1.1.2.5 1.1.2.6, 1.1.2.7 1.1.3.1, 1.1.3.2 1.1.3.3, 1.1.4.1 1.1.4.2, 1.1.5.4 1.1.6.1, 1.1.6.2
3.3.6	The 'RDA Graph'	The need to accommodate a wider set of entities, concepts, and relations emanating from RDA within its landscape and ecosystem has to be accommodated: largely by providing access to additional schema, vocabularies and registries as a basis for annotation.	<ul> <li>Consistency</li> <li>Machine- readability</li> <li>Landscape integration</li> <li>Discoverability</li> <li>Usability</li> <li>Efficiency</li> <li>Cross-workgroup benefits</li> </ul>	1.1.1.1, 1.1.2.1 1.1.2.2, 1.1.2.3 1.1.2.5, 1.1.2.6 1.1.2.7, 1.1.3.1 1.1.4.1, 1.1.4.2 1.1.4.3, 1.1.5.2

### 3.4 Guidance and Support

#	Solution	Description	Benefits	Problem(s) Solved
3.4.1	Guidelines: Workgroup Establishment	<ul> <li>Description of life cycle, which solution elements are available for and should be used in each step of the life cycle, and how to</li> <li>Generate the Output Definition Report;</li> <li>Create outputs and support materials that are publication and machine-ready.</li> <li>Annotate and reference individual elements of outputs in a reusable way.</li> </ul>	<ul> <li>Consistency</li> <li>Landscape integration</li> <li>Discoverability</li> <li>Usability</li> <li>Efficiency</li> <li>Cross-workgroup benefits</li> <li>Defined process</li> <li>Improved Planning</li> </ul>	1.1.2.2, 1.1.2.5 1.1.5.3, 1.1.6.1 1.1.6.2
3.4.2	Guidelines: Maximising Workgroup Effort	Includes landscape analysis support, annotation best practices, structuring of outputs and supplementary materials, making effort reusable -	<ul> <li>Consistency</li> <li>Landscape integration</li> <li>Efficiency</li> </ul>	1.1.2.1, 1.1.2.2 1.1.2.3, 1.1.2.4 1.1.2.5, 1.1.2.6 1.1.2.7, 1.1.3.1





		overlaps with the previous, but adds guidelines for annotation of web resources.	•	Cross-workgroup benefits Defined process	1.1.3.2, 1.1.3.3 1.1.4.1, 1.1.4.2 1.1.5.2, 1.1.5.3
3.4.3	Documenting Standards and Specifications	It is anticipated that standards and specifications will be occasional outputs from working groups, and to ensure that these are documented in a consistent and publishable way, a template with attendant guidance will be developed for this aspect. WP4 will endeavour to make use of EU-funded support services for this <sup>4</sup> .	•	Consistency Usability Efficiency Cross-workgroup benefits Defined process Machine- readability	1.1.2.1, 1.1.2.4 1.1.2.5, 1.1.2.6 1.1.2.7, 1.1.3.1 1.1.4.2, 1.1.4.3 1.1.5.3, 1.1.6.1 1.1.6.2
3.4.4	Standards Authority Adoption Guidelines	Further support of the process of creating and disseminating standards and specifications: guidance will be developed, in consultation with e.g., ISO and W3C to guide working groups on processes to formally adopt their work. Will also make use of EU- funded support services where applicable <sup>4</sup> .	•	Usability Efficiency Cross-workgroup benefits Defined process	1.1.2.4, 1.1.2.5 1.1.2.7, 1.1.3.3 1.1.4.3, 1.1.5.3 1.1.6.1, 1.1.6.2
3.4.5	Informal Contributions to the 'RDA Graph'	It is possible and desirable for individual RDA members to annotate, rate, and comment materials referenced and produced by RDA. To assist with this, a guidance document will be developed.	• • • • •	Consistency Landscape integration Discoverability Usability Efficiency Cross-workgroup benefits Defined process	1.1.2.2, 1.1.2.3 1.1.2.7, 1.1.4.1 1.1.4.3, 1.1.5.2 1.1.6.2

### 3.5 Capacity Enhancement, Training, and Service Assistance

Table 11.	Capacity Solutions
-----------	--------------------

#	Solution	Description	Benefits	Problem(s) Solved
3.5.1	Workshops: Workgroup Output Support	Training materials and plenary- aligned workshops to highlight and introduce the support mechanisms developed by RDA TIGER.	<ul> <li>Consistency</li> <li>Efficiency</li> <li>Cross-workgroup benefits</li> <li>Defined process</li> <li>Improved Planning</li> </ul>	1.1.2.2, 1.1.2.3 1.1.2.5, 1.1.2.6 1.1.2.7, 1.1.3.2 1.1.3.3, 1.1.4.1 1.1.5.1, 1.1.5.3 1.1.6.1, 1.1.6.2

<sup>&</sup>lt;sup>4</sup> https://www.hsbooster.eu/



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



3.5.2	Service Assistance: Curation	Curation activities for Working Groups to publish their efforts and outputs in the recommended way. Include into plenary workshops.	<ul> <li>Capacity Building</li> <li>Examples and Reference Implementation</li> </ul>	1.1.2.5, 1.1.2.6 1.1.2.7, 1.1.3.2 1.1.3.3, 1.1.4.1 1.1.5.1, 1.1.5.3 1.1.5.4, 1.1.6.2
3.5.2	Alignment with other TIGER initiatives	The solution elements should be included into and align with relevant TIGER deliverables - see 4.2 below.	<ul> <li>Consistency</li> <li>Efficiency</li> <li>Cross-workgroup benefits</li> <li>Capacity Building</li> <li>Examples and Reference Implementation</li> </ul>	1.1.2.2, 1.1.2.3 1.1.2.5, 1.1.2.6 1.1.2.7, 1.1.3.2 1.1.3.3, 1.1.4.1 1.1.5.1, 1.1.5.3 1.1.6.1, 1.1.6.2

### 3.6 ROM-BoK: The Research Object Management Body of Knowledge

A large investment is underway in the development of knowledge in respect of research data management, starting with funding agency grants for applied research in the field, and increasingly extending to other research outputs (code, semantic artefacts, methodologies, instruments, and so on) as a result of the Open Science driver.

The results of these investments are typically captured in reports, and contain a mixture of guidance, recommendations, templates, specifications, designs, and best practices. As a rule, these reports are published through channels ranging from project websites to peer-reviewed journals.

The proliferation of such guidance, and the fact that the metadata, keyword conventions, and vocabularies used for tagging and description are not standardised leads to a complex situation facing those that want to implement best practices.

To address the issue, a parallel effort to which RDA TIGER, the EOSC-focused HE projects and EOSC Task Forces, and RDA stakeholders such as ReSA can contribute in a non-binding way is needed.

There is a precedent in several fields for creation of community-maintained 'Bodies of Knowledge' (BoK), which serve as a summary and overarching guide to the formal knowledge about the subject. We propose a similar solution but designed to be both human- and machine usable, in Appendix D: The Research Outputs Management Body of Knowledge, or ROM-BoK. The proposed effort does not create any dependency in respect of the solution components created by RDA TIGER, but the RDA TIGER solution elements can fit directly into such a body of knowledge.





