

RDA Europe Task Force on Sustainability

Pre-final Report

|  |  |
| --- | --- |
| Lead Author | Fotis Karayannis |
| Version | 1.0  |
| Version info | Pre-final version; based on RDA EU BoD, Task Force feedback & RDA Europe 4 inputs |
| Date | 23 February 2018 |
| Dissemination level | Restricted to RDA Europe/TF members – RDA Governance Bodies |
| Task Force members | Juan Bicarregui, Timea Biro\*, Sandra Collins, Ingrid Dillo\*, Fabrizio Gagliardi, Hilary Hanahoe\*, Leif Laaksonen, Mark Parsons\*, Raphael Ritz, Peter Wittenburg and Fotis Karayannis (Chair)\*Mark Parsons until June 2017, replaced by Ingrid Dillo; Hilary Hanahoe until May 2017, replaced by Timea Biro*Input received from the RDA Europe Board of Directors including RDA Council EU Members and TF meetings*  |

**ABSTRACT**

This document summarises the output of the RDA Europe Task Force on Sustainability, assessing the status of RDA Global sustainability from the RDA Europe perspective in terms of funding-financing and governance, analysing the overall landscape in which RDA & RDA Europe are operating, and proposing a set of recommendations for the future.

 RDA Europe receives funding from the European Union Horizon 2020 Programme (H2020/2014-2020) under grant agreement n° 653194.

Table of Contents

[Executive Summary 3](#_Toc499263616)

[1. Introduction 5](#_Toc499263617)

[1.1 Background 5](#_Toc499263618)

[1.2 Objective 5](#_Toc499263619)

[1.3 Target Audiences 5](#_Toc499263620)

[1.4 Definitions and related work 6](#_Toc499263621)

[2. RDA and the data landscape 7](#_Toc499263622)

[2.1 RDA in need of evolution 7](#_Toc499263623)

[2.2 Global landscape 7](#_Toc499263624)

[2.3 Europe 10](#_Toc499263625)

[2.3.1 RDA in Europe: National vs. European approaches 13](#_Toc499263626)

[2.3.2 RDA Europe 4 project 15](#_Toc499263627)

[3. RDA Global sustainability challenges 18](#_Toc499263628)

[3.1. Governance structures and processes 18](#_Toc499263629)

[3.2 RDA Core Activities and their Funding-Financing 19](#_Toc499263630)

[4. RDA value proposition and assets –Proposed directions 23](#_Toc499263631)

[4.1 RDA value proposition and its assets 23](#_Toc499263632)

[4.2 SWOT analysis 24](#_Toc499263633)

[4.3 Sustainability elements and potential directions 25](#_Toc499263634)

[4.3.1 Looking for Comparisons – Well-packaged -marketed messages- Services and new RDA business models? 26](#_Toc499263635)

[4.3.2 Engagement with Industry? 27](#_Toc499263636)

[4.3.3 Testbeds and EOSC 28](#_Toc499263637)

[4.3.4 RDA continues with public funding for funders to remove silos? 29](#_Toc499263638)

[4.3.5 The role of RDA individual members’ organisations 29](#_Toc499263639)

[5. Conclusions and Recommendations 30](#_Toc499263640)

# Executive Summary

This document is the outcome of the RDA Europe Task Force on Sustainability that was created following the first RDA Europe 3 project review by independent experts. The aim of the Task Force and the document outlined in section 1 is to contribute to the **assessment of** **RDA Global sustainability,** mainlyfrom the RDA Europe perspective. **The document is an internal document**, mainly targeted towards its funders and the RDA Governance, including the RDA Europe ones. It will be considered in the future whether a public version of the document needs to be developed.

In section 2, a summary analysis of the **overall context and the data landscape around RDA** is provided, at both the global and European levels. Emphasis is also given in the developments of “RDA in Europe”, including the national vs. European approaches. The document also summarises the main points of the RDA Europe 4 project approach on sustainability and the related outcome of the discussions of the Task Force members.

In section 3, **the RDA Global sustainability challenges, in terms of governance structures and processes, but also in terms of finances**, are presented. These include among others the limited resources of RDA TAB and Secretariat, who need to cope with the growing RDA community and the corresponding growth of the RDA Working and Interest Groups. The challenges of the RDA Foundation are also outlined. The RDA core activities to sustain RDA along with a summary of RDA Global budgets, current and future (desirable) are also summarised. The available future budgets for the next years and especially the income streams after 2020 are not yet included, and this may be attempted in the next versions with the support of RDA Global governance.

In section 4, **the RDA value proposition, along with main “exploitable assets”** are presented and evaluated, along with a Strengths/Weaknesses/Opportunities/Threats (SWOT) analysis. The sustainability assets and potential directions/scenarios are then described.

The document closes with an initial set of directions and recommendation areas, along with a set of recommendations. As the environment around and inside RDA is dynamic, **the document should be considered as a snapshot at the end of the RDA Europe 3 project, namely February 2018.** The document was shared with the RDA Council and Secretariat for feedback but also in an effort to feed into a similar exercise that the RDA Council is working on, developing a strategy for scaling and growth management. A version of the document was also also communicated to the European Commission around mid-December.

Some main points and conclusions from the document taken from section 5, are also presented in this section. The RDA value proposition and its exploitable assets that can make RDA sustainable have been debated and the most prevalent ones have been recorded. In brief, these include:

* **The RDA Community: human network and bottom-up / neutral platform**
* **RDA Outputs/Recommendations and EU ICT technical specifications**
* **Services from RDA outputs and related test-beds**
* **RDA Governance as policy makers for the data commons – Engagement with EOSC**
* **Training and consulting, especially for industrial entities/users**

RDA Europe 4 has also proposed a financial viability plan, including several measures:

* **Light RDA certification schemes, i.e. performing accreditation of compliance for adopters of RDA Recommendations and ICT technical specifications for a fee.**
* **Royalties from potential services related to RDA outputs or partners.**
* **Research and Innovation Grants, from proposals submitted to the next Framework Programme and to other national / regional funding schemes related to RDA and its outputs. Project like StandICT.eu can provide grants for data specialists working on standards.**
* **Contributions from Member States, in particular the national/regional nodes.**
* **Donations from Organisations and Individuals (for the latter à la Wikipedia)**
* **Trainings and workshops for data scientists and other professionals from industry**

Most of the RDA assets are not directly exploitable (i.e. that can be monetised easily), such as the community or the policy making. Other ones, which seem more directly exploitable, such as the RDA recommendations / outputs and ICT Technical Specifications and the RDA-related services require substantial investigation, in some cases even feasibility studies and business/delivery models. The same is true for some of the concrete RDA Europe 4 financial viability measures, such as the certification schemes, the royalties from potential services, the national contributions and donations.

**The Task Force members believe that a stream of funding via grants provided by the main regional funders, including the European Commission, is currently essential to sustain RDA. In Europe, a GÉANT like model, possibly via operational grants, or the engagement of RDA in the European Open Science Cloud with a key facilitator role in European and national data/e-Infrastructure coordination can contribute towards the sustainability of RDA. To keep RDA relevant in the long term for its community, a dynamic approach is needed, capturing the changing community requirements over time via frequent community coordination exercises and feedback consultations.**

At the time of submission of this report, a timely addition that involves RDA and reinforces the efforts already taken place in the ICT technical specifications, directly has been communicated from European Commission which has recommended that the newly funded EOSC-Hub and OpenAIRE-Advance H2020 projects jointly cooperate actively with RDA and already begin finding appropriate working groups or interest groups to channel their challenges they face that may be further discussed during RDA Plenary 11 in Berlin (March 2018). As recently as January 2018, the External Board of the joint collaboration agreement between EOSC-Hub & OpenAIRE-Advance mandated under DG Connect Excellence in Science have written in their consolidated assessment (period 15/12/2017-09/01/2018) that “**the *External board supports the Commission’s proposal that EOSC-Hub & OpenAIRE-Advance start cooperating with RDA and lead the definition of ICT technical specifications for future referencing in public procurement****.”*

An set of recommendation areas, along with **recommendations** is attempted at the end of section 5. Main areas of recommendations include the RDA Governance, RDA finances, RDA engagement and possibly the RDA scope itself. These include:

* The evaluation of RDA Governance, and in particular the establishment of a formal RDA Funders body (with only the ones who provide actual funding to RDA)
* The development of strong national nodes, who can play a key role in national developments (in synergy or integrated with EOSC national nodes in Europe)
* The recruitment of professional effort for business development and marketing, in particular around RDA business models and marketing of outputs.
* The further elaboration of a set of RDA Global budgets for the short and long term, with the help of experts
* The engagement of RDA with key stakeholders, namely funding agencies (see also above recommendation on formal funders’ body), RDA individual members’ organisations, targeted industrial sectors and big research organisations.
* Finally, RDA may consider expanding its scope to cover other areas around big data and public data (Public Sector Information-PSI), to better market itself and exploit the data economy around these

# Introduction

## Background

During the first RDA-Europe 3.0 informal review, the reviewers came up with a set of recommendations, the first of which was about the RDA Europe perspective on RDA (Global) sustainability.

The formulation of the recommendation was that *“RDA Europe 3 should develop short-term (2 to 5 years) and longer-term sustainability goals that will make RDA a centrepiece of European strategy for open science”*.

