

# Self-Assessment of the value and impact of European Support to the Research Data Alliance

---

## Executive Summary

This document presents an assessment of the RDA global alliance and the European support project – RDA Europe (phase 3 Sept 2015 – Feb 2018). The partners of the RDA Europe consortium and some members of RDA global, particularly the Secretary General – Mark Parsons, have contributed to this document. Based on a set of questions posed to the consortium by the European Commission, the answers are outlined below with a series of references and notes to other sections of the document. The achievements of RDA global which are the basis for the majority of RDA Europe activities are outlined in section 2, a wide variety of statistics about outreach and impact of RDA in Europe as a result of the specific RDA Europe project activities are outlined in section 3.

Based on the data, statistics and information included in this document, we conclude the following answers to the questions posed.

### 1. What is the value of the RDA organisation and how does the European Support action, RDA Europe, contribute to creating this value?

#### a. Value of the RDA organisation

- *It is the only organisation currently organising a global and cross-disciplinary community of data practitioners devoted to overcoming the huge inefficiencies in data management, access and re-use and offering a platform for dense interactions and solution finding. The statistics (3.1.1/2, 3.6) show that RDA has a growing attractiveness for data practitioners worldwide and that it is agile in stimulating new challenges to be addressed.*
- *The statistics (chairs 3.1.3, plenaries 3.3.13) show that RDA is attracting most of the major research domains. Two of the major communities still under-represented are the engineering and High Energy Physics communities.*
- *Within 4 years it has come up with 17 flagship recommendations (2.1.2, 2.1.3) that have shown their impact on the changes within many communities and which are adopted widely (over 75 documented cases) with many adoption and implementation activities not quantifiable in the sense of KPIs (2.1.3, 2.1.4) and difficult to document.*
- *RDA collaborates with many other international organisations, evidently some of these collaborations can be improved and others can be established. Success here depends on RDA's capability to define its "Grand Core Messages" in an efficient and comprehensive way.*
- *RDA is still based on major funding from US, EU and Australian institutions. While other regions and countries are contributing (Japan, Canada, UK), investment needs to be made to engage other countries to offer substantial financial support. In addition, the number of global and particularly, European, organisational members needs to increase.*

#### b. Role of RDA Europe

*RDA Europe is a fundamental mediator and stimulator of RDA, the RDA Europe project has:*

- *actively engaged in formulating the RDA governance and steering its procedures by board contributions and stimulating strategic thinking*

- *created awareness about challenges, stimulated the participation in solution finding, disseminated results, supported and stimulated adoption plans, carried out training courses etc.*<sup>1</sup>
- *stimulated the testing of RDA results and the formation of testbeds of a combined set of RDA results*
- *supported, where possible, the work of the most active people being chairs and board members mostly working on own funds*
- *conceived and designed the Atlas of Knowledge that has a potential to more easily understand the relationships between the different RDA Groups and to facilitate navigation*
- *RDA Europe applied a wide spectrum of methods to reach out to stakeholders and scientific communities (3.3, 3.4) and also stimulated and supported the organisations of national meetings as important keystones in community forming.*
- *RDA Europe took the role to also interact with countries outside of Europe to stimulate participation and form global communities.*
- *RDA Europe also took steps to liaise with other organisations such as ITU, W3C-WoT, IoT Forum, DONA, BitKom and others to come to continuous collaborations*<sup>2</sup>.
- *RDA Europe also reacted on new developments such as Digital Single Market, EU Copyright regulations and European Open Science Cloud (EOSC), defining RDA's role and establishing a reputation and concrete interaction with policy makers.*

## 2. What are the main actions that RDA Europe has put in place in order to:

### a. engage European researchers in the RDA process?

- *RDA Europe used a wide spectrum of measures to engage European researchers*
  - *(co) organising **outreach meetings** of different sorts including intensive interviews, organisation of joint meetings, participation in meetings by giving talks, etc.*
  - *organising **Science Workshops** to remain in a regular interaction with leading domain scientists*
  - *stimulating **national meetings** and organising the support of European and global experts*
  - *organising **technology working meetings** to better integrate computer scientists and RDA practitioners*
  - *organising **training meetings** (=> 2 per month) that include webinars and f2f meetings*
  - *organising **meetings with hands-on** activities (datathons, summer schools)*
  - *financing & supporting **collaboration projects** to stimulate solution adoption and testing*
  - *forming and moderating the **Group of European Data Experts (GEDE)** as an interaction platform of practitioners from large European infrastructures*
  - *stimulating **European organisations** to become RDA organisational members*
  - *running early career programmes to support researchers & scientists to participate in **RDA plenaries***
  - *intensifying the **ambassador** program in Europe*
  - *creating and distributing **dissemination material** including "**success stories**" which are also appealing for the scientific communities to showcase successful adoptions*<sup>3</sup>

<sup>1</sup> This role may change over time dependent on the progress of solution finding.