### 4. Implementation Planning

### 4.1 An Opportunity

The work in RDA TIGER WP4 fits into a wider context, where we are aligning our milestones and deliverables to feed into RDA Plenary events, funding opportunities, and ongoing EU-funded projects within the EOSC initiative. These interrelated opportunities all, in principle, contribute to the establishment of an RDA body of knowledge, which we describe in more detail in Appendix D. This will happen irrespective of how the body of knowledge is presented and whether it is formalised or not.

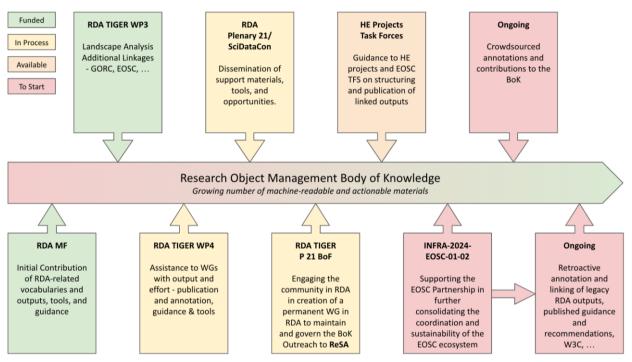


Figure 3: Overview of Opportunities to complete the Body of Knowledge

#### 4.2 Timelines and Dependencies

The two funded activities that will allow this process of knowledge encoding and formalisation to commence, and is described in this report, are the RDA TIGER project, and the project to establish the RDA Maintenance Facility. The interdependencies between RDA TIGER WP tasks, RDA Maintenance Facility Tasks, and RDA Plenary events have been determined, and documented in Appendix C, together with the expected solutions (procedures, policies, tools, standards, and similar) that will be available at the time of each plenary for dissemination to and use by RDA Working Groups.

In summary terms, these objectives are as follows:





Plenary	Main Objectives
P21	Disseminate procedures in respect of Research Processes and Curation of RDA Collections Standardise Preferred Formats for Dissemination and Preservation, and Minimum Metadata Schema Implement the RDA Maintenance Platform, Helpdesk, and Helpdesk Knowledge Base Publish guidelines for Workgroup Establishment Conduct one or more workshops to introduce the above to Working Groups Service assistance with curation of RDA materials
P22	Implement the Annotation Platform Publish guidelines for Maximising Workgroup Effort Operationalise and explain Informal Contributions to the 'RDA Graph' Conduct one or more workshops to introduce the above to Working Groups Service assistance with curation of RDA materials
P23	Disseminate procedures for Workgroup Creation and Life Cycle, and for Selection and Maintenance of Vocabulary Publish and introduce community maintenance of Tag Lists, Registries, and Vocabularies, as well as Annotation Standards and Schema Release an updated version of the RDA Maintenance Platform, and the Helpdesk Knowledge Base Publish guidelines on Documenting Standards and Specifications, and on Standards Authority Adoption Conduct one or more workshops to introduce the above to Working Groups Service assistance with curation of RDA materials
P24	Publish and disseminate an RDA Policy on Output and an updated Minimum Metadata Schema Release integration of Reference Managers and external Annotation Platforms Conduct one or more workshops to introduce the above to Working Groups Service assistance with curation of RDA materials

The work also needs to align with other RDA TIGER initiatives and contribute to them as applicable. The following table summarises the identified opportunities and dependencies:

Milestone/ Deliverable	RDA Tiger Task	Dependencies and Alignments
M15	1st update of the handbook for facilitation	Include references to support materials and solution elements that are available at that time.
M16, 17, 18	Updates to the service quality	Include references to support materials and solution

Table 13. Dependencies and Alignments





	handbook	elements that are available at that time.
D3.4	Landscape analyses and potential engagement targets	Provide templates, tools and structuring of analysis, and support landscape analyses with graph data.

### References

- [1] Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions (RDA-TIGER), Project Proposal
- [2] Research Data Alliance, Formal Outputs Collection, https://zenodo.org/communities/rda/?page=1&size=20
- [3] Research Data Alliance, earliest deposit to formal outputs, https://zenodo.org/communities/rda/search?page=1&size=20&sort=-mostrecent
- [4] Research Data Alliance
- [5] RDA TIGER D3.1 OS Landscape analysis and potential engagement targets, 10.5281/zenodo.8096612
- [6] Principles of Open Scholarly Infrastructure, <u>https://openscholarlyinfrastructure.org/</u>
- [7] W Hugo, M Buys: Compliance Case Study Analysis: POSI, 2023, http://doi.org/10.5281/zenodo.7933879 (under review)
- [8] ACTRIS, 2020: G 440.4.4.1 Information Management Blueprint, Internal Report
- [9] Research Software Alliance, <u>https://www.researchsoft.org/</u>
- [10] IEEE, Software Engineering Body of Knowledge, <u>https://www.computer.org/education/bodies-of-knowledge/software-engineering</u>
- [11] PMBOK<sup>®</sup> Guide (2021). Project Management Institute, The Project Management Body of Knowledge, <u>https://www.pmi.org/pmbok-guide-standards/foundational/PMBOK</u>
- [12] INCOSE, The Systems Engineering Body of Knowledge, https://sebokwiki.org/wiki/Guide\_to\_the\_Systems\_Engineering\_Body\_of\_Knowledge\_(SEBoK)
- [13] RDA: GORC International Model Working Group, <u>https://www.rd-alliance.org/groups/gorc-international-model-wg</u>
- [14] RDA TIGER D6.1 Initial Quality Control Processes, 10.5281/zenodo.8016331





### Appendix A: RDA Maintenance Platform Details

### A.1 Core Components

#### A.1.1 Problem Statement

Most organisations struggle with knowledge management, largely due to a set of interrelated problems:

- Knowledge is contained in many types of digital output, distributed in multiple internal and external repositories.
- The semantic context of the knowledge is a combination of facts that are digitally encoded and possibly discoverable, and facts that are in the minds of employees, end users, or collaborators.
- The semantic context depends at least in part on the user: a 'foundational' semantic context often needs to be extended and refined for specialised use cases.

RDA wants to combine the formal organisation and classification of its outputs - done by staff or IG/ WG chairs - with the feedback, annotations, linkages, and context provided by the broader community in RDA and beyond.

#### A.1.2 Concept

Semantic annotation of RDA outputs will be supported in three broad categories:

- formal classification and 'filing' of outputs by staff and workgroup leads based on one or more vocabularies developed by RDA for that purpose,
- linking of outputs to external semantic resources such as vocabularies, registry PIDs, and ontologies for the purpose of integration of RDA outputs into the wider scope of research graphs and linked open data, and
- enabling the structured and unstructured annotation of outputs by end users. The latter can also be used to obtain usability and quality feedback from the user community.