The rationale and the main actions are described below:

* *Rationale: As data-intensive research continues to expand in scale across disciplines and countries, knowledge infrastructures must expand in scale and sophistication to keep Europe competitive.*
* *RDA efforts will continue to be valuable for all research communities. RDA Europe 3 members are in the best position to* ***design a sustainability plan****.*
* *The financial situation of RDA is currently complex to assess. Establishing a comprehensive* ***budget*** *for RDA global and RDA Europe is now necessary.*

## Objective

The main objectives of the Task Force are:

* to contribute to RDA Global sustainability assessment from the RDA Europe perspective, including provisions for governance, funding and fiscal policies.
* to determine the place of RDA in the European infrastructure and data landscape

## Target Audiences

The document will mainly be targeted to the main stakeholders below, so that they can be assisted in their related actions, including the common aim towards RDA global sustainability:

1. Funding agencies, policy and decision makers.
	1. RDA Europe funders, namely the European Commission and national funders
	2. RDA Council and related RDA Council Subcommittees including the one on Sustainability and Funding
	3. The RDA Funders Forum, and in particular the funding agencies contributing to the RDA Global budget
2. Other relevant stakeholders
	1. RDA Europe 3 and RDA Europe 4 project partners, and other stakeholders involved in the next RDA Europe phases.

The goal within RDA EU 3 has been to write a concise document containing major assertions and not a comprehensive document. As the document contains budgeting / funding information, it is not a public document, however it will be considered whether a public version will be created removing sensitive information.

## Definitions and related work

In the business context, sustainability is defined as the ability to maintain an activity over the long term[[1]](#footnote-2). **For RDA, sustainability entails the components of financial viability and scalable development[[2]](#footnote-3), provided that its mission continues to be relevant/pertinent for its community and stakeholders.** Thepoint of scalable development has been undertaken by the **RDA Council**, which has developed a strategy for scaling and growth management (including reviewing hierarchical, distributed and regional models). Underpinning the new RDA strategy will include the recognised global Unique Selling Points (USPs) for RDA, i.e. being a neutral, independent, open and global forum to address data sharing. The **RDA Council Strategy Subcommittee** has provided elements for the strategy, and in addition, the **RDA Council Sustainability and Funding Subcommittee** has also addressed global sustainability. The RDA Europe 3 Task Force worked closely with RDA Global, especially via the interim Secretary General in 2017. This document provides the RDA Europe point of view, and reciprocally, it has also provided input to RDA Council work, especially given that an early draft of this document has been shared before the Council’s work output.

The Task Force recommendations are also broadly aligned with the directions of the RDA Europe 4, namely the development of national nodes and the strengthening of relations with key industrial players, as well as some key research initiatives, projects or fora including EOSC and ESFRIs (see sections 2.3 and 2.3.2).

It is noted that the RDA global budget currency has been agreed inside RDA to be the US dollar (USD), and the Organisational Member fees are also expressed in USD. The RDA Foundation budget (which is only a part of the RDA global budget) on the other hand is worked out in GB pounds (GBP) as the Foundation is a UK entity. Given the fact that the document deals with the RDA Global sustainability and its budget, it thus adopts the RDA Global decision on budgeting in USD, and does not use Euros or any other currencies.

# RDA and the data landscape

## RDA in need of evolution

Since its beginning in 2013, RDA has been working as a community-driven global forum that facilitates disciplinary and multidisciplinary research data interoperability and data re-use. When it comes to cross-disciplinary activities, RDA has been one of the key drivers for the development of "**building blocks**" of common data infrastructures and "**data bridges**" for reducing data complexity and enabling data sharing. RDA is already in a unique position to synchronise between the geographical, disciplinary and sector stakeholders to support the transition towards FAIR data through the RDA outputs and recommendations. The first RDA recommendations were acknowledged by the European Multi Stakeholder Platform (MSP) on Information and Communications Technologies (ICT) Standarsidation[[3]](#footnote-4) as ICT Technical Specifications in July 2017[[4]](#footnote-5), while a second batch was approved in 2017 and is expected to be officially published in 2018. The RDA approach is focused on overcoming concrete barriers and building technical and sociological bridges using a growing global forum.

RDA has a grass-roots, inclusive approach covering all data lifecycle stages, engaging data producers, users and stewards, addressing data exchange, processing, and storage. It succeeded in creating the **neutral social platform** where international research data experts meet to exchange views and to agree on topics including social hurdles on data sharing, education and training challenges, data management plans and certification of data repositories, disciplinary and interdisciplinary interoperability, as well as technological aspects. This platform covers as of February 2018 135 countries with over 6600 individual members, of which 50% are European.

On the other hand, after more than four years of operations, RDA is facing considerable challenges. The number of Working Groups (WG) and Interest Groups (IG), which organise the RDA grass-root activities, has grown considerably (it is expected to celebrate 100 groups at the RDA 11 Plenary in March 2018) and the support provided by RDA is starting to reach its limits in several areas, especially in the Secretariat and Technical Advisory Board (TAB). Thus, more funded and professional staff is required. Then, following the start-up phase, RDA needs to refine its mission and stay relevant for the whole society around it, of which key components are the data community and its funders. The RDA outputs span multiple layers of disciplines, geographies and sectors, and it may take a long time to create a coherent usable “grid” of specifications. Still, individual adoption cases show great interest in many of the areas and there are good examples of WG results linking with the relevant organisations and gradually rooting and weaving through to sustainable paths, for instance in the agriculture domain. But again, more professional effort from the RDA is needed for this to happen, to sustain and market the outputs.

RDA needs to work towards the next step, otherwise there is a clear risk of fading away. A single way ahead is not straightforward and this document tries to analyse potential paths.

## Global landscape

This section includes a summary analysis of the overall context and the data landscape at global level. The analysis does not claim to be complete, nor comprehensive, and rather focuses on the players around RDA. The document reuses some material from the UK Open Research Data Taskforce landscape report on Research Data Infrastructures in the UK[[5]](#footnote-6). The latter report provides a rather complete landscape analysis and can be consulted for detailed information around data policies and services, along with a set of technical, social and organisational issues.

[CODATA](http://www.codata.org/about-codata) is an organisation of the International Council for Science (ICSU). CODATA advocates data sharing via “policy, technological and cultural changes necessary to make research data more widely available”. It sets up committees and groups dealing with data issues around guidelines, education and training and coordination of data projects. RDA and CODATA work closely in many areas and there are several joint CODATA/RDA groups. CODATA is an affiliate member of RDA and RDA is also represented in the CODATA governing structures. RDA expressed the need for “building bridges” to enable both data sharing and re-use. CODATA is focusing more on global policy-level activities, while RDA is more centred on organising the discussion between data professionals that work in the various initiatives as data scientists, data managers, librarians, etc., to overcome concrete barriers on the way to FAIR data and beyond. Furthermore, CODATA is organised in a top-down process to achieve its goals, while RDA is relying on a bottom-up process. Overall, there is good complementarity, collaboration and synergies among the two bodies.

ICSU’s [World Data System](https://www.icsu-wds.org/organization/intro-to-wds) (WDS) is the evolution of the World Data Centre and is a system of systems composed of accredited data centres and network members. It advocates the curation, access and long-term preservation of scientific data and services across disciplines. It provides the accreditation for its data centres and a gateway service for access the data. WDS is another affiliate RDA members with cross-representation in each other’s governance structures. There are also joint groups with RDA on common areas, in particular around certification and registration of repositories and data publication. The Data Seal of Approval (DSA) and the Wold Data System (WDS) used RDA as an umbrella to join forces towards a global certification standard for repositories, which is now called CoreTrustSeal. Other relevant players in the area, are the [US DataBib](file:///C%3A%5CDisk%20D%5CDocuments%5CMyDocs%5CProjects%5CRDA-Europe-iCORDI%5CRDA%20Europe%203.0%5CProject%5CTF%20on%20Sustainability%5Cdatabib.org) initiative and the [re3data](http://www.re3data.org/) repository initiated in Germany. They joined forces under the RDA umbrella and the auspices of DataCite to create a registry of repositories that also provides information about trustworthiness. Key players in RDA are now busy to study how certification may need to be adapted to fit with all FAIR principles.

[DataCite](https://www.datacite.org/mission.html) is a global organisation providing persistent identification for research data aiming at helping the research community to “locate, identify, and cite research data with confidence”. It develops search facilities and policies to promote data citation. There are worldwide partners responsible for allocating DOIs at the regional level. DataCite is another RDA affiliate member and there are efforts towards more concrete collaboration in RDA based on a Memorandum of Understanding (MoU).

[ORCID](https://orcid.org/) provides an open registry of unique, persistent identifiers for researchers and works with the research community to ensure these identifiers are embedded in research systems and workflows. ORCID is an open, independent, non-profit, community-driven organisation, working across disciplines, research sectors and national boundaries. ORCID identifiers are being integrated into university faculty information systems, manuscript submissions, grant applications, and association management systems, among others. ORCID is another RDA affiliated member and RDA is an ORCID partner working towards the creation of "interoperable data infrastructures through adoption and use of trusted persistent identifiers and standard vocabularies and record formats to promote data quality in the collection, management, exchange and aggregation of research information".

The [FORCE11](https://www.force11.org/) group should be also highlighted as they came up with the excellently formulated “FAIR” principles. At least at the level of principles it seems that "FAIR" established a common language with broad acceptance, to describe principles already in use in some communities.