<sup>2</sup> In case of international organisations it would be the task of RDA global to come to agreements once trust relationships have been established.

<sup>3</sup> It is obvious that it requires different types of material to reach out to the practitioners compared to those directed towards policy makers. Yet the group of practitioners is heterogeneous and often more interested in hands-on material.

*Practice has shown that not all measures were equally successful and at the end of the RDA EU3 project an open evaluation will be required. Most important were the national meetings and the f2f meetings and those with hands-on components. The GEDE platform also seems to fulfil an important gap in the interaction pattern. Collaboration projects have an important role in demonstrating the usefulness and effectiveness of RDA results.*

**b. ensure that the European scientific communities benefit from the outcomes of RDA?**

- Most of the measures mentioned above are focused on ensuring that European communities profit from RDA's work and results.
- These are in particular
  - distribution of information about RDA results, trends and recommendations at various layers (Europe wide, regional (e.g. South East Europe, Nordic countries) national, community wide)
  - offering training courses of different types addressing different kinds of experts
  - giving concrete advice on data challenges and possible solutions
  - establishing interaction platforms to intensively discuss data issues driven by the concrete practices within the infrastructures
  - obtaining recognition of 4 RDA recommendations as ICT technical specifications in Europe to be referenced in public procurement.

**3. Which are the indicators you have that show the success of RDA Europe's actions?**

- The summary matrix in 3.6 shows that RDA Europe was very active and successful. In detail we can refer to the following points:
  - RDA engages European representatives from almost all scientific communities in RDA related activities and we facilitated convergence building in some data issues through very intensive interaction patterns.
  - RDA is now very well-known in most scientific domains and the view on what RDA can be used for as a bottom-up mechanism has become clearer.
  - Many communities already changed their practices by adopting fundamental models and principles - in fact almost all communities agree now for example that they need to build trustworthy repositories assessed by DSA/WDS, that they need to assign PIDs and metadata with their digital objects, that PIDs are used as handles to access data, etc.
  - Many communities now create Data Management Plans and funders request to submit DMPs with proposals.
  - The number of European organisational members committing to the alliance from Europe has grown.

**4. What are the stakeholders / scientific communities that you serve so far in Europe? How well do you serve them? (Provide evidence based on the above-mentioned indicators or on other objective evaluation criteria)**

**a. Which stakeholders/communities do we serve in Europe?**

- The matrix in 3.6 gives a good overview which scientific domains we are serving to which degree.
- The list of groups involved in GEDE (Appendix A) is also a good indication of which large infrastructures in Europe accept RDA as a moderator to bring together data practitioners and to formulate improved recommendations.
- As the matrix shows we also continue to have intensive interactions with policy stakeholders to continue a bi-directional information exchange. Many of these interactions happen also in advisory boards and national interactions with policy makers which are difficult to

document. In those European countries with an active RDA community RDA members are voices people listen to since RDA is seen as a neutral place for interactions.

**b. How well do we serve them?**

- RDA Europe has the clear feeling that we were very successful in our policy level work until now although some work needs to be intensified (see below). The "data challenges" are now subject of all kinds of policy level and advisory board meetings at European, national and organisational level and certainly RDA EU contributed to this change of mind-set.
- RDA Europe did more than could be achieved with just the project funds with respect to stimulating and supporting adoptions, starting larger testbeds of combined configurations, running intensive f2f training courses, giving advice, disseminating relevant information, etc<sup>4</sup>.
- We see areas where the intensive interactions we had in Europe with many different scientific communities led to an increased convergence in data solutions which will reduce the costs for the involved communities and increase interoperability. Yet we see the need to make steps beyond recommendations and invest much more funds to test out complex configurations which is beyond the potential of the project. Nevertheless, RDA turns out to be an excellent platform to engage a few globally acting communities to take real steps.
- During the last year RDA Europe intensified its interaction with industry which is not easy given the economic expectations of companies. However, there are clear indications that industry is recognising the value of a neutral broker of standards and best practices to make data work much more efficient than it is currently the case.

Therefore, in general RDA Europe is satisfied with its results and the broad acceptance in many communities of its role and network capacities.