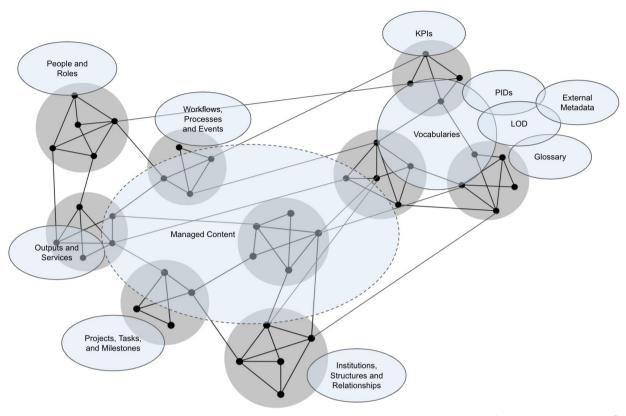
Conceptually, all resources that are of interest to RDA for knowledge management and retention purposes should be viewed as nodes in a graph.

This is important, since it allows that each of these resources can have one or more relationships to any other node. Naming such relationships according to a formal vocabulary is one of the cornerstones of knowledge management technology.

We can also link to or label ('tag') nodes using vocabulary, and in combination with named relations this provides a powerful basis for organising knowledge, and for building tools for it. We can use free-text annotations to label nodes and to describe relations, but these are much less useful.







Typical Example - Not RDA-Specific Figure A.1.2.1 - Vocabularies and 'Managed Content' - such as RDA Outputs

Nodes are generally required for all the concepts and things that RDA works with in executing its tasks and producing its outputs. We call the inventory of all these things and concepts the RDA Vocabulary', but it is actually a virtual collection of vocabulary services, internal and external, registries that manage PIDs, and of services that expose subsets of graph data (such as LOD or PID Graph data) relevant to RDA. Metadata can also be used to create graph content for use by RDA.

Some special subsets or collections of vocabulary can be given additional labels, for instance by defining a glossary or a metadata catalogue as a subset of the RDA Vocabulary. In broad terms, traditional mechanisms for 'filing' objects and resources correspond to one or more hierarchical lists in the 'RDA Vocabulary'.

Some vocabulary contributions come from systems or sources in RDA (the IG and WG structure, participating organisations, ...) and others from external sources such as, for example, ORCID.

#### A.1.3 Additional Criteria

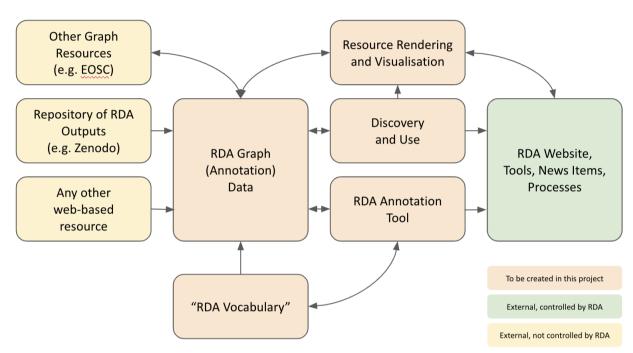
Solutions need to satisfy some practical criteria:



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



- Preferably based on well-supported, community-maintained code that requires little extension initially to be useful;
- Capable of deployment in a variety of environments, such as local servers, containers, and VMs;
- Integrated with popular and recent versions of web browsers, ideally allowing it to be used as a plugin that is accessible across multiple sources of digital content.



#### A.1.4 RDA Maintenance Platform Components

Figure A.1.4.1 RDA Maintenance Platform Components

The RDA Maintenance Platform establishes a number of infrastructure components, as follows:

- 1. A portfolio of '**RDA Vocabularies**', that in practice needs to be hosted in an RDA-managed environment and is maintained by RDA. It is likely that at least some of these vocabularies will be provided and maintained by the RDA website or its successor, but the balance will be hosted by the RDA Maintenance Platform until a decision is made about its hosting and maintenance.
- 2. An **Annotation Tool** links web resources, including RDA outputs, to these vocabularies, and writes the resulting triples into a graph database ('**RDA Graph**'). The RDA Graph can contain any triples that are of interest to and are useful to the RDA community, and we make proposals about its extension in Appendices B and D. The annotation tool can be deployed both as a browser plug-in and as a form in a website (for example the RDA website).
- 3. A search and **Discovery Tool** that supports both search capabilities and presentation of individual search results ('**Research Rendering and Visualisation**'). The discovery interfaces are designed to support multiple entry points, for example by configuring one or more dashboard views that serve as summary views of the content of the underlying graph.



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



The RDA Maintenance Platform is designed to integrate with the RDA website (current, and also its future replacement), and it will use existing infrastructure for archiving and long-term preservation of resources - at the moment, these choices are Zenodo for document and report resources, and the DANS VAult for long-term preservation.

Finally, in all cases where the RDA graph uses existing LOD references of any kind (community vocabularies, registries of resources, services, people, or institutions) it will already be able to provide additional context about the entities referenced in the graph. We discuss this and possible extensions in the next section.

### A.2 Extension to the 'RDA Graph' and ROM-BoK

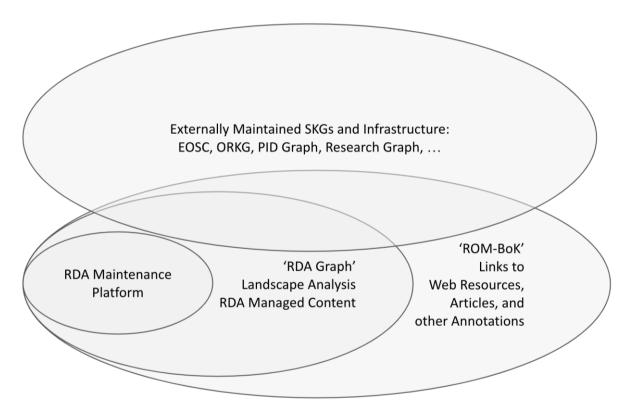


Figure A.2.1: The relative scope of graph-based resources in RDA context

The diagram describes the relationships between different graph-based data collections and the overlaps that may exist. The RDA Maintenance Platform will provide a kernel of graph-based triples that link entities (resources, people, working groups, topics, and other contextual information) to one another, but is only mandated to provide this information for RDA outputs and its supporting resources. It is quite simple to extend the scope of this graph to also include other contextual information and entities of interest to RDA in a wider context, for example how RDA efforts are linked to the features and attributes of GORC, or how it links to EOSC. This extension will be possible in the RDA TIGER project.





A third and even wider perspective is possible by linking RDA entities and concepts to a wider research data or research object context - for example by referencing W3C and ISO specifications where relevant, and by incrementally extending the graph to include community-developed recommendations, best practices, specifications, and guidance - typically these are generated not only by RDA, but also by funded projects (such as EOSC-focused Horizon Europe funding) and other forums (for example the EOSC Task Forces). Discussions with stakeholders in this regard (for example in the FAIRCORE4EOSC, FAIR-IMPACT, and selected EOSC Task Forces) have already started and will shortly be extended to the HE Technical Forum.