The [Internet Engineering Task Force](https://www.ietf.org/) (IETF) deals with the Internet standards. Although not directly relevant to RDA, it can provide an inspiring paradigm for RDA sustainability. The Internet industry got heavily engaged in IETF around protocols for networking and other equipment, and it can constitute a path for RDA of how it can organise itself to address its challenges. In particular, the example of IETF “Area Directors” is being discussed inside RDA as possible way of coordination of WG/IG activities, providing guidance and complementing the bottom-up approach of RDA. IETF however cannot be an absolute reference for RDA, since dealing with data is less predominantly technological than the IETF domain.

The [World Wide Web Consortium](http://www.w3.org/) (W3C) deals with web standards. It is composed of member organisations paying contributions and has technical board and experts, meeting with high frequency, to discuss web and data standards. Besides the initial web and semantic web specifications, W3C has expanded towards “Data on the Web Best Practice”, schema.org and other specifications, developing standards and schemas for structured data. The review and evaluation approaches of W3C standards have been discussed inside RDA, yet there is no consensus on taking such a path fully, nor funding to support the evaluation of RDA outputs/recommendations. However, the RDA adopted Request for Comments mechanisms for Group proposals and proposed recommendations, and its processes have been accepted in the ICT specification review.

Other RDA Affiliate members, working at global level and having a complementary scope with RDA include the [Consortia Advancing Standards in Research Administration](http://casrai.org/) (CASRAI) and the [International Council for Scientific and Technical Information](http://www.icsti.org/) (ICSTI). CASRAI it is a non-profit standards development organisation that has built a mostly Canada-UK community of leading research funders and institutions collaborating to ensure seamless interoperability of research administration information. ICSTI it is an international membership organisation working with stakeholders in all disciplines engaged in the scientific communication process with the aim of improving the efficiency of scientific research, without political or commercial agendas.

RDA has also established MoUs in disciplinary areas, such as Earth Observation and Agriculture & Nutrition. Two RDA affiliate members in these areas are the [Group on Earth Observations](https://www.earthobservations.org/index.php) (GEO) and the [Global Initiative for Open Data in Agriculture and Nutrition](http://www.godan.info/) (GODAN). GEO is coordinating efforts to build a Global Earth Observations System of Systems (GEOSS), while GODAN seeks to support global efforts to make agricultural and nutrition-related data available, accessible, and usable for unrestricted use worldwide, advocating towards open data.

Collaborations with other global organisations and initiatives are being pursued. In particular, the [Institute of Electrical and Electronics Engineers](https://www.ieee.org/index.html) (IEEE) needs to be mentioned, with emphasis on big data governance and metadata management. IEEE is a not-for-profit corporation based in the US and is the world’s largest engineering organisation, with over 420,000 members in more than 160 countries. It focuses on advancing the theory and practice of electrical technologies, electronics, computer engineering, computer science, and other related technologies.

[International Telecommunications Union](http://www.itu.int/en/Pages/default.aspx) is the United Nations specialised agency for Information and Communications Technologies and is focussing on global needs and standards. Recently created working groups extended the scope of the ITU work on Internet of Things in Study Group 20 of which data management is an important part. The interactions have indicated that there is a large interest of ITU to discuss issues such as global identification of digital objects and standards around this to make PIDs globally interoperable as shown by the ITU X.1255 standard. There is a great interest in ITU to come to a formal MoU with RDA.

[The DONA Foundation](https://www.dona.net/) has been recently established as a Swiss Foundation to guide the future of the globally available Handle System as part of the Digital Object Architecture. Global persistent identification is a core pillar of the data landscape as contained in the FAIR principles. The DONA work has many relationships to several groups in RDA that deal with digital objects and persistent identifiers and there is a continuous cross-fertilisation between RDA and DONA.

The Open Archives Initiative (OAI) develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. It is producing a number of proposals for best practices which have been widely used such as the Object Reuse and Exchange (OAI-ORE) and the OAI-Protocol for Metadata Harvesting (OAI-PMH). OAI-ORE defines standards for the description and exchange of aggregations of Web resources and OAI PMH is used by many repositories to offer their metadata sets in a standard way to metadata services providers. OAI projects are relevant and being discussed into several RDA WG/IGs.

RDA Europe members have also been asked to join expert groups and working groups in other major umbrella organisations also turning their attention to the tasks of data interoperability, accessibility and re-use. The [OECD Global Science Forum](http://www.oecd.org/sti/sci-tech/oecdglobalscienceforum.htm) in conjunction with CODATA created an expert group on [Business Models for Sustainable Data Repositories](http://www.oecd-ilibrary.org/docserver/download/302b12bb-en.pdf?expires=1516885781&id=id&accname=guest&checksum=EB344086646DC44E978BF0D2D76D5B79), which built on the work of a previous RDA-WDS Working Group on Cost Recovery for Data Centres.

The [ALLEA (All European Academies)](http://www.allea.org/working-groups/overview/working-group-e-humanities/) E-Humanities Working Group, also dedicated to data interoperability, accessibility, and change, has developed a new programme of work that promotes the adoption of RDA outputs across Academies.

## Europe

This section includes a summary analysis of the overall context and the data landscape at European level. Once again, the analysis does not claim to be complete, nor comprehensive, and rather focus on the players around RDA. This section reuses material from the landscape analysis of the [e-IRG Roadmap](http://e-irg.eu/roadmap). It assesses among others the relations of RDA Europe with RDA Global, RDA national/regional entities/partners and the emerging data infrastructure eco-system in Europe.

At the European level, there is number of projects and initiatives that are related to RDA. The [European Open Science Cloud](https://ec.europa.eu/digital-single-market/en/european-open-science-cloud) (EOSC) is an EC initiative envisaging a “research data commons, widely inclusive of all disciplines and (EU) Member States, sustainable in the long-term”. In June 2017, the EOSC Summit brought together key players from across Europe committed to make the EOSC a reality by 2020. An “EOSC Declaration” gathered the key messages and commitments from all stakeholders. EC Director General of DG Research and Innovation (DG RTD) Robert-Jan Smits wrote in his note on the "EOSC Summit - The way forward" that EOSC is an initiative to join forces and that stakeholders should take some actions forward - so EOSC is defined basically as a process.

The key EOSC stakeholders have been approached to provide feedback on the EOSC Declaration in an effort of convergence and support to the EOSC Vision. RDA was represented at the EOSC Summit and was cited in the Declaration, as a key forum to reach consensus on practical implementations at European and global level. RDA Europe developed a paper entitled “RDA in Europe as an EOSC Stakeholder to achieve FAIR at all levels” to outline a clear value proposition of RDA and RDA in Europe for the EOSC.

The paper advocates that the EOSC community can immediately benefit from the RDA approach, adopt its recommendations and outputs, and learn from the broad range of past experiences of RDA contributors and early adopters of RDA outputs in building EOSC services, structures or frameworks. The EOSC community can leverage the problem-solving social structure provided by RDA, its community and the series of RDA Plenaries.

An EOSC Roadmap has been drafted and is being discussed with EU Member States outlining among others the governance and funding of the EOSC initiative, and is expected to be published later in 2018.

The EC has already initiated EOSC related projects, namely the [EOSCpilot](https://eoscpilot.eu/) and the [EOSC-hub](http://eosc-hub.eu/) projects, the first preparing the ground for EOSC in terms of community requirements capture, governance models and stakeholder engagement and the latter integrating some of the major European e-Infrastructure projects including [EGI](https://www.egi.eu/), [EUDAT](https://eudat.eu/) and [INDIGO – Data Cloud](https://www.indigo-datacloud.eu/). The [EOSC-Pilot](https://eoscpilot.eu) project that started in January 2017 has already developed a governance model for EOSC, has been developing a number of demonstrators from different domains and has been working on engaging relevant stakeholders, including RDA. Some of its goals are similar with RDA, such as reducing fragmentation between data infrastructures by working across scientific and economic domains, countries and governance models, and improving interoperability between data infrastructures.

The EC-funded project [EOSC-hub](http://eosc-hub.eu/eosc-hub-integrated-services-european-open-science-cloud) project started in January 2018, bringing together an extensive group of national and international service providers to create the Hub: a central contact point for European researchers and innovators to discover, access, use and reuse a broad spectrum of resources for advanced data-driven research. The EOSC-hub project mobilises providers from the EGI Federation, EUDAT CDI, INDIGO-DataCloud and other major European research infrastructures to deliver a common catalogue of research data, services and software for research.

Furthermore, it is planning a series of funded calls around EOSC in its next Work Programme 2018-2020. The main areas include a marketplace-type of access portal, connecting thematic clouds, prototyping new services, accessing commercial services, connecting big user communities and research infrastructures and finally supporting the governance of the whole ecosystem. The funding level is expected to be of the order of 190 million Euros and all EU efforts will be centred around EOSC.

**As the EOSC Research Data Commons vision is aligned with the RDA vision, it is crucial to understand whether and how the EOSC can act as the service vehicle at European level to implement at least part of the RDA principles and exploit the RDA neutral forum capacities. Already EOSC-hub, Freya and [OpenAIRE-Advance](https://www.openaire.eu/advance) have agreed to cooperate actively with RDA and channel their challenges through existing or new RDA WG/IGs and deliver outputs that can be considered as ICT Technical Specifications.**

In parallel, [GO FAIR](https://www.dtls.nl/go-fair/), which stands for “Global Open FAIR” has been initiated as a global initiative addressing the connecting, re-enforcing and building of strong national “Open Science Cloud” nodes. GO FAIR, thus is working at the national level, proposing “an inclusive, open and practical implementation of the recommendations of the first [EOSC High Level Expert Group](https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3353)[[6]](#footnote-7) through a federated approach, making optimal use of initiatives and infrastructures that already exist in the EU member states”. The Netherlands has initiated and co-leads the early development of the GO FAIR initiative. It aims at covering the gov-user-einfra triangle. In May 2017 Germany and the Netherlands signed a declaration calling for rapid action on EOSC. GO FAIR[[7]](#footnote-8) is not intended to build infrastructure, rather collecting best practices on content, process and implementation, influencing governments with regards to Open Science and FAIR principles, influencing training curricula, influencing/promoting cultural changes towards FAIR data and working towards a FAIR-framework.