**5. Are there relevant stakeholders/scientific communities that are not served or sufficiently engaged in RDA so far? What are your plans to approach those stakeholders/scientific communities and by when (a roadmap will be appreciated)**

**a. Which stakeholders/communities are not served well?**

- As explained in the document a few stakeholders/communities are not served as well for different reasons:
  - The **HEP community** has established its own standards and practices for primary data streams influenced by CERN and the various tier level centres give data services to the science community. CERN library experts are deeply involved in RDA discussing issues of secondary data and data publishing.
  - Some of the big data centres involved in **EIROFORUM** are not as active in RDA as one would wish to bring in their expertise and knowledge.
  - Some communities such as for example **bio-informatics** and **neuro-informatics** have established strong global interactions talking about standards and best practices. They are represented in RDA and collaborate in different working and interest groups. Here the task is always to closely synchronise the activities.
  - The **engineering** community is hardly involved in RDA global activities although at national level (see matrix in 6.3) some cross-fertilisation is happening.
  - The number of **organisational members** from Europe grew during the last months but more need to join.

---

<sup>4</sup> It should be noted that a) some partners invest much more time as contracted with the project and b) people outside of the project participate in training and outreach events. Due to this effect the EC gets much more value for every Euro it spends than is usual for projects (see also 3.1.4).

- RDA's contribution and benefit for the **EOSC** is being strengthened. As EOSC is still under development (HLEG, EOSC pilot, national FAIR ports, etc.), this is still work in progress and will evolve over the next 2 years.
- **Industry** involvement is still limited. Concrete plans are now in place to establish a balanced Industrial Advisory Board that will focus on SMEs. The unique selling points, in particular in form of Grand Core Messages, of RDA for industry, in particular large industry, need to be defined.

**b. Which plans to include the missing stakeholders/communities?**

- The HEP community is being contacted. A strategic meeting should be organised in the first half of 2017 to discuss with policy level experts at CERN and from the HEP scientists.
- First interactions between EIROFORUM and RDA Europe took place. A strategic meeting should be organised in the first half of 2017.
- An intensive interaction with bioinformatics is ongoing and it would be good to plan a strategic meeting with the leading persons of ELIXIR to assess the state of interactions and the mutual expectations. The same holds for the neuro-informatics community centred around INCF (International Neuroinformatics Coordination Facility) and in Europe now partly also around the Human Brain Project (HBP)<sup>5</sup>.
- Engineering departments are highly involved in IoT. Since RDA and RDA Europe are now investing more time on addressing the data needs within IoT, there could be a basis for improved interaction. A meeting to be held at the IoT week in Geneva in June 2017 together with industry will give more insight in this respect. Also engineering communities are typically organised nationally due to their close relation to industry. As started in Germany we will continue to interact and strengthen the relationships.
- The EOSC pilot project started in January 2017 and both projects have agreed to set up a task force to define the synergies and areas of concrete contribution in 2017.
- It is certainly a priority for the coming year to convince more organisations to join as organisational members of RDA. We currently have 20 European members and the goal is to increase this number to 30 until February 2018.
- As indicated new steps are being taken to improve the engagement with industry. At the Barcelona plenary RDA is planning to bring together experts who will talk about "data Markets" and for June the IOT Week workshop is being planned. In a few countries such as Germany close and regular interactions between RDA EU and German Industrial Organisations such as BitKom have been agreed upon.

## Contents

Executive Summary .....	1
1. Introduction.....	8
Classification System .....	8
2. Achievements .....	8
2.1 RDA Results .....	9
2.1.1 General .....	9
2.1.2 Concrete Recommendations & Impact .....	10
2.1.3 Quantifiable and known Adoption .....	11
2.1.4 Hidden Adoption and Impact .....	12
3. Statistics and Outreach.....	14
3.1 RDA Global.....	14
3.1.1 RDA Membership Development.....	14
3.1.2 RDA Plenary and Group Engagement.....	16
3.1.3 Chairs Engagement across disciplines .....	17
3.1.4 Estimation of Personnel Effort in RDA.....	18
3.2 RDA Europe .....	19
3.2.1 RDA Europe Membership.....	19
3.2.2 RDA Europe Engagement .....	20
3.3 Outreach Activities of RDA Europe.....	21
3.3.1 Outreach Meetings.....	21
3.3.2 Science Workshops.....	24
3.3.3 National/Regional Meetings (Overview) .....	24
3.3.4 Industry Meetings.....	25
3.3.5 Outreach beyond Europe .....	25
3.3.6 RDA Ambassadors .....	26
3.3.7 Technology Working Meetings.....	27
3.3.8 Webinars .....	27
3.3.9 F2f Training Meetings (training, SS/WS/Datathons) .....	28
3.3.10 Collaboration Projects .....	29
3.3.11 GEDE Group (Group of European Data Experts) .....	30
3.3.12 European Organisational RDA Members.....	31
3.3.13 European Participation in Denver Week.....	31
3.4 National/Regional RDA Meetings.....	32
3.4.1 Finland/Nordic.....	32