#### A.2.1 Design Considerations<sup>5</sup>

The following design considerations apply for a database to support the RDA Maintenance Platform, RDA TIGER landscape analysis, and future work linked to and associated with the landscape analysis:

- 1. **Flexibility**: the landscape analysis results, and any knowledge extraction and organisation resulting from it, needs to be flexible enough to allow a mixture of formal and informal classification, tagging, and contextualisation. The need is driven by three generic use cases:
  - a. RDA itself, who wants to be in a position to classify resources and knowledge elements fairly formally in respect of its own vocabularies and semantics;
  - b. The broader community, who will in principle benefit from the ability to link RDA-related resources and knowledge elements to a broader set of semantics, typically present in Scientific Knowledge Graphs and widely used in the landscape; and
  - c. Individual end users in specific communities, who may want to link and annotate resources formally to their specific semantics, or informally to reflect individual assessments of the resources or knowledge elements.
- 2. **Practical application**: We have, to date, identified some specific applications of the type of data envisaged for the landscape analysis result [5].
- 3. **Machine and Human-Readable:** the outputs and efforts resulting from the landscape analysis should ideally be maximally machine-readable, in addition to being human-readable.
- 4. **Re-use of infrastructure**: The RDA Maintenance Platform Project and support actions in RDA TIGER (WP4) will create and operationalise a platform for the management, classification, and publication of RDA outputs, resources, and supplementary materials. This infrastructure is based on a graph database that contains at least minimum applicable metadata for each type of resource but can already be extended and supplemented by formal and informal 'tagging' of resources. There is significant benefit to be gained if the landscape analysis outputs and results can be linked to this infrastructure, including:
- 5.
- a. Re-use and inheritance of relations that already exist in the graph database;
- b. No need to duplicate any infrastructure;
- c. Leverage the vocabularies, tag lists, and registry sources already identified and operationally available in the RDA Maintenance Platform.
- d. Re-use of machine and human interfaces wherever possible.

<sup>&</sup>lt;sup>5</sup> This section is reproduced with some amendments from RDA TIGER Deliverable D3.1 [5]



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



#### A.2.2 Options for Implementation

The following broad alternatives are available for publication and dissemination of the landscape analysis:

- 1. **Publication of a narrative report**: Such a report serves as a definitive digital object that can be cited and is persistently available, but is not machine-actionable in any meaningful way.
- 2. Using existing scholarly annotation platforms: The <u>ORKG</u> and <u>hypothes.is</u> are examples of existing platforms that can be used to annotate and classify web resources both formally and informally. In both instances, one is not completely assured of the sustainability of these platforms, and as a long-term solution, it may not be completely satisfactory. It is expedient to investigate and implement integration with these platforms if it proves feasible.
- 3. **The RDA Maintenance Platform**: This is designed to store, link, and annotate RDA outputs, resources and supplementary materials using RDA-defined vocabularies and tag lists. There are no practical reasons why knowledge elements and nodes defining the RDA understanding of the landscape cannot be accommodated in the same graph. The longer-term sustainability is no more certain than the options in (2), but it is under RDA control.
- 4. **Hybrid solution**: We consider a hybrid solution, as described below, as a feasible and flexible option that satisfies the design considerations.

#### A.2.3 Hybrid Solution

A hybrid solution will have the following components and features:

- Landscape Ontology: The RDA TIGER WP3 identifies foundational vocabularies and classification trees that adequately describe the landscape (an emerging list of these overlaps with the vocabularies, tag lists and registry references identified for the RDA Maintenance Platform [7]). There is also an overlap with the Global Open Research Commons (GORC) Essential Elements of a research common currently published as Typology and Definitions<sup>6</sup>.
- 2. These vocabularies can be configured as an annotation form for referencing and linking information identified in any web-based resource in essence the typical bibliographic references that will form part of a narrative report. Such an annotation is stored in a graph database with the appropriate vocabulary and tag links, which means that:
  - a. It is machine readable, and
  - b. A reference can be linked to multiple vocabulary elements or tags, and hence can support, assert, elaborate, or contextualise more than one concept in the landscape.
- 3. A narrative report is still useful and required mostly for non-machine purposes but using annotation of snippets in the narrative report, one can also link the report to concepts in the landscape, which will by default also link these report elements to supporting references that have been annotated with the same concept elements.

The RDA Maintenance Platform will only become operational in the second half of 2023, and that will initially not be available to support the landscape analysis effort in RDA TIGER. It is, however, possible and feasible to use tools such as hypothes.is in the interim to annotate web references and resources with vocabulary elements and tags. Converting these annotations to a format that can be imported into the RDA Maintenance Platform Graph later on should not present any technical difficulties.

<sup>&</sup>lt;sup>6</sup> https://www.rd-alliance.org/group/global-open-research-commons-ig/outcomes/gorc-ig-typology-and-definitions



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



#### A.2.4 Conceptual Data Model

The conceptual data model is work in progress but is already broadly representative of the concepts, things, and relations that are needed to properly describe and contextualise the work of RDA.

The conceptual model allows linking any number of curated tag lists (for example the GORC Elements, Features, and Attributes of Data Infrastructure, or EOSC terminology) to the outputs, structures, and vocabulary in use by RDA and included in the RDA Maintenance Platform. On this basis, we will be able to link elements of landscape analysis to any or all of these concepts and entities as required.

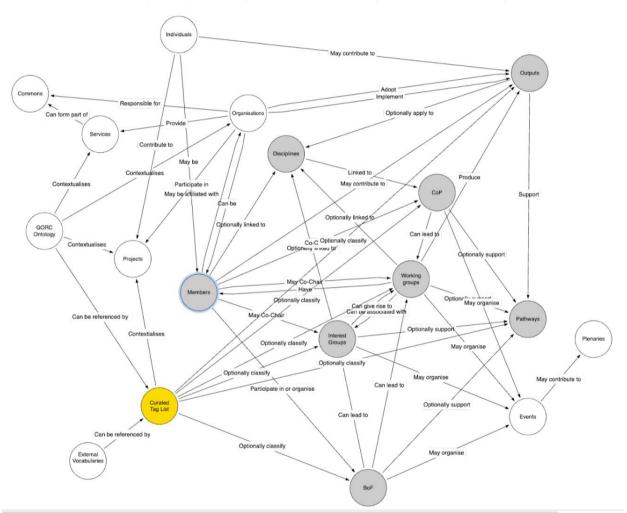


Figure 2.2.4.1: RDA Maintenance Platform Conceptual Data Model





# Appendix B: Metadata Continuum and the 'RDA Graph'

### B.1 The Metadata Continuum

All digital objects that are of interest or use to RDA have to be managed in a consistent way. At a minimum, we need to be able to identify, resolve, and unambiguously reference each digital object. If the objects were created by RDA, there may also be a need to preserve it properly.