In the meantime the [new High Level Expert Group on EOSC](https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud-hleg) was set up and already input has been provided by its Chair (RDA Europe 4 Coordinator). Trust-IT will be leading the RDA Europe 4 project starting March 2018.

Besides EOSC and GO FAIR, there are other initiatives and projects, working on services and related policies. [OpenAIRE](https://www.openaire.eu/) has expanded beyond open access to publications towards workflows and services to facilitate the harvesting of research artefacts and outputs including data. As reported at the e-IRG Roadmap 2016, OpenAIRE thus enables researchers to deposit research publications and data into Open Access repositories and provides support to researchers at the national, institutional and local level to guide them on how to publish in Open Access (OA) and how to manage the long tail of science data within the institution environment. **Disciplinary and generic national and local repositories** play a key role in particular for preservation and dissemination of the so-called long tail of research data. If researchers have no access to an institutional or a subject repository, [Zenodo](http://www.zenodo.org/), hosted by CERN and which exposes its contents to OpenAIRE, enables them to deposit their articles, research data and software. One important aspect for all these repositories is in particular the creation of persistent identifiers. OpenAIRE and its members are involved in RDA and related Working and Interest Groups around data-publication linking, legal interoperability and others. In January 2018 the [OpenAIRE-Advance](https://www.openaire.eu/advance) project started that continues the mission of OpenAIRE to support the Open Access/Open Data mandates in Europe. By sustaining the current successful infrastructure, comprised of a human network and robust technical services, it consolidates its achievements while working to shift the momentum among its communities to Open Science, aiming to be a trusted e-Infrastructure within the realms of the European Open Science Cloud

[EUDAT](https://eudat.eu/eudat-cdi) aims to move towards a sustainable research data infrastructure. The agreement signed in September 2016 formalises the roles and responsibilities of the service providers constituting the CDI (EUDAT Collaborative Data Infrastructure). Covering both access and deposit, from informal data sharing to long-term archiving, and addressing identification, discoverability and computability of both long-tail and big data, EUDAT services aim to address the full lifecycle of research data. EUDAT has developed services which are FAIR compliant, and its basic data model is compliant with the DFT core model. Furthermore, EUDAT experts were amongst the founders of RDA and participated in many RDA groups also as co-chairs. EUDAT is an RDA Organisational Member.

The European Strategy Forum on Research Infrastructures (ESFRI) plays a key role in policy-making on Research Infrastructures in Europe in several thematic areas including Energy, Environment, Health and Food, Physical Sciences and Engineering, and Data, Computing and Digital Research Infrastructures (or e-Infrastructures). In particular, ESFRI contributes to the development of a strategic roadmap that identifies vital new European RIs for the next 10-20 years. ESFRI is also working closely with the [e-Infrastructure Reflection Group](http://e-irg.eu/) (e-IRG) on electronic (or digital) Infrastructures (e-Infrastructures). e-IRG is a strategy body aiming at facilitating integration in the area of European e-Infrastructures and connected services, within and between member states, at the European level and globally. Following the [GEANT 2020 European Communications Commons](https://geant3plus.archive.geant.net/About/European_e-Infrastructure/Pages/GEANT_Expert_Group.aspx)[[8]](#footnote-9), it came up with the notion of [e-Infrastructure Commons](http://e-irg.eu/commons), which was then taken up further by the European Open Science Cloud. EOSC can be considered as the instantiation of the e-IRG e-Infrastructure Commons, which also includes or is expanded towards the Data Commons.

A number of disciplines have been developing advanced data infrastructures, at the international or European levels. Some of them, such as DARIAH or ELIXIR, are landmarks of the European Strategy Forum on Research Infrastructures (ESFRI) Roadmap, others, such as in astronomy, are global disciplinary data frameworks. [ELIXIR](https://www.elixir-europe.org/about-us) coordinates, integrates and sustains bioinformatics resources across its member states and enables users in academia and industry to access vital data, tools, standards, compute and training services for their research. ELIXIR is leading and is involved in related RDA groups. The language community, organised in [CLARIN](https://www.clarin.eu/), has established a network of FAIR compliant and quality assessed repositories as backbone of their infrastructure. They have implemented some of the core models and results of RDA such as the DFT core model, the certification according to DSA/WDS criteria, have designed an open and flexible metadata framework, registered their metadata categories in open registries and many other aspects that are relevant for RDA work. CLARIN from the beginning was very active in RDA to push cross-disciplinary best practices and to learn from other disciplines. Their members co-chaired a number of groups. Another very active community in RDA is the international climate modelling community which has its European branch in the [ENES](https://portal.enes.org/) infrastructure. It is developing its 6th version of the CMIP protocol which is the basis of the data exchange between the global partners. The work of this community is fully FAIR compliant, they are active also as co-chairs in many RDA groups and are always early adopters of some advanced concepts coming from RDA groups. Finally, the [International Virtual Observatory Alliance (IVOA)](http://www.ivoa.net/) is an organisation that agrees technical standards aiming at facilitating international coordination and collaboration to enable global and integrated access to data gathered by astronomical observatories. An information system allowing such an access is called a Virtual Observatory.

RDA in Europe has managed to bring together delegates from 47 large research infrastructures in the [Group of European Data Experts in RDA](https://www.rd-alliance.org/groups/gede-group-european-data-experts-rda) (GEDE) to address **practical problems in infrastructure building** to promote the data re-use and the support of the RDA Global ambassador program. A lot of representatives are coming from projects included in the [ESFRI](http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri) Roadmap. During 2017, GEDE members produced the report "Persistent identifiers: Consolidated assertions"[[9]](#footnote-10). As outlined in the report abstract, "Experts from 47 European research infrastructure initiatives and ERICs have agreed on a set of assertions about the nature, the creation and the usage of Persistent Identifiers (PIDs). This work was done in close synchronisation with the RDA Data Fabric Interest Group (DFIG) ensuring a global validation of the assertions. The intention of this cross-disciplinary report is to overcome still existing confusions about PIDs and the lack of detail knowledge in many disciplines.

### 2.3.1 RDA in Europe: National vs. European approaches

Given the development of a series of RDA national/regional nodes in Europe, discussions have been initiated inside RDA Europe structures on the relations of such national nodes with RDA Global, and the role or future of RDA in Europe in the next years and beyond 2020. National/regional events and in some cases also national/regional initiatives have already taken place/developed. A concrete example is RDA Germany that has been established as an entity[[10]](#footnote-11) (association) and it is currently seeking to establish an official link with RDA Global via an MoU to be negotiated with the RDA Council. Other countries/regions working in this direction include Austria, Finland, Iberia (Spain and Portugal) and United Kingdom. There are also broader national initiatives starting to appear towards national e-Infrastructure nodes or national EOSC nodes, where coordination among all e-Infrastructure and data related entities in a country, including RDA and OpenAIRE nodes, are being worked out. Such an initiative has been launched in Austria recently[[11]](#footnote-12) (November 2017).

Italy is also working in a similar direction towards the launch of an Italian Data and Compute Infrastructure initiative (ICDI) that is expected to contribute to EOSC. RDA is included in the set of initiatives that should be under this umbrella. RDA-Italy may act as a facilitator for ICDI, not only at national level, but also as an instrument for liaising with other national initiatives and discussing concrete topics, as this is currently done among the OpenAIRE National Open Access Desks (NOADs).

In parallel, the RDA Europe 4 project, which kicks-off in March 2018, has developed a plan towards the establishment of European national nodes along with the potential establishment of an RDA Europe legal entity. Other main directions of RDA Europe 4 are the undertaking of a key role in EOSC and the establishment of synergies with ESFRI Research Infrastructures and their legal entities. A summary analysis of the RDA Europe 4 project directions and sustainability plans is presented in the next section.

Although all the above elements will be dealt in depth as part of the RDA Europe 4 project and its structures, it is expected that an RDA Europe structure is maintained, at least for the remaining of the RDA Europe 4 project. An RDA Europe 4 Governance Board, composed of representatives of all project partners and of the European national nodes, is expected to take up from the RDA Europe 3 Board of Directors, acting as the main decision-making body in Europe. Beyond RDA Europe 4, it is currently unclear what type of coordination will be used at European level and it will depend on the outcome of the RDA EU 4 project and the national nodes/initiatives and the proposed establishment of the RDA Europe legal entity. In any case, it is envisaged that for several activities coordination at European level would still make sense, achieving economies of scale and pooling resources, even if funding gradually becomes available at national level. Such activities include communications and events, training, EU-wide information access such as the Atlas of Knowledge and its successor. It may be possible however to integrate some of these efforts at the global level.