3.4.2 France .....	33
3.4.3 Germany .....	34
3.4.4 Greece .....	34
3.4.5 Ireland.....	35
3.4.6 Italy .....	36
3.4.7 UK .....	36
3.5 Not-well represented communities .....	36
3.5.1 Computer Scientists.....	36
3.5.2 Large European Entities.....	37
3.5.3 Engineering.....	37
3.5.4 High Energy Physics Community .....	38
3.6 Summary Engagement Matrix.....	39
Appendix A: GEDE Group Classification .....	40

## 1. Introduction

The following sections summarise the achievements and impact of RDA in Europe and globally.

### Classification System

As basis for our classifications into scientific domains, for the purpose of this report, we used the revised OECD Frascati Manual 2002<sup>6</sup> scientific classification system. Given the data driven activities in RDA, which is not purely oriented on science but also needs to indicate other contributions the following changes were applied:

- introduction of categories for “eScience, data, compute centres” and “libraries/archives”
- division of the “natural sciences” category into “physical & chemical sciences”, “earth, geo and environmental sciences” and “biological sciences”
- addition of categories for industry and organisations which do not always apply
- addition of “neuro and cognitive sciences” to “medicine and health”

This classification scheme in most instances gives balanced overviews about the engagement level without spreading to overly granular categories.

## 2. Achievements

RDA was setup to promote a change with the large inefficiencies and the resulting huge costs in data management, access, re-use and archive. About 80% of data is not accessible any longer after some time, about 80% of the precious time the scientists invest on typical data management tasks (finding, converting, curating, etc.) and integrating data from different sources (projects, disciplines, countries) is often not done, since the integration costs are simply too high due to the fragmentation. A global and cross-disciplinary approach will help to overcome the many barriers. These inefficiencies will become even more severe given the trends driven by the billions of smart devices that will be deployed in the coming years and that create continuous data streams.

A number of factors are key for the success of RDA and thus need to be considered in this document:

1. Firstly RDA is a **global and cross-disciplinary interaction platform** bringing many data practitioners together to exchange opinions and solutions about how to make data work much more efficiently. This factor is difficult to concretely measure since all events at global, regional and national level organised and stimulated by RDA and RDA Europe contribute to this.
  - a. Events where RDA topics are being addressed help to transmit RDA results, but also help in identifying new challenges and solutions to be taken up within RDA.
  - b. Events organised in collaboration with scientific communities are of great relevance independent of their geographical scope (global, regional, national).
  - c. The most productive events are those that lead to a convergence of opinions and/or to concrete actions, but this cannot always be predicted and may depend on charismatic persons.
  - d. Often the impact of meetings can only be seen months later when communities start changing their practices incrementally.
  - e. The number of meetings alone is not a sufficient KPI, but it is a necessary criterion to establish an intensive dialogue between key people and reach critical mass.
2. The second relevant point for RDA success is the **productivity and creativity of the working groups** that need to create specifications for components, processes, guidelines etc. to

---

<sup>6</sup>

<http://www.oecd.org/sti/inno/frascaticmanualproposedstandardpracticeforsurveysonresearchandexperimentaldevelopment6thedition.htm#fos>

improve data practices within 18 months. As is shown in chapter 3.1.2 we are satisfied with progress to date since the working groups, which were initiated by senior practitioners, are now attracting a new generation of practitioners who are starting to take the lead, and influence their communities.

In addition to the WG results, **adoption of the results** is crucial to indicate that they indeed solve a problem shared by several organisations. It is worth noting here that interest groups are also invited to identify their outputs of interest for a wider audience, which are publicized as 'RDA Supporting Outputs', some of them with a high impact such as from the Legal Interoperability, the Data Publishing and Data Fabric IGs.

- a. Adoption however can only be indicated partly in quantitative ways as will be shown in chapter 2.1.3/4.
3. The third relevant point for RDA success is the **continuation of activity in interest groups, in organising Birds of a Feather (BoFs)<sup>7</sup> meetings or in organising other kinds of facilitating activities** that also have an impact on the change of data practices. Here in particular we refer to those activities in RDA that are clustering RDA topics into focus areas and thus go beyond the production of individual WG results. After the first years, the creation of first WG results was the focus, since the Tokyo plenary (March 2016), more activities of clustering towards broader testbeds have been initiated.