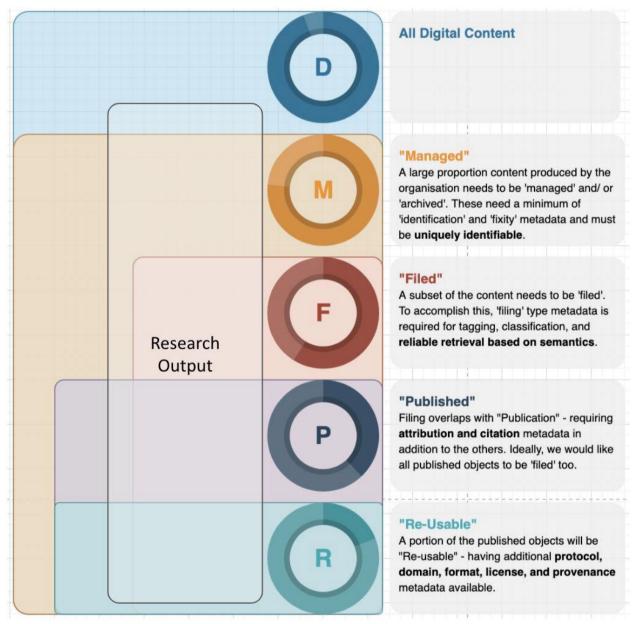


Figure B.1.1: The Metadata Continuum for Managed Content [8]





The digital object does not need to be physically stored by RDA: in fact, it will most likely be a web resource or publication hosted outside the RDA infrastructure, and potentially owned by an external entity, or it can be a document hosted in an RPO document management system or a third-party repository such as Zenodo.

Figure B.1.1 provides a summary of the scope of different sets of managed digital content, and the typical metadata requirements for each of these [8].

- "Managed ": A large proportion of content produced by any organisation needs to be 'managed' and/ or 'archived'. These need a minimum of 'identification' and 'fixity' metadata and must be uniquely identifiable. If external resources have PIDs, the fixity, persistence, and resolvability of the resources are maximised.
- "Filed": A subset of the content needs to be 'filed'. To accomplish this, 'filing' type metadata is required for tagging, classification, and reliable retrieval based on semantics. This latter part depends heavily on agreed vocabularies within RDA, together with mechanisms to manage these vocabularies properly.
- "Published": Filing overlaps with "Publication" requiring attribution and citation metadata in addition to the above. Ideally, we would want all published objects to be 'filed' too.
- "Re-usable": A portion of the published objects will be 'Re-usable' having additional protocol, domain, format, licence, and provenance metadata available.

The above provides a metadata continuum that can be applied across all digital content - including publications and outputs created especially for RDA.

Superimposed on this continuum is a set of structured and unstructured annotations of digital objects in the web. In some cases, if the same vocabulary is used for such annotations, the objects are directly linked to the managed content created for and by RDA, and if the digital objects being annotated coincide with RDA content, it enhances the contextual information available for those objects.

### B.2 The Need for Agreed Vocabularies

It should be clear from the discussion above that we need to be able to reliably refer to objects, but that we also need to be able to reliably link those objects to all kinds of subjects, other objects, and context: this accomplished through controlled vocabularies and registries. This section discusses the scope and application of vocabularies.

Firstly, we need to discuss the types, structures, and sources of vocabularies. We will typically be reliant on the following sources:

- Structure of Vocabularies:
  - Vocabulary can be simple (glossary-like definition lists), or hierarchical/ linked-lists, or a graph (for example the graph of scientific domains of study and their relationships).





- Ontologies are complex vocabularies where the nature of the relationship between entries have been expressly defined, and thesauri are ontologies that support a specific set of lexical relationships.
- Sources of Vocabularies:
  - Internal vocabularies: these are used by no-one else, and RDA must actively manage and agree on these vocabularies.
  - External Vocabularies: these can come from several sources.
  - Global Registries: these include authoritative lists and PIDs of, for example, research organisations, researchers (e.g. ORCID), funders, and so on.
  - Global Repositories: containing PID references to digital content, and optionally the content itself - examples are DOIs provided by DataCite, repository URIs provided by GitHub, etc.
  - Community Vocabulary Services: Many of our definitions derive from community consensus - for example FAIR principles and the attendant criteria. These are often maintained by standards organisations or international bodies on behalf of the community.
  - Linked Open Data: This is a broader, non-community specific resource that, for example, allows machine-based access to knowledge repositories such as Wikipedia, and unambiguously references a topic and its content.
- Types of Vocabularies: There are several subtypes involved.
  - Structures and Classifications: these include examples such as the RDA organisational structure, the roles and responsibilities of personnel and how they report to one another, or are organised in teams, the structure of one or more filing systems, KPIs, and so on. The main purpose of these is to classify things 'belonging to a group'.
  - Processes and workflows: These define the relationships between steps in a process. We usually have different types of formality to these processes:
    - Formal workflows: these are distinguished by a statutory, legal, or business requirement to follow a specific process or decision pathway. Examples output adoption workflows, procurement via tender or open call, and elections for officers of RDA structures.
    - Standard Procedures: these are documented procedures and require more than one working group, interest group, or multiple members to contribute in a fairly well-defined, but flexible process. Examples include solicitation and selection of session proposals for a plenary meeting, completion of monthly and annual reports and KPI assessments, publication of reports, outputs and deliverables, and similar.
    - Ad-Hoc or Informal Workflows: workflows that require multiple points of action across the organisation, but are expedient based on circumstances or context. Examples include a work plan or task list agreed for a working group and published in ots case statement.



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



To make vocabularies work in RDA, we require a mechanism to agree on them. This mechanism is proposed to be a working group-like activity but with permanent mandate and could function as a sub-group of the Technical Advisory Board.

### B.3 The 'RDA Graph'

The RDA Maintenance Platform, as described in Annexure A, will create a minimum version of a graph database with a scope that covers RDA outputs, supplementary materials, and the vocabularies required to properly supply FAIR metadata, and contextualisation in the RDA landscape.

The tools provided by the RDA Maintenance Platform can, however, support a wider scope of vocabularies and links, web-based resources and objects, and contextual information. As an example, the work done in RDA TIGER WP3 (Landscape Analysis) [5] can be annotated using the same tools, but focused on slightly different objects and using a slightly amended vocabulary collection - for example by including GORC-related elements and features as controlled vocabulary. In doing so, one will be in a position to expand the minimum version of the graph to provide connections between RDA outputs and supporting materials, GORC-related resources and landscape assessments, and other elements of externally maintained Scientific Knowledge Graphs (SKGs).

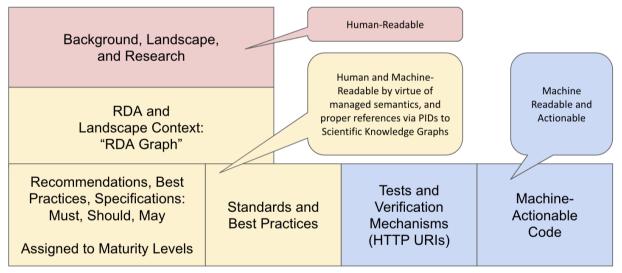


Figure B.3.1: RDA Graph, Machine- and Human-Readable Views

The diagrams below provide some sense of the current scope of vocabularies being developed and populated. The first diagram defines the vocabularies that are 'native' to RDA and will in all likelihood be maintained either in the website or in the RDA Maintenance Platform. The second diagram provides an overview of the additional tag lists (also provided by RDA but more similar to code or tag lists, and more likely to change over time) together with external vocabularies that emerge from the TIGER Landscape Analysis and our engagement with external stakeholders.