On the other hand, there are voices inside Europe advocating a more direct link between the European national nodes and RDA Global, relying more at the national level and its structures, especially in the cases where there is substantial national funding. It appears that as long as there is EC funding in the RDA Europe 4 project (roughly until May 2020), the organisation at the European level via the RDA Europe Governance Board is given. Still, there may be national RDA constituents evaluating more direct links with RDA Global, as currently RDA Germany is doing, in a way enforcing a hybrid scenario. In case there are different needs and priorities between the national community and the EU along with its obligations (including Key Performance Indicators-KPIs from the EU grant), these need to be tackled and balanced. This issue may not come up though if the EU KPIs are in-line with the community needs, as the main purpose of such KPIs would be to serve the community. In any case, it is clear that RDA structures, national, European or global, have to serve the community and an effort to record their needs is essential to be able to serve them well and remain relevant. The issue of national vs. European approaches becomes more relevant after the end of the RDA Europe 4 project, and there may be different approaches in different countries. The Task Force members thought that a good level of EC funding, possibly in the form of an Operational Grant would be necessary to sustain some key EU structures, even lightweight. Where there is clear consensus among the TF members is the need to maintain the “RDA in Europe” community forum, a forum to gather all major European representatives engaged in RDA to reflect on key points. In a way, such a Forum has worked across the RDA EU projects, using the projects to fund different boards and efforts. The forum took different shapes along the projects, initially as the RDA Europe Strategy Forum (in RDA Europe 1) and later on (RDA Europe 2 and 3) via the RDA Synchronisation Assembly (SyA), with an effort to engage active RDA WG/IG co-Chairs and other stakeholders. In RDA EU 3 the GEDE network was created to organise the data practitioners from the various ESFRIs. And finally, there are also efforts to engage with key industrial stakeholders in Europe, although this is challenging and slow.

While the outcome of all these efforts has still to be seen and evaluated, the important common ground of such attempts has been to create an active "RDA in Europe". RDA Europe members need to draw conclusions from these efforts based on the current experiences and reflect on the optimal form of this "RDA in Europe".

### 2.3.2 RDA Europe 4 project

In this section, the RDA Europe 4 approach on sustainability is presented in more detail, along with the main conclusions from the discussions during the RDA EU 3 BoD and TF meetings. RDA EU 4 wants to build on the successes of RDA as a global and neutral forum acting as a problem-solving structure to enable data sharing and interoperability and on its Recommendations and outputs, some of which have already been recognised as ICT Technical Specifications. RDA and its European constituent need to continue synchronising between the geographical (national, European, global), disciplinary and sector stakeholders and deliver both at global and local levels. RDA EU 4 proposes an open and dynamic structure to expand the RDA community in Europe; it builds on top of existing national and regional networks to create a European network of national/regional nodes to enhance member state cooperation and support of RDA in many European countries, and vice versa, i.e. to enhance the representation of EU countries in RDA Global.



Figure – RDA Europe 4 partners and national nodes (Source RDA Europe 4 project).

The RDA Europe 4 is organised with 5 core partners (yellow colour) as beneficiaries including the RDA Foundation, and an initial set of 9 national nodes (green colour) as third parties, largely based on the RDA Europe 3 consortium partners. In addition, further support to RDA Global and European levels is planned via a series of open calls (called cascading grants), where existing and other third parties may be able to receive funding from the beneficiaries.

The key components and potential revenue streams described in the RDA EU 4 financial viability plan, which are analysed further below, include:

* Light RDA certification schemes, i.e. performing accreditation of compliance for adopters of RDA Recommendations and ICT technical specifications for a fee.
* Royalties from potential services related to RDA outputs or partners.
* Research and Innovation Grants, from proposals submitted to the next Framework Programme and to other national / regional funding schemes related to RDA and its outputs. Project like StandICT.eu can provide grants for data specialists working on standards.
* Contributions from Member States, in particular the national/regional nodes.
* Donations from Organisations and Individuals (for the latter à la Wikipedia)
* Trainings and workshops for data scientists and other professionals from industry

Figure –RDA Europe 4 sustainability model at the end of the project (Source RDA Europe 4 project)

The Task Force members discussed the above measures and it was clear that all of them required considerable preparatory efforts, in some cases even feasibility studies, as for the light certification schemes. The experience from the CoreTrustSeal as a global certification standard for repositories developed jointly by the Data Seal of Approval (DSA) and the World Data System (WDS) as part of RDA shows that it is quite challenging to come up with even low revenues.

Exploiting services related to RDA outputs or partners requires also considerable thinking. Such an idea has been discussed at the RDA Organisational Assembly and Organisational Advisory Board since the Barcelona plenary as a way to offer services from one Organisational member to another, possibly for a discounted fee. In addition, similar ideas have been discussed as part of the RDA WG/IG Chairs meetings for the RDA outputs. There are several issues to tackle, such as the fact that the Intellectual Property Rights (IPR) of services related to outputs, does not belong to RDA, rather to the organisation developing the service. Thus, a business model has to be developed of how RDA can possibly exploit services related to its outputs, possibly in collaboration with third party organisations.

Contributions and donations/micro-payments were also debated by the TF members. It is more likely that Member States support the national rather the global level. And there may be provisions in the national entity statutes, competing or even preventing such (double) fees. The amounts that can be gathered may also not be substantial. Already the fees from the RDA Organisational members do not get more than 100.000 USD per year and some of the members paying 10.000 USD a year, may have issues sustaining their contributions.

It may be also challenging to produce revenue from trainings and workshops for data scientists, even coming from industry. Preparing professional training courses for industry is not trivial either. Current trainings and webinars presenting RDA outputs struggle to gather enough interest and participation, even without any fee. On the other hand, this is an area where cooperation across Europe and beyond makes sense and RDA in Europe should invest in it.

**All in all, the above proposed measures require substantial work including value proposition statements and business plus delivery models. The Task Force members believe that a stream of funding via European Commission and other regions’ grants is essential to sustain RDA in Europe and overall. A GÉANT like model, possibly via operational grants, is needed for RDA Sustainability. A possible RDA in Europe role in the European Open Science Cloud may also put RDA in a key position, playing a facilitator role in national data/e-Infrastructure coordination. Such an opportunity can provide a handle for continuing European funding support towards better RDA sustainability.**

# RDA Global sustainability challenges

## 3.1. Governance structures and processes

As identified in section 1.4, one of the components of RDA sustainability is its scalable development, making sure that its mission continues to be relevant/pertinent for its community and stakeholders. This means that its current mission[[12]](#footnote-13) may need to be adapted in the future to stay in pace with the evolution of the context and of RDA activities. Furthermore, based on the RDA growth -mainly in terms of its members and the number of Working and Interest Groups (WG/IGs) that it supports-, the RDA Governance structures and processes may need to be adapted. This mainly includes the RDA Secretariat and the RDA Technical Advisory Board (TAB), while the RDA Foundation also needs to be evaluated in the upcoming years.

In particular, both the RDA Secretariat and RDA TAB are under resourced, requiring more professional and paid staff. This was brought up at the Barcelona RDA community coordination exercise[[13]](#footnote-14), and discussed during the 7th WG/IG Chair meeting in Gothenburg. The understaffing of the Secretariat is well documented. Another point from the same meeting was an annual road-mapping exercise of RDA next activities to be developed by TAB and the Chairs. There have been discussions about the future of the RDA TAB as part of TAB itself (referred to as “TAB 2.0”) and already it was decided to gradually increase the number of TAB members to 15. Other considerations include that TAB becomes (partially-)funded, with more and/or dedicated (support) persons, possibly including area directors in complement to the volunteer work.

The RDA Foundation – a charity based in England/UK- and its role in RDA may also need to be optimised. There are also considerations of creating a new legal entity and a position statement on this is also part of the recently approved RDA Europe 4.0 project[[14]](#footnote-15). “Brexit” also raises some concerns in terms of getting financial EU support. Main challenges include the fact that the RDA Foundation does not directly employ any personnel, and it mainly subcontracts (part of) the Secretary General salary (to its home organisation). Most contributions are in kind, or paid via other organisations, in particular the Secretariat staff. So, only a small part of the RDA budget goes through the RDA Foundation. The implication is that there have to be multiple budgets and there is no single control point. Furthermore, the Secretary General has no direct control over the RDA secretariat members, who are paid via their home organisations and also report to their organisations’ managers. The flow of money from the different regions via the RDA Foundation via subcontracts to cover the Secretary Generals’ salaries raises some concerns. The latter is also dealt with as part of section 3.3. In RDA Europe 4 however the RDA Foundation is a direct grant beneficiary, covering part of the Secretary General’s salary.

On the other hand, the RDA national constituents, including the ones in Europe, which are expected to be further developed as part of RDA Europe 4.0 project to further sustainability, may complicate the RDA landscape, including related processes and funding paths. Still, it appears that the RDA Foundation or a new legal entity should provide a better construct, eliminating the above drawbacks.

## 3.2 RDA Core Activities and their Funding-Financing

The main purpose of this section is to contribute towards establishing an overall budget plan for RDA Global to be sustainable at the short and longer-term. RDA Europe contributions are also estimated. To do this, a list of core services that need to be maintained in the future are presented. This brings in the renewed RDA Strategy that will be developed by the RDA Council. Note that a list of core services already exists, along with corresponding costs/budget[[15]](#footnote-16). This initial list is updated and presented in this section. Note that, as already stressed, RDA keeps multiple budgets, since not all funding flows through the RDA Foundation. The whole RDA budget should then be considered, and it is dealt with in this section.