## 2.1 RDA Results

Chapters 2.1.1 and 2.1.2 are based on the statements made by RDA's secretary general at the visits at the EC in November 2016.

### 2.1.1 General

- **Membership** - RDA has shown by its growth of engaged experts that it is accepted as a useful global platform of interaction between data practitioners<sup>8</sup> from a broad range of disciplines. These data practitioners are the key drivers behind organising and changing data practices in the scientific communities and in data/information infrastructures. The main target is not scholars<sup>9</sup>, since most of them will rely on the data practitioners as mediators, but disciplinary expertise is very present
- **Promotion of good practices** - RDA helped a lot in popularising the idea of "sharing data" (in the broad sense of making it available, discoverable and re-usable) and in disseminating the messages about basic pillars that are required to build a seamless functioning global data infrastructure, such as PIDs, core metadata, certified trusted repositories, trusted registries, etc. It also helped convincing stakeholders to make Data Management Plans active and that research and e-Infrastructures take data preservation seriously.
- **Beyond principles** - Due to its approach to look for concrete solutions, RDA, however, goes beyond stating principles but tries to turn discussions globally to come to changes in data practices.
- **Changing Minds** - The global interaction is stimulating a discussion about the necessity of tools and infrastructures between key stakeholders from relevant countries and regions that could change current practices. Yet we are not far enough along to convince big industry of the need for increased interoperability. Initiatives such as EOSC have the potential to change minds and RDA has the potential to be one of the EOSC pillars.
- **Integrative platform** - RDA helped bring together rather heterogeneous global communities such as the agro-community with its many initiatives and organisations to work on standards, i.e. it

---

<sup>7</sup> RDA encourages potential new working of interest groups to run a BoF at the plenary meeting to understand the interest in the topic, recruit members, identify potential overlaps and synergies with existing groups, etc.

<sup>8</sup> On purpose we are using the term „data practitioners“ which is incorporating yet not well defined terms such as data managers, data librarians, data scientists, data service providers, etc.

<sup>9</sup> Scholars are those researchers that are focusing on the creation of highly rated scientific papers.

has shown its potential to act as a platform for global interactions for global challenges such as the food supply.

- **Adoption/testing/training** - RDA has put the need for adoption, testing and training of new recommendations on the agenda of infrastructures globally which implies an enormous impulse to globally synchronise activities and discussions including not only US, Australia and Europe, but also countries such as China, India and Japan. Due to joint activities with CODATA there is also an outreach to Africa.
- **Evolution of Principles** - At the pre-ICRI meeting in Copenhagen in March 2012 which was a first meeting of the group of experts that were instrumental in the launch of RDA, a first discussion took place where the important layers were presented and discussed: discovery, access, interoperability, re-use. These found their way in slightly differing variations into the G8 principles on data, the core data model presented by the RDA Data Foundation & Terminology group presented at the Amsterdam plenary (March 2013) and then into the now widely supported FAIR statements. It is great to see a global coherence and RDA has contributed to this development (see more detail in 2.1.4).

### 2.1.2 Concrete Recommendations & Impact

- To date RDA has released 17 formal recommendations and several supporting outputs. These outputs have been adopted and concretely put into use by at least 76 organisations<sup>10</sup>: <https://www.rd-alliance.org/recommendations-outputs/adoption-recommendations>
- The core data model of RDA's Data Foundation & Terminology group has now been accepted by nearly all research domains and is being put in place - it simply states "register all your digital objects in trustworthy repositories, associate PIDs and metadata with all your digital objects"<sup>11</sup>.
- Four of the 17 recommendations have been formally accepted as ICT technical specifications for reference in public procurement in Europe —highlighting the quality of RDA's work and the growing trust in and reliance on RDA.
- These original four recommendations are frequently used together by early adopters, and they form the basis for implementing a 'Global Digital Object Cloud' by a number of globally acting communities (climate modelling, biodiversity and natural museums, material science) that now seems to attract more communities and also be interesting for industry.
- Persistent identifiers (for data, documents, people, organisations, workflows, etc.) have become a central, core component for data sharing. RDA is working with many European research infrastructures and others to put this into working practice. Publishers, authors, and repositories, all now expect data to have a PID.
- Due to RDA's concerted efforts, PIDs may be approaching the tipping point of becoming the generic interoperability solution we aim for. In fact, many scientific communities up to the key researchers have accepted the message that PIDs will be key for a more efficient data practice.
- Under the leadership of RDA and others, data citation is becoming a routine practice. Now RDA is working out the details and actively addressing the "reproducibility crisis".
- RDA has demonstrated the power of building a community and an ability to reach across disciplines, geographies, and other types of difference. Its community itself is perhaps the most valued aspect of RDA to date.
- RDA partners effectively with like-minded organisations to amplify our collective impact. For example, we have partnered with CODATA in developing a successful data management curriculum for researchers, especially those in the developing world. CODATA and RDA are working together strategically to engage Africa, and financial support from South Africa is a potential outcome. Other active partnerships include GEO, WDS, and DataCite.