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.





Figure B.3.2 - RDA-Owned Vocabularies - Current Scope Blue Tags: First versions of these vocabularies are available for implementation

The GORC Elements, Features, and Attributes are significantly more detailed than shown in Figure B.3.3, and is currently being standardised by TIGER WP4 for use as a vocabulary in its own right.

Similarly, community expectations such as FAIR, CARE, and TRUST have individual criteria associated with principles that will be included in the detailed vocabularies. Preferably, these community expectations vocabulary can and will be maintained externally.

Both diagrams can be accessed in an interactive web application.





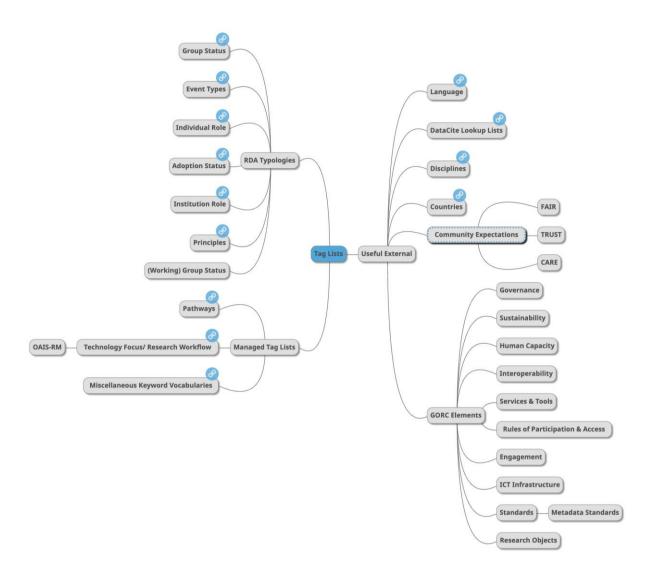


Figure B.3.3 - Vocabularies Linking to the Landscape Blue Tags: First versions of these vocabularies are available for implementation.

# Appendix C: Support Responses and Solutions - Timeline

Section 3 of the document summarises the scope of solution elements that we have identified. This appendix provides additional details about the support responses and solutions that will be developed and implemented in RDA TIGER WP4. These responses address the concerns, requirements, and issues identified in Section 2.

### C.1 Detailed Milestones and Plenary Alignment

In the following table, the assumption is made that the first working groups to be supported by RDA TIGER will be able to receive such support by M9 of the project, and that the first outputs to be accommodated via processes defined by RDA TIGER will be required by M21.





We also need to align with the RDA Plenary cycle: P21 = M9, P22 = M15, etc. Dissemination Opportunities are shown as  $\exists$ 

#	Solution	Description	Milestone(s) Plenaries	Major Dependencies
3.1.1	Policy - Outputs	RDA will benefit from a policy in respect of output production and publication - dealing with aspects such as licensing and access, branding, ownership, attribution, and maintenance.	M24 P24 <b></b> <u></u>	A policy will have to be adopted and endorsed by RDA Secretariat and Board
3.1.2	Procedure: Workgroup Creation and Life Cycle	This exists to some degree, but must be reviewed in the light of additional support materials, guidelines, mechanisms, and tools required for the ' <b>Output Definition Report</b> ', and must align with policy.	M9 M21 P23 <b></b> <u></u>	A procedure will have to be reviewed and agreed by RDA structures - Secretariat, Board, TAB.
3.1.3	Procedure: Research Processes	Procedures for annotation and referencing of resources and publications in the course of the workgroup's activities	M9 P21 <b></b> <u>⊒</u>	No major dependencies
3.1.4	Procedure: Publication of Outputs	RDA working groups need to adopt a common procedure for publication of outputs and supplementary materials, based on the RDA Maintenance Platform.	M12 P22 <b></b> <u>⊒</u>	For working groups supported by RDA TIGER, this is M21, but it will be useful as soon as it is available.
3.1.5	Procedure: Curation of RDA Collections	RDA collections in Zenodo need accession procedures that will ensure consistency of materials and resources in the collections.	M9 P21 <b></b> <u>⊒</u>	This has to be discussed and agreed with the RDA Secretariat and the responsibility defined.
3.1.6	Procedure: Selection and Maintenance of Vocabulary	The process whereby RDA identifies vocabulary and how it is governed and maintained must be documented.	M18 P23 <b></b> <u></u>	
3.2.1	Tag Lists, Registries, and Vocabularies	RDA needs to use standard APIs and encodings for its own vocabularies. These need to extend to and allow the linking of external resources as intended in O1 and O3.	M21 P23 <b></b> <u>⊒</u>	Dependent to some extent on 3.1.6

Table 14. Solution dependencies and alignment





3.2.2	Preferred Formats for Dissemination and Preservation	Theses are needed to maximise utility for the community and to ensure long-term preservation	M9 P21 <b></b> <u>⊒</u>	Not complex, but needs an endorsement and confirmation process in RDA
3.2.3	Annotation Standards and Schema	RDA will define a number of standard schemas: metadata, rating, and landscape context are examples.	M12 P22 <b>⊒</b> M21 P23⊒	Agreement in RDA and RDA TIGER by month 9, revision by community
3.2.4	Minimum Metadata Schema	These have to be differentiated for different output and resource types, while satisfying repository and RDA requirements.	M9 P21 <u></u> M24 P24 <u></u>	Needed earlier than other schemas. Will need formal adoption in RDA by milestone 2
3.2.5	Templates for Output Reports, Recommendations, Standards and Specifications	A number of standardised templates are needed for typical RDA outputs and supporting materials. One objective is to simplify annotation and machine-readability.	M12 P22 <b></b> <u>⊒</u>	No specific dependencies, except engagement with community and desk research
3.3.1	RDA Maintenance Platform	The RDA Maintenance Platform plays a central role in support of WGs and associated RDA activities: publication, annotation, discovery, and preservation. See Appendix A.	M6 P21 <u></u> M12/15 P23 <u></u> झ	Intended release for use by beta testers in M6. Expected integration into new website M12
3.3.2	Reference Managers	Integration between mainstream Reference Managers and the RDA MP will benefit end-users.	M24 P24 <b></b> <u></u>	Not urgent
3.3.3	Annotation Platforms	Integration with annotation platforms such as OKRG and hypothes.is will broaden the scope of end users and extend the RDA graph/ knowledge base.	M15 P22 <u></u> M24 P24 <u></u> ₽24	Dependent on agreed schemas: 3.2.3
3.3.4	Helpdesk	The Helpdesk assists working groups and other RDA members to request assistance with tools, software, templates, or any other solution element	M9 P21 <b></b> <u>⊒</u>	Needs to be available early in the lifetime of the project.
3.3.5	Helpdesk Knowledge Base	The guidance and support materials (below) serve as an entry point to the policies, procedures, standards, tools, and best practices available to working groups. These are made available as HelpDesk knowledge base articles, in addition to being published and annotated.	M9 P21 <u></u> M21 P23	At the latest, full set to be available by M21 First release M9