RDA deals with two main lines of activities: the actual work of the WG/IGs building the RDA bridges, referred to as “**the** **RDA Work”**, and the support/facilitation of the RDA Work, referred to as “**the RDA Business**”. The RDA work is mostly undertaken by the RDA Community volunteers, while the RDA Business is partly funded (e.g. RDA Secretary General and Secretariat).

Leaving aside the cost of the RDA (voluntary) Work, the core RDA Business activities include the following areas[[16]](#footnote-17):

* Run plenaries
* Provide communication, collaboration, discussion tools for RDA, including the website, discussion support, email support, remote meeting support, etc.
* Facilitate and advise on the creation, operation, and outputs of the groups
* Endorse the outputs of groups and support Recommendation adoption and dissemination of outputs
* General Operations, i.e.,
	+ Ensure organisational governance
	+ Administer the global organisation
	+ Undertake business development and fundraising
	+ Reporting to stakeholders: Membership, Funders, Organisers, etc.

Other areas may need to be factored in, such as ensuring RDA presence in relevant events and fora, which is deemed important. The current vs. future budget is presented in the table below in US dollars[[17]](#footnote-18), which has been the currency that the RDA Global budget has been operating. The “Current” budget column is an estimation of the 2016 budget[[18]](#footnote-19), while the “Future” budget column is an estimation of the required budget for the future, given a series of assumptions to be presented later. The Personnel FTE column provides the estimated FTEs to accomplish the corresponding core RDA Business task (including both funded and in-kind effort), while the Personnel US$ column estimates the budget required for the previous FTE column. There is also an “Other budget” column that includes all other budget besides FTEs (such as registration fees, travel etc.). In summary, the current costs/budget is estimated to be around 7,55 FTEs[[19]](#footnote-20), of which 3,8 funded and 3,75 in-kind. The total budget is ~1,78M USD (funded part ~1,14 M USD).



**Estimated Annual “RDA Business” Budget -Current (2016)**

The future budget (required budget for the future), is estimated 13,45 FTEs[[20]](#footnote-21), out of which 7,8 is funded and 5,65 is in-kind.



**Estimated Annual “RDA Business” Budget - future**

The total budget required is ~2,49M USD (funded part ~1,81 M USD). Note however that the 1,81 Million USD is considered as the required future budget, supposing that 5,65 FTEs will be still offered as in-kind contribution (the same way that currently 3,75 FTEs are provided in kind).



**Estimated Annual “RDA Business” Budget – current vs. future**

By comparing the current with the future budget, it is required that some expenditures raise considerably in the future, such as the personnel and effort around supporting the outputs and recommendations. Furthermore, some areas in the operations would also require to raise, namely recruiting a professional business manager for undertaking several financial and business model aspects of RDA and its legal entity, currently scattered around several overqualified RDA secretariat members, or more professional effort around marketing or RDA outputs. A rough regional allocation of the upcoming future RDA income is given below, with RDA Europe and US, along with plenary registration fees and sponsorships being the big part of it. An estimation beyond 2020 should also be approached by RDA Council, given the scenarios presented in the next chapter.



**Estimated Annual “RDA Business” Budget – current vs. future**

Assumptions need to be made on how the RDA income streams may evolve from 2018 onwards and especially after 2020, documenting a few scenarios. Such scenarios may include the level of engagement from current major funders, (Australia, EC, NSF, Jisc or other regions/funding agencies) who may gradually reduce their contributions, the possible engagement of industry and other revenue streams outlined in RDA Europe 4.

In addition, the engagement of national nodes, European or other, along with possible contributions to RDA Global (if any) should be considered and document expectations. This should give a rough idea of the future outlook beyond 2020. Given the uncertainties around this, e.g. in getting concrete figures from some of the RDA regions for , such an exercise should be attempted by RDA Global.

What has been already performed by the RDA Council Finance Subcommittee with the support of the RDA secretariat is the estimation of the RDA Foundation forecast income for 2018 and 2019, in order to secure the necessary funding towards the Secretary General’s salary. The outcome is that the Secretary General’s salary is covered for these two years, without taking into account a significant 2017 bank carryover and with a conservative estimation of Organisational Membership fees, leaving also a surplus. This surplus however may shrink from possible Value Added Tax obligations and related accountant’s fees.

# RDA value proposition and assets –Proposed directions

## 4.1 RDA value proposition and its assets

The key point that needs to be answered using business terms is what is the RDA value proposition[[21]](#footnote-22) and its exploitable assets that can make it sustainable. In the RDA case, the value proposition should be used to highlight its strengths and its unique “selling” points that can sustain RDA (at least cover its costs) and keep it relevant in the long term for its community. A dynamic approach is needed as the community requirements may change over time and thus the requirements need to be collected possibly by means of frequent community coordination and feedback exercises and consultations. Such a community coordination exercise has already been performed once during the Barcelona plenary[[22]](#footnote-23).

So, what is RDA value proposition and its key assets? An attempt is done in this subsection to summarise the main points.

1. **The RDA Community: human network and neutral platform:** Over the years the community of RDA has built a strong, global, bottom-up, multifaceted **human network** composed of all different stakeholders around data sharing and interoperability, from data experts and practitioners to policy makers and funders. RDA has become the **neutral place** where anyone can create a group where a series of actions can take place ranging from discussions on social aspects to problem-solving structures, services or software around data, based upon and also acting as open standards.

*🡪 The RDA community is not a typical asset that can be easily exploited and monetised, rather a concrete point that brings high value to RDA*. *However, a potential exploitation path of the RDA platform in Europe, may be its use as part of the European Open Science Cloud implementation, where significant funding is planned in the coming years.*

1. **RDA Outputs and EU ICT specifications:** RDA Outputs are the technical and social infrastructure solutions developed by RDA Working Groups or Interest Groups that enable data sharing, exchange, and interoperability.The RDA community has developed a series of RDA outputs, ranging from the flagship RDA-endorsed adoptable **Recommendations** to supporting and other non-endorsed outputs like white papers, guidelines and reports. Furthermore, RDA Europe has promoted several recommendations to be recognised as **ICT technical specifications** by the European Multi-Stakeholders Platform (MSP) on ICT standardisation. Initially the outputs were isolated, but gradually a collection of around 10 RDA Recommendations has been built and another 5 is in the process of approval. As soon as a broader “network“ of Recommendations is built it may be easier to incentivise potential adopters; still over 75 adoption cases of RDA outputs have been reported.

*🡪 RDA outputs are a more direct exploitable asset as they can be used by anyone and can create an* *open “data economy”*. *However, monetisation via certification schemes and accreditation of adopters or engagement with/adoption by industry requires significant preparatory work, if not feasibility studies, as analysed in the RDA Europe 4 subsection.*

1. **Services from RDA outputs:** RDA recommendations, besides specifications, can be data models, schemas, ontologies, workflows or even software. There are cases where related services, tools or models have been developed as part of the RDA outputs. The tricky point though is that the Intellectual Property Rights (IPR) of these belong to a third party, i.e. the organisation that has actually developed the tools and services. Discussions around test-beds for adoption of RDA outputs or other more general test-beds around data sharing and interoperability have taken place as part of RDA governance or consultation fora, yet there is not a straightforward path to exploitation. Organisational members are also discussing as part of the OAB about providing discounted services to each other, but there is much work ahead before a list of such services can be created and a concrete proposal can be made.

*🡪 RDA-related services are thus not directly exploitable and a business-exploitation model with the engagement of third parties or brokers is required to take this further.*

1. **RDA Governance – policy making for the data commons:** Although RDA is not a policy body (rather a bottom-up initiative), its global and regional governance bodies are policy-making vehicles that pave the way towards data sharing and interoperability at the global and regional levels and strategy setting towards the data commons. As an example, the RDA Europe Board of Directors has been dealing with such strategy-setting in the area, via position statements, reports and papers. Again, this is not a direct exploitable asset, but there may be areas of exploitation such as the EOSC in Europe.

**🡪** *The RDA regional and global governing boards play a policy and strategy role that may indirectly be exploited in certain areas, such the EOSC and its governing boards in Europe.*

1. **Training and Consulting:** Although training is not directly part of the RDA mandate, RDA can record experiences and promote best practices around training as part of its relevant groups, possibly in cooperation with other groups. In addition, training about RDA recommendations and outputs is also taking place, especially at regional levels. The national RDA constituents also have a future role in this arena. Liaison with less developed areas and other organisations is also important. Furthermore, certain RDA governance groups have developed expertise in several areas that could potentially be exploited in the form of consulting. Although revenue streams in the open data, open science environment may be challenging, if not unsuitable, there may be areas such as industry where such instruments can be exploited.

🡪 *Training on RDA outputs and consulting on data practices given the RDA experience may not be mainstream exploitable assets, but in specific cases, such as industry, these means can be used to train or consult data professionals.*

## 4.2 SWOT analysis

The RDA “SWOT” analysis (Strengths, Weaknesses, Opportunities and Threats), as documented in the RDA Europe 4 project is also outlined in this sub-section. In terms of **Strengths,** RDA is new, timely and relevant, with excellent growth and strong brand, having delivered concrete outputs and recommendations. It also has a high profile with funders. On the other hand, in terms of **Weaknesses**, RDA has an under-resourced and distributed administration and coordination, low industrial engagement, tensions between bottom-up and top-down, under representation from the southern hemisphere and not fully-developed processes. In terms of **Opportunities**, RDA has created several strategic alliances while it doesn’t have clear competitors, is working on a vast arena and has support from funders. In terms of **Threats,** these are the RDA broad focus, the fact that it can overtaken by disciplinary or commercial solutions, that the solutions may be irrelevant and that the funders and participants become disillusioned. Other SWOT points, in particular Strengths-Opportunities, have been raised above, include the neutral multi-faceted platform, the potential of expanding to services and testbeds, RDA exploitation in EOSC, the opportunity of policy-making, education and consulting.