---

<sup>10</sup> It needs to be accepted that RDA as a lean organisation does not know about all activities in all countries and that neither the group of chairs nor the secretariat can track all usages, in particular when the outputs become part of the 'normal' landscape. RDA outputs are open and anyone can use them in their own context.

<sup>11</sup> Only some communities probably have read the full set of documents, but due to the many meetings where the basics have been mentioned in different wordings the effect can clearly be seen.

- RDA interacts with initiatives such as W3C and makes not only use of well-established standards such as RDF and PROV, but also discusses gaps of such standards as for example for PROV and also is taking up new standards such as ResourceSync<sup>12</sup>.
- The ISO 5127 (Information and documentation vocabulary<sup>13</sup>) update included recommendations from the Data Foundation & Terminology outputs.
- RDA is closely working together also with international organisations such as ITU where for example solutions for a global identification mechanism are being discussed and seem to lead to a global debate.
- The OECD has adopted the RDA cost recovery model and has built an expert group on top of RDA's work to develop recommendations for sustaining data.
- Springer/Nature uses the Data Publishing Workflows reference model as a business tool to see how they compare to their competition.
- The Data Seal of Approval and the World Data System came together in RDA to harmonize their repository certification schemes which can be seen as a globally accepted basic certification option. They are implementing together this new certification framework.
- "23 Things Libraries can do to get involved with data management" has become a "best seller" in 11 languages. It formed the basis for a highly successful training course for institutional libraries across Australia.
- There is growing research discipline engagement with the model defined by the Agricultural IG. Their wheat interoperability recommendation has led to similar work with rice, fisheries and more general and broadly applicable solutions.

#### **Impact (2.1.1 & 2.1.2)**

- *The number and quality of results are remarkable for an organisation that started work just 4 years ago (March 2013), that first had to set up a smoothly functioning light organisation and invested 2 years to understand its way of functioning as a body and to stabilise the processes.*
- *Even large international organisations are aware of RDA results, try to incorporate them and build on them.*

#### **2.1.3 Quantifiable and known Adoption**

Every Working Group must identify "early adopters" that are participating in the group and ensure they test the results and where possible to develop software demonstrators. These "early adopters" are well-known to the chairs and RDA and mostly active in adoption as indicated in the case statements.

Other activities where RDA secretariat is aware of adoptions are:

- The collaboration projects that are being supported by RDA Europe (see 3.3.10) and RDA US.
- The adoptions cases presented during the plenaries.

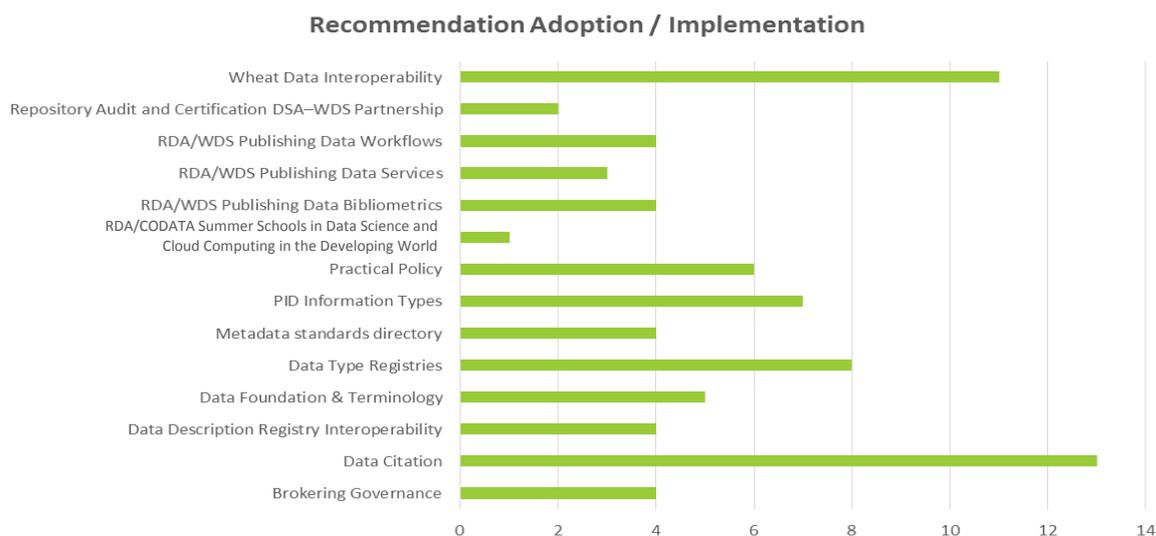
All these "known" adoption cases are part of the statistics below, but these are just the tip of the "iceberg" since much adoption cannot be quantified in simple ways and since much adoption is simply unknown to RDA. The chairs of the Working Groups are important mediators about adoptions but a) they do not know all cases and b) participation in RDA is not their main job and hardly have the time to check what happens with their results. In some cases, already adoption of RDA results disseminates naturally in and between communities without specific RDA trigger<sup>14</sup>.