3.3.6	The 'RDA Graph'	The need to accommodate a wider set of entities, concepts, and relations emanating from RDA within its landscape and ecosystem has to be accommodated: largely by providing access to additional schema, vocabularies and registries as a basis for annotation.	M9 P21 <b></b> <u></u>	Dependent on 3.3.1 (RDA Maintenance Platform) beta release
3.4.1	Guidelines: Workgroup Establishment	<ul> <li>Description of life cycle, which solution elements are available for and should be used in each step of the life cycle, and how to</li> <li>Generate the Output Definition Report;</li> <li>Create outputs and support materials that are publication and machine-ready.</li> <li>Annotate and reference individual elements of outputs in a reusable way</li> </ul>	M9 P21 <b></b> ₫	First iteration in M9, but it has dependencies on many other solution elements: 3.1.1 3.1.2
3.4.2	Guidelines: Maximising Workgroup Effort	Includes landscape analysis support, annotation best practices, structuring of outputs and supplementary materials, making effort reusable - overlaps with the previous, but adds guidelines for annotation of web resources.	M12 P22 <b></b> <u></u>	3.1.3 3.1.4
3.4.3	Documenting Standards and Specifications	It is anticipated that standards and specifications will be occasional outputs from working groups, and to ensure that these are documented in a consistent and publishable way, a template with attendant guidance will be developed for this aspect.	M18 P23 <b></b> <u>⊒</u>	No major internal dependencies, but requires desk research and community engagement.
3.4.4	Standards Authority Adoption Guidelines	Further support of the process of creating and disseminating standards and specifications: guidance will be developed, in consultation with e.g. ISO and W3C to guide working groups on processes to formally adopt their work.	M21 P23 <b></b> <u></u>	3.4.3 is a prerequisite Dependency on successful engagement with ISO and/ or W3C
3.4.5	Informal Contributions to the 'RDA Graph'	It is possible and desirable for individual RDA members to annotate, rate, and comment materials referenced and produced by RDA. To assist with this, a guidance document will be developed.	M15 P22 <b></b> <u></u>	Must be available to coincide with Plenary 22





3.5.1	Workshops: Workgroup Output Support	Training materials and plenary- aligned workshops to highlight and introduce the support mechanisms developed by RDA TIGER	M9 P21 <b></b> <u></u>	First workshops for TIGER-supported WGs at Plenary 21
3.5.2	Service Assistance: Curation	Curation activities for Working Groups to publish their efforts and outputs in the recommended way.	M21 P23 <b></b> <u></u>	First workshops for TIGER-supported WGs at Plenary 23

### C.2 Plenary Targets

Solution elements assigned to Plenaries:

Plenary	Available Solution Elements		
P21	<ul> <li>3.1.3 Procedure: Research Processes</li> <li>3.1.5 Procedure: Curation of RDA Collections</li> <li>3.2.2 Preferred Formats for Dissemination and Preservation</li> <li>3.2.4 Minimum Metadata Schema</li> <li>3.3.1 RDA Maintenance Platform</li> <li>3.3.4 Helpdesk</li> <li>3.3.5 Helpdesk Knowledge Base</li> <li>3.3.6 The 'RDA Graph'</li> <li>3.4.1 Guidelines: Workgroup Establishment</li> <li>3.5.1 Workshops: Workgroup Output Support</li> </ul>		
P22	<ul><li>3.3.3 Annotation Platforms</li><li>3.4.2 Guidelines: Maximising Workgroup Effort</li><li>3.4.5 Informal Contributions to the 'RDA Graph'</li></ul>		
P23	<ul> <li>3.1.2 Procedure: Workgroup Creation and Life Cycle</li> <li>3.1.6 Procedure: Selection and Maintenance of Vocabulary</li> <li>3.2.1 Tag Lists, Registries, and Vocabularies</li> <li>3.2.3 Annotation Standards and Schema</li> <li>3.3.1 RDA Maintenance Platform</li> <li>3.3.5 Helpdesk Knowledge Base</li> <li>3.4.3 Documenting Standards and Specifications</li> <li>3.4.4 Standards Authority Adoption Guidelines</li> <li>3.5.2 Service Assistance: Curation</li> </ul>		
P24	<ul> <li>3.1.1 Policy - Outputs</li> <li>3.2.4 Minimum Metadata Schema</li> <li>3.3.2 Reference Managers Integrated</li> <li>3.3.3 Annotation Platforms Integrated</li> </ul>		

Table 15. Solution elements assigned to Plenaries.

Note: some solution elements are listed more than once due to multiple releases of the same element.





## Appendix D: ROM-BoK

### D.1 Context

There is a well-established precedent for communities in the engineering domain to create and maintain 'Bodies of Knowledge': we present three such examples that could serve as a basis for a similar community effort emerging from RDA and ReSA activities.

- The Project Management Body of Knowledge: PMBOK [11]: "A Guide to the Project Management Body of Knowledge (PMBOK<sup>®</sup> Guide) is [the] PMI's flagship publication and is a fundamental resource for effective project management in any industry. Over the years, business has changed considerably, but projects remain critical drivers of business success."
- 2. The Software Engineering Body of Knowledge: SWEBOK [10]: "The Guide to the Software Engineering Body of Knowledge (SWEBOK) describes generally accepted knowledge about software engineering. Its 15 knowledge areas (KAs) summarise key concepts and include a reference list for detailed information."
- 3. The Systems Engineering Body of Knowledge: **SEBOK** [12]: "The SEBoK provides a guide to the key knowledge sources and references of systems engineering organised and explained to assist a wide variety of individuals. It is a living product, accepting community input continuously, with regular refreshes and updates. The SEBoK is not a compendium but instead references existing literature."

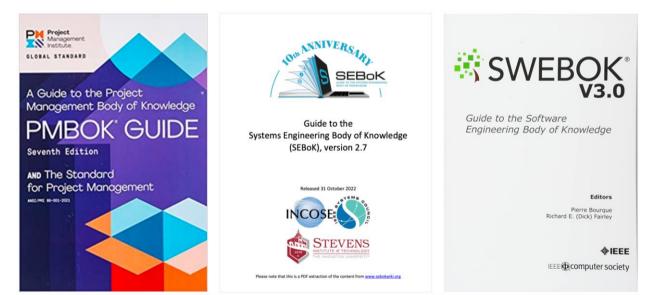


Figure D.1.1 Examples of Bodies of Knowledge



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



We believe that a similar effort is required and will provide benefits to the community active in RDA and its broad ecosystem. Existing initiatives in RDA, for example GORC [13], can serve as an important effort and contribution to standardise and describe knowledge areas in the research data management domain.