Figure 3 –RDA SWOT analysis (Source RDA Europe 4 project)

## 4.3 Sustainability elements and potential directions

This section outlines some proposed directions for the future, formulated in different scenarios based on the RDA value proposition and exploitable assets. It should be understood that RDA should work towards making the next steps, otherwise there is a clear risk of fading away. This section tries to analyse potential paths, as a single way ahead is not straightforward. The different paths may not be all complementary or even compatible with each other, so at the end a selection of a subset of directions may have to be made. Finally, currently there is no convergence at the RDA governance structures for some of the directions below, so further deliberation by RDA Global is required. Hence, this document also aims to stimulate further discussions by RDA global.

### 4.3.1 Looking for Comparisons – Well-packaged -marketed messages- Services and new RDA business models

If we raise the question what will make RDA sustainable it is always good to look for comparable activities although drawing conclusions from other examples may be misleading. The IETF is often taken as an exmaple for the RDA. A few decades ago, experts started to discuss the interconnection of computers to exchange information. Many different technologies were being designed and tested amongst circuit switching networks influenced by the experiences from telephone networks which followed different principles than the packet switching networks. Big companies such as IBM and DEC developed their own proprietary networks and thought that their solution might be the one to adopt. Finally, reference architectures such as ISO-OSI where in competition with more straight-forward designs such as TCP/IP. After some years of discussions finally the "Internet" concept stabilised being based on the packet switching concept, a unified structured numbering system including a hierarchy and a few basic protocols. Obviously advanced research needs led to first adopters of early Internet and it took a while until industry not only adopted it, but came into the driving seat. We still take profit from broad agreements on these simple basics and there is no doubt that Internet changed research methods fundamentally. As a consequence, many other approaches and initiatives failed and died.

RDA should compare the current situation in the data domain with such intiatives although data is much more complex. The overview includes:

* a landscape with silo solutions in research, industry and the public sector,
* creative people in many scientific domains continuously inventing new solutions and implementing them to the benefit of their own work,
* initiatives that have understood the inefficiencies we currently are suffering from and the need for improved solutions to meet the future challenges,
* initiatives working at different levels (policy vs. practitioners, generic vs. discipline-specific, etc.) and following different approaches (bottom-up, top-down, etc.)
* current lack of a break-through that is broadly agreed.

The great effort of the FORCE 11 group summarised earlier discussions and implementations and came up with the excellently formulated and marketed FAIR principles. At least at the level of principles it seems that "FAIR" established a common language globally with broad acceptance. One could argue that other initiatives such as CODATA and RDA have not yet turned their insights into such nicely packaged and marketed formulations of principles. Early and definitive conclusions should be avoided at the moment since FAIR principles do not cover all aspects of data management and re-use and are not a blueprint for designing systems and infrastructures. FAIR has the potential to increase interoperability as there is broad acceptance of assigning persistent identifiers and creating metadata systematically, something some initiatives implemented a decade ago, but in isolation.

The global community must make the next steps and initiatives such as EOSC, GO-FAIR, RDA etc. are working on these. ***Their contributions will have many different expressions including***:

* continuation of policy level discussions focusing on data as policy entities,
* continuation of sector and research domain discussions focusing on data as content entities,
* continuation of practitioner level discussions across and within sectors/domains focusing on data as digital objects,
* testbed implementations and conclusion forming about relevance.

As FAIR principles have shown it is important to create simple, well-packaged and well-marketed messages for gaining broad acceptance and to influence practices broadly. With respect to implementations we need to come to simple principles and technologies that have the potential to gain momentum. However, RDA goes beyond technology, i.e. in the areas of social aspects or also content. So, it is not clear whether the above points will be enough or even necessary for RDA success.

***To remain sustainable, i.e. have an impact on practices in the above sense, RDA may have to go beyond the current set of outputs that are partly excellent, but hard to sell. RDA must be capable of stimulating related services such as broad testbeds and learn from comparisons with other solutions. The exact model of how this should be implemented within RDA needs urgent attention. It may be the case that a new proxy entity, constituted by RDA and external service providers, is needed to explore new business models for sustainability. For the time being, an annual roadmap process is being discussed[[23]](#footnote-24). However, this roadmap process may not be bold enough for RDA to be sustained in the future, and concrete decisions are needed.***

### 4.3.2 Engagement with Industry

A recent joint workshop from RDA Europe and the IoT Forum has basically shown that data industry[[24]](#footnote-25) suffers from the same problems of a lack of willingness to share data and of interoperability at all levels compared to science. In fact, lack of interoperability is for various companies a business model since integrating data from different sources requires complex software solutions which guarantee income. However, this model is not scalable which is increasingly seen in data industry as well. While advanced data science is a place where many different solutions are being generated, implemented and tested funded by the governments to guarantee the best facilities for competitive science, industry needs to have a basis for investments with well-calculated financial risks.

There are different approaches in industry to tackle these data challenges:

* starting forums to present and discuss a variety of solutions being implemented
* working on reference architectures that describe the data domain holistically and based on this top-down view derive components that may have a value for the future
* developing platforms and solutions that have the chance to influence markets and gain a share of the data economy

There are very recent signs from industry that a bottom up solution such as RDA working on a variety of barriers to develop a cross-silo technology is seen as useful. However, there is a growing fear in industry that the lead companies will be able to impose their solutions on the rest of the players active in the data market. Data economy is a place of strong competition and many realise that they will not be powerful enough to confront the GAFAs[[25]](#footnote-26). Losing this competition may be a disaster for many of the national data economies.

Here is a great chance for RDA, since it is a neutral, bottom-up place for getting people together to work on components based on **open standards** that can be used by everyone. Working out such components would create an **open data economy** again, would allow also smaller companies in the various countries to participate, would enable the participation of many startups to enter this competition without being totally dependent from a few big companies, and would offer chances to establish national data economies.

The Data Innovation Forum[[26]](#footnote-27), that is planned to take place in Brussels on 30 January 2018, along with RDA pre-Plenary industrial event in Berlin in March 2018, are good examples of industry engagement, not only bringing the data economy perspective of RDA as a general point, but also mapping concrete recommendations and outputs of RDA to industrial use cases.

RDA, however, needs to take actions to be seen as a relevant player:

* it must be open and attractive to data experts from industry who need to push results in short timeframes that have a chance to be accepted as open standards
* it must have forums supported by professionals that analyse and evaluate the various results, determine their market relevance and their place in the landscape of reference architectures which are of great relevance for industry
* it needs to support a process to stimulate pivotal idea and extract/generate "selling messages" that can be compared with the grand messages mentioned above.

***After 4 years of work in RDA it is obvious that sustainability of RDA will largely depend on its capability to also address the needs of industry and to adapt its image. RDA must learn to sell its best results in such a way that its potential impacts become evident. An evaluation/advisory process must be designed and set-up urgently. Industry will be involved only if there is a business model for them to make profits.***

### 4.3.3 Testbeds and EOSC

Key for the success of TCP/IP was the successful demonstration of its working in large testbed projects. RDA Europe and RDA US have managed to organise funds for small testbed projects. This was and is important, however may not be sufficient. To show the relevance of its results larger testbeds involving a number of results need to be designed and carried out. In this respect RDA has not progressed much in this area and a number of reasons could be discussed

* there is a lack of broad support for activities such as the ones started in the Data Fabric IG, which may be partly due to the lack of an overarching advisory group discussing the potential of certain results
* the EOSC discussions in Europe are currently so dominant that all initiatives are trying to fit into the EOSC structures which has the danger of blocking initiatives that do not find their way in; on the other hand there are also great potential and opportunities given the high funding and also the focus of the new efforts in the EOSC integration
* RDA must get the support of relevant lobby organisations. This can be difficult as often they primarily promote their own convictions and solutions.

***RDA sustainability may depend on being able to show success of its solutions through running larger testbeds. It may be difficult to get this approach through the European policy makers and structures, such as the European Open Science Cloud. On the other hand, there is a great opportunity and potential, including funding, and it has to be understood whether EOSC can act as the service vehicle for the RDA platform. The design of a useful testbed project centered around promising RDA results can be thus introduced in EOSC. This will be crucial for RDA in Europe.***

### 4.3.4 RDA continues with public funding for funders to remove silos

Funding agencies can benefit from RDA in removing the silos and bridge the barriers among regions, disciplines and cultures. RDA can address the need of researchers to share data, or at least educate/consult some of the (less progressed) communities to work towards these directions and make research more reliable and reproducible. Some of the big communities and players may be happy and insist on their own solutions and silos, not putting any money on RDA.