---

<sup>12</sup> RDA colleagues pushing the Global Digital Object Cloud work now also together with the Web of Things experts to co-organise meetings.

<sup>13</sup> [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=33636](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=33636)

<sup>14</sup> For instance, outputs of the Wheat Data WG are known worldwide in the agriculture community including all its different organisations.



The above diagram indicates the number of known adoption cases. It is worth mentioning that adoption within Europe accounts for 42% of these cases which reflects the general impression of a very active European RDA community.

All these recommendations are directed towards different groups of users. While the results of PID Information Types is mainly addressing the services from the limited set of PID service providers and the results of the Practical Policies (PP) group those of data repositories, the Metadata Standards Directory (MSD) can be used by everyone looking for a suitable metadata standard. This implies that for RDA it is hardly measurable how often repositories make use of the PP cookbook – a 150 page document with recipes of how to carry out essential management operations – or how often people make use of the MSD. It does not make sense to calculate every inspection of these two results as an adoption.

For the joint DSA-WDS partnership on repository certification 4 adoption cases are mentioned. They can better be described as demonstration cases. DSA and WDS are widely used in their respective fields and infrastructures, e.g. CLARIN adopted both standards and asked its data centres to be compliant too and in the meantime more than 20 centres in Europe have done so. The most important achievement of the RDA WG however was to bring these two initiatives together to join on one set of criteria with the goal:

- to overcome the confusion in the communities which one should be used
- to have ONE standard with worldwide impact.

The impact cannot be measured in adoption cases but in global relevance: almost everyone globally who is building a trustworthy repository knows about the DSA/WDS certification criteria. The impact goes beyond the definition of a common set of criteria since the two organisations are now building a common certification framework, and through collaboration this one aligned framework is set to be an improvement on its two predecessors.

Therefore an evaluation of the impact of results can be made by looking at a few examples of hidden adoption which are ‘top stories’ as well.

### 2.1.4 Hidden Adoption and Impact

Two examples of typical cases in which hidden adoption have a huge impact are outlined below.

#### *Data Foundation & Terminology*

One of the biggest impacts of RDA can be seen in the diagram which indicates the evolution of basic principles and their implementation and which led to the broad acceptance of the FAIR principles. At

the first informal “RDA” meeting – a pre-workshop to the ICRI meeting 2012 in Copenhagen – Larry Lannom presented his view on the “access layers” for efficient dealing with data and he distinguished 4 basic layers. The following discussions led to a) the formulation of the principles from the G8 Data Group and b) to the core data model of RDA’s Data Foundation & Terminology (DFT) Working Group which is based on more than 20 use cases from different scientific fields. Ongoing discussions in RDA and other contexts led to the refined and now broadly accepted FAIR principles. All these principles and underlying models overlap where the DFT group statements are more operational than the other formulations. DFT states that one will need trustworthy repositories to find and access data, that one needs to assign persistent identifiers to access/find exactly those digital objects that are meant, and that metadata of different sorts will be required to support finding, interoperability and re-usability.

These messages, including the ones of DFT, have been presented and discussed at so many different meetings that we can claim that these principles and the core model have reached out to almost all scientific communities in Europe. RDA Europe has supported for example the implementation of these principles and the core model in repository systems such as FEDORA to showcase how to actually build trustworthy and well-organised repositories, and is using this in training courses.