### D.2 Concept

ROM-BoK can be seen as an ongoing process of standardisation of resources and materials available to the community, based on agreed categorisations and definition of 'Knowledge Areas' - essentially a set of classifications and vocabularies that can be used to properly contextualise any relevant resource.

There are two levels of granularity to consider for standardisation:

- 1. **Resource-Level**: this entails the annotation, linking, and metadata schema-based publication of web resources, including RDA outputs. Reusability depends on the extent to which web resources are provided with persistent identifiers, but in its absence, URI references will have to be used.
- 2. Element-Level: within web resources, specific elements or assertions are of interest. These could be elements such as individual recommendations, best practices, principles, criteria, and so on. At present, these elements are not optimally structured for annotation and re-use, although annotation is already possible using tools such as ORKG and hypothes.is. The RDA Maintenance Platform (Appendix A) will build on both these tools to make element-level annotation and structured linking simpler. Coupled to structuring of document resources in such a way as to support annotation, element-level resources can be mobilised and become a powerful cross-cutting body of knowledge.

The RDA Maintenance Platform will establish the basic infrastructure (graph database platform, annotation and linking tools, search and discovery interfaces) required to support a wider implementation such as foreseen for the ROM-BoK. In addition, RDA TIGER is in a position to contribute to this resource by structuring its outputs and efforts to align with the RDA Maintenance Platform.

The list below provides a partial overview of projects and initiatives that can contribute to the ROM-BoK if mobilised effectively:

- Ongoing projects, funding, and initiatives
  - RDA Maintenance Platform (Creates Vocabulary, Data, and Infrastructure)
  - FC4E Compliance Assessment Toolkit (Creates Vocabulary, Data, and Infrastructure)
  - FAIR-IMPACT/ FC4E/ EOSC TF Collaboration on PID Policy (Creates Vocabulary and Data, provides an example or pilot of HE project integration)
  - o RDA TIGER Landscape Assessment WP3 (Creates Vocabulary and Data)
  - EOSC TF Technical Interoperability (Creates Vocabulary and Data, provides and example or pilot of task force integration)
  - RDA TIGER Output Support WP4 (Creates tools, guidance, procedures, and standards to integrate RDA Workgroup effort and outputs into the ROM-BoK)



RDA TIGER "Research Data Alliance facilitation of Targeted International working Groups for EOSC-related Research solutions" is funded by HORIZON-INFRA-2022-EOSC-0 Project: 101094406.



- The RDA GORC International Model [13] Vocabulary serves as a basis for broad knowledge organisation
- Serves as Examples/ Creates Tooling and Guidance For
  - FAIR-IMPACT and FAIR's FAIR Recommendations, Guidance, Best Practice
  - FC4E Vocabularies, Specifications
  - Other Horizon Europe-funded projects, especially in support of EOSC.
  - EOSC Task Force Outputs
  - RDA Outputs Detailed Analysis

The diagram below illustrates the scope of potential contributions in the ecosystem that can be aligned in support of the ROM-BoK.

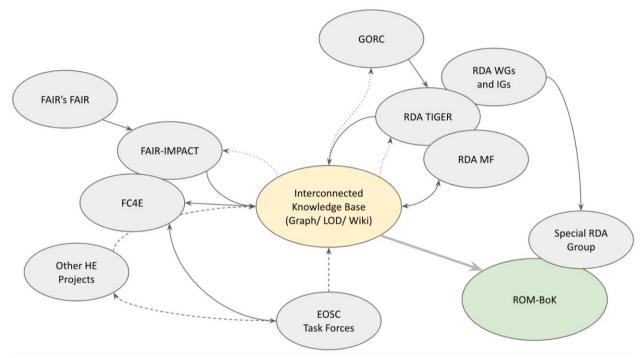


Figure D.2.1 Leveraging efforts in the ecosystem to contribute to ROM-BoK

It is proposed that RDA considers a special and permanent working group to serve as an implementing and governance structure for the ROM-BoK: such a group will be required, albeit with a narrower mandate, to govern and manage the RDA Maintenance Platform and its vocabularies. Looking after ROM-BoK and the RDA contributions to it is a natural extension of this mandate.

#### D.3 Outcome

The following positive outcomes are envisaged for the ROM-BoK:

• Create a machine-actionable body of knowledge that aggregates and links contributions from major regional and global projects, RDA efforts, networks, and infrastructures.





- Simplifies discovery and inclusion of guidance, best practice, recommendations, and specifications into training materials, applications, and services.
- Assists with alignment and deduplication of vocabularies and terminology in use by the community.
- Living resource that can be published from time to time as a snapshot of the ROM-BoK.

### **D.4 Implementation**

The following important milestones and opportunities need to be taken advantage of in respect of both the activities of RDA TIGER WP4, and the possible implementation of the ROM-BoK:

- Session at International Data Week (SciDataCon), October 2023: A session proposal has been submitted for this event, and RDA TIGER will have an opportunity to share ideas and invite community members and stakeholders to contribute (Session proposal to be submitted).
- **BoF at RDA Plenary 21 in Salzburg, October 2023**: Propose the formation of a special RDA group to analyse RDA outputs and identify/ govern vocabularies and tag lists, and to foster increased use of the tools and facilities available for linking web resources into the RDA Graph and the Body of Knowledge (Session proposal submitted).
- Grant Funding Proposal HORIZON-INFRA-2024-EOSC-01-02: Supporting the EOSC Partnership in further consolidating the coordination and sustainability of the EOSC ecosystem - creating mechanisms of mutual learning, replication of best practices and joint activities. "This expected outcome will be achieved mainly through financial support to third parties in the form of cascading grants" (applicable consortium and contribution to be investigated).

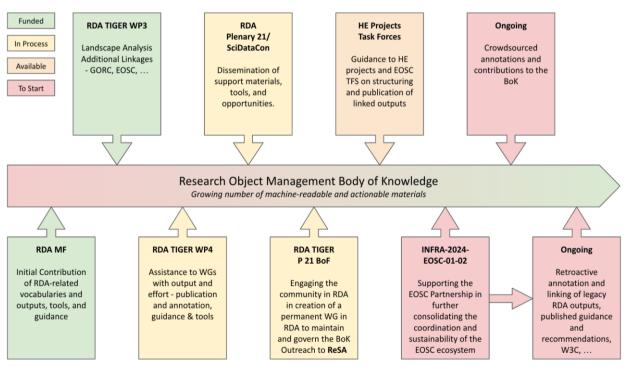


Figure D.4.1: Timeline of Important Contributions to ROM-BoK.





- Contributions from **RDA TIGER WP3**, **WP4**, and the **RDA Maintenance Platform** will be directly usable and are already funded.
- It is hoped that some momentum can be built within RDA and the ecosystem (HE-funded EOSC projects, the task forces, and similar initiatives) to contribute to the ROM-BoK via crowdsourcing, and an increasing adoption of mechanisms and procedures created by RDA TIGER WP4 in everyday applied research activity.
- If a permanent working group can be established in RDA, and especially if funding can be obtained via HE, there is a real **prospect of annotating and properly structuring legacy RDA outputs**, as well as resources created by ongoing and previous HE-funded projects.