***Funding agencies should avoid financially supporting silos,.*** ***RDA can thus contribute towards data sharing and building the bridges via its neutral problem-solving framework, organising cross-disciplinary groups, events, and providing outputs. It is not yet clear however whether this direction alone would be enough to make RDA sustainable, as it will depend on the transient decisions and segmented funding of the funders. Closer interactions with the funders is needed. The RDA funders forum should evolve into or at least complemented by a formal funders body composed only of current actual funders to discuss such a direction. A longer-term commitment may be needed for the next 7-10 years, including the next European Framework Programme.***

### 4.3.5 The role of RDA individual members’ organisations

The main stakeholder in the RDA ecosystem are the individual members, and in particular the active members rather than the passive observers. Although it is difficult to make such an evaluation and data analysis, especially as there may be privacy concerns because of analytics in personal data, it may be worth attempting this. Furthermore, it has to be understood whether the employers of individual members are backing their employees in their RDA engagement or not. It appears that this is broadly the case, although there have been possibly isolated examples of where employers did not back their employees and they had to work on RDA on their (non-funded) extra time.

***The organisations of individual members may thus be key stakeholders for the sustainability of RDA***. ***Understanding their views and backing to the RDA work can be an important sustainability path for RDA. This point can be possibly combined with other points, especially on funding agencies and working closely with the EOSC in Europe. Having a clear RDA strategy on these areas, may be much more important than refining the RDA mission***.

# Conclusions and Recommendations

This section outlines some first conclusions, along with an initial set of recommendation areas. Further recommendations will be provided in the next versions.

In this document the RDA value proposition and its exploitable assets that can make RDA sustainable have been identified. In brief, these include:

* **The RDA Community: human network and bottom-up / neutral platform**
* **RDA Outputs/Recommendations and EU ICT technical specifications**
* **Services from RDA outputs and related test-beds**
* **RDA Governance as policy makers for the data commons – Engagement with EOSC**
* **Training and consulting, especially for industrial entities/users**

RDA Europe 4 has also proposed a financial viability plan, including several measures:

* **Light RDA certification schemes, i.e. performing accreditation of compliance for adopters of RDA Recommendations and ICT technical specifications for a fee.**
* **Royalties from potential services related to RDA outputs or partners.**
* **Research and Innovation Grants, from proposals submitted to the next Framework Programme and to other national / regional funding schemes related to RDA and its outputs. Project like StandICT.eu can provide grants for data specialists working on standards.**
* **Contributions from Member States, in particular the national/regional nodes.**
* **Donations from Organisations and Individuals (for the latter à la Wikipedia)**
* **Trainings and workshops for data scientists and other professionals from industry**

Most of the RDA assets however are not directly exploitable (i.e. that can be monetised easily), such as the community or the policy making. Other ones, which seem more directly exploitable, such as the RDA outputs and ICT Technical Specifications and the RDA-related services require substantial investigation, in some cases even feasibility studies and business/delivery models. The same is true for some of the concrete RDA Europe 4 financial viability measures, such as the certification schemes, the royalties from potential services, the national contributions and donations. **The Task Force members believe that a stream of funding via grants provided by the main regional funders, including the European Commission, is currently essential to sustain RDA. In Europe, a GÉANT like model, possibly via operational grants, or the engagement of RDA in the European Open Science Cloud with a key facilitator role in European and national data/e-Infrastructure coordination can contribute towards the sustainability of RDA. To keep RDA relevant in the long term for its community, a dynamic approach is needed, capturing the changing community requirements over time via frequent community coordination exercises and feedback consultations.**

A set of recommendation areas, along with corresponding recommendations is attempted below. Main areas of recommendations include the RDA Governance, RDA finances, RDA engagement and possibly the RDA scope.

**RDA Governance**

* Evaluate the RDA Governance given the RDA growth
	+ RDA Global entities
	+ RDA Foundation legal entity
* Develop strong RDA national nodes, who can play a key role in the national developments around data and e-Infrastructures (e.g. EOSC national nodes), exploiting the RDA value proposition, and in particular its bottom-up community and neutral platform.
* The RDA funders forum should evolve into (or at least complemented by) a formal funders body composed only of current actual funders.
	+ Longer-term funding decisions and planning are needed for the next 7-10 years, including the next European Framework Programme
* Professional effort, including a business manager/analyst and marketing experts, are needed to complement the current teams, in order to work on business and delivery models for sustainability, along with marketing of RDA Recommendations.

**RDA Finances**

* RDA Global bodies (RDA Council and its subcommittees, RDA Secretary General and RDA Secretariat) in collaboration with the RDA regions should better elaborate and update the current set of budgets with professional accountants and business experts, in order to better understand what would be needed to sustain RDA in the short and long term (i.e. towards and beyond 2020)

**RDA Engagements**

* Engage more with:
	+ RDA funding agencies to better plan for the future
	+ RDA individual members’ organisations and understand their interest in sustaining RDA; an evaluation/advisory process may be needed
	+ Targeted industrial sectors (e.g. IoT, publishers, SMEs, IT) who can profit from the RDA neutral platform and open standards
	+ Big research organisations with interest in data sharing and building bridges and other horizontal groups such as computer scientists

**RDA Scope**

* Consider expanding from research data to open data (and communicating about it), taking into account big data and public data (also referred to as Public Sector Information – PSI) where SMEs and industry is heavily involved. Exploit the data economy and the neutral RDA platform towards the data economy.
1. <http://www.businessdictionary.com/definition/sustainability.html> [↑](#footnote-ref-2)
2. This should also deal with the challenges raised by DG Connect Head of Unit “e-Infrastructures and Science Cloud”, Augusto Burgueño Arjona in his presentation: [https://www.rd-alliance.org/sites/default/files/attachment/RDA9thPlenary\_keynote\_RDA\_a bottom-up initiative\_AugustoBurguenoArjona.pdf](https://www.rd-alliance.org/sites/default/files/attachment/RDA9thPlenary_keynote_RDA_a%20bottom-up%20initiative_AugustoBurguenoArjona.pdf) [↑](#footnote-ref-3)
3. https://ec.europa.eu/digital-single-market/en/european-multi-stakeholder-platform-ict-standardisation [↑](#footnote-ref-4)
4. http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1500883656509&uri=CELEX:32017D1358 [↑](#footnote-ref-5)
5. <http://www.universitiesuk.ac.uk/policy-and-analysis/research-policy/open-science/Documents/ORDTF%20report%20nr%201%20final%2030%2006%202017.pdf> [↑](#footnote-ref-6)
6. <https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf#view=fit&pagemode=none> [↑](#footnote-ref-7)
7. [http://e-irg.eu/documents/10920/382077/The+GO+FAIR+initiative+-+e-IRG+workshop+June+9+2017.pdf](http://e-irg.eu/documents/10920/382077/The%2BGO%2BFAIR%2Binitiative%2B-%2Be-IRG%2Bworkshop%2BJune%2B9%2B2017.pdf) [↑](#footnote-ref-8)
8. The GÉANT Expert Group (GEG) was established by the European Commission in December 2010, with the mandate to articulate a 2020 vision for European Research and Education networking and identify an action plan for realising this vision. On October 4, 2011 the GEG presented its vision and recommendations in its Report ‘Knowledge without Borders: GÉANT 2020 as the European Communications Commons’ to Commissioner and Vice-President for the Digital Agenda Neelie Kroes in Brussels. [↑](#footnote-ref-9)
9. <https://zenodo.org/record/1116189> [↑](#footnote-ref-10)
10. <http://www.forschungsdaten.org/index.php/RDA-DE> [↑](#footnote-ref-11)
11. http://www.ifs.tuwien.ac.at/dp/rda17-workshop-vienna/ [↑](#footnote-ref-12)
12. <https://www.rd-alliance.org/about-rda>: The Research Data Alliance (RDA) builds the social and technical bridges that enable open sharing of data. [↑](#footnote-ref-13)
13. <https://www.rd-alliance.org/rda-communication-coordination> [↑](#footnote-ref-14)
14. The Grant Agreement is not yet signed; expected late 2017. [↑](#footnote-ref-15)
15. Developed by the former Secretary General, Mark Parsons [↑](#footnote-ref-16)
16. Input by former Secretary General, Mark Parsons [↑](#footnote-ref-17)
17. Developed by former Secretary General, Mark Parsons. Does not take into account the switch to interim SG (who is not full time and thus only partially compensated). [↑](#footnote-ref-18)
18. There was not enough time to process the 2017 budget given the TF timeframe [↑](#footnote-ref-19)
19. Includes all governance bodies’ funded and unfunded effort, i.e. Council, TAB, OAB, Secretariat, plus a fraction of RDA WG/IG chairs. [↑](#footnote-ref-20)
20. Includes all governance bodies’ funded and unfunded effort, i.e. Council, TAB, OAB, Secretariat, plus a fraction of RDA WG/IG chairs. In this case, part of TAB-OAB costs is funded. [↑](#footnote-ref-21)
21. A value proposition is used mainly in start-ups or companies acting as a statement to highlight why consumers should buy their products or use their services, and is employed to show that the companies can make profit. [↑](#footnote-ref-22)
22. https://docs.google.com/presentation/d/1cpe3ifAvrgKEp\_JH-tiZI7n2TgVHk-G66ChCmMwX-Sc/edit#slide=id.p4 [↑](#footnote-ref-23)
23. At the RDA chairs meeting in Gothenburg, Sweden, in June 2017, this issue was in the focus of discussion. The need for intensified roadmapping was widely supported by the participating chairs, while the notion of "grand messages" was debated. [↑](#footnote-ref-24)
24. We define "data industry" as those companies that help solving data integration and data analytics challenges which is in contrast to the side market of "data publication". [↑](#footnote-ref-25)
25. GAFAs: Google, Apple, Facebook, Amazon [↑](#footnote-ref-26)
26. https://www.rd-alliance.org/rdaeu-data-innov-forum-2018 [↑](#footnote-ref-27)