It can be mentioned that a few communities started early with designing and implementing systematic approaches to proper data organisation which are fully FAIR compliant. Two examples can be given. The CLARIN community specified already in 2009 that recognised data centres need to assign PIDs and metadata to digital objects to make them discoverable, accessible, interoperable and re-usable, i.e. the core messages spread out across the whole language resource community throughout Europe and even beyond. The EUDAT data infrastructure applied the same principles and model in some of their core services such as B2SHARE and B2SAFE resulting in broad discussions, awareness and acceptance in many of the major data centres in Europe and in 5 communities (physiology, climate modelling, linguistics, seismology/volcanology, biodiversity).

Therefore due in particular to the work of RDA and RDA Europe these basic principles and this core model have reached out to a large number of data practitioners in almost all scientific communities in Europe and are being implemented in concrete data organisations by a number of initiatives. The practices are changing stepwise and there is no way to measure this change except to study the practices in the communities which RDA Europe does by its various stakeholder meetings. Only few communities will notify these stepwise changes explicitly as in the case for the CMIP6 data model in climate science. Many others will not be part of adoption stories.

### **Practical Policies**

The Practical Policies Working Group collected use cases along 11 fields of typical data management operations such as replication, integrity checks, etc. At an abstract layer the various tasks are described with the help of templates. To make them operational, however, these abstract templates

needed to be translated into code snippets that reflect the heterogeneity of the operating systems, of the storage systems, of the data organisations etc.

In the EUDAT project the most time-consuming task was to implement a service that allows communities to replicate large data collections from a community centre to the core network of service providers. The reason could be found in exactly the heterogeneity of solutions at various levels. It was not possible to specify "one simple API" (application programming interface) and then ask everyone to just "quickly" adopt to it. This approach has been chosen in the DataONE project for example with the consequence that it cost two years of code development to map a single large data collection in such a way that the API could be used. Similar challenges can be identified in large initiatives such as the Human Brain Project. The current state of the art in cross-border data management and access is highly inefficient and too expensive. Therefore, the cookbook from the PP group is an excellent document for all data managers to look at and we know that it is widely spread in the community that is using for example the iRODS tool which is the most popular data management tool for data federations worldwide. Adoption in the PP case can mean

- to look at the abstract templates to get inspired about how to carry out a certain management task
- to look at code snippets to see how to implement certain management tasks given certain software and data organisations

We need to accept that no practitioner will signal this kind of usage to RDA. People will simply reuse either the documented templates or the code snippets.

### **Impact of RDA Results (2.1.3 & 2.1.4)**

*From these two chapters about adoption we can conclude that*

- *RDA had already an enormous effect on changing the awareness about basic requirements of data organisations which goes far beyond the mentioned adoption cases.*
- *The reported adoption cases are important for potential users to look at and see how results are being implemented.*
- *In some communities the FAIR principles have already been put in place due to the ongoing discussions stimulated amongst others by RDA during the recent years.*

## **3. Statistics and Outreach**

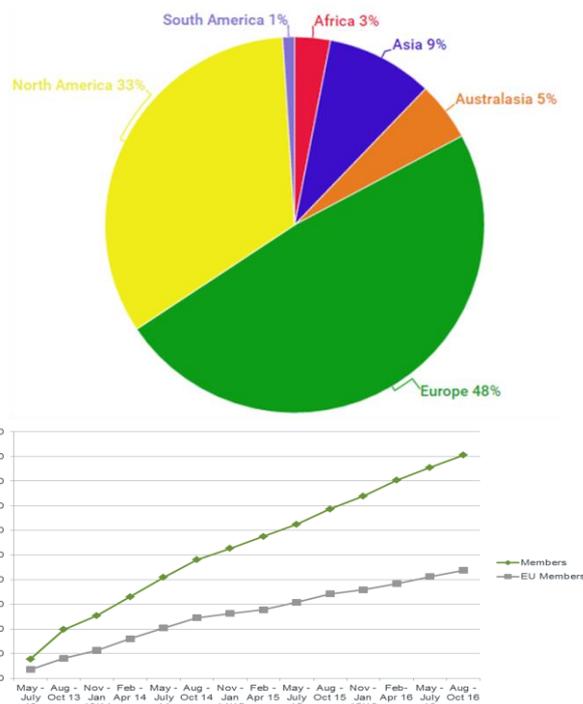
The statistics outlined in this chapter cover RDA Global, RDA Europe, RDA Outreach and national RDA activities, since they all contribute to the value and impact of RDA and RDA Europe.

### **3.1 RDA Global**

This section presents the development of membership at RDA global level indicating the attractiveness of RDA and the general knowledge about RDA results and relevance.

#### **3.1.1 RDA Membership Development**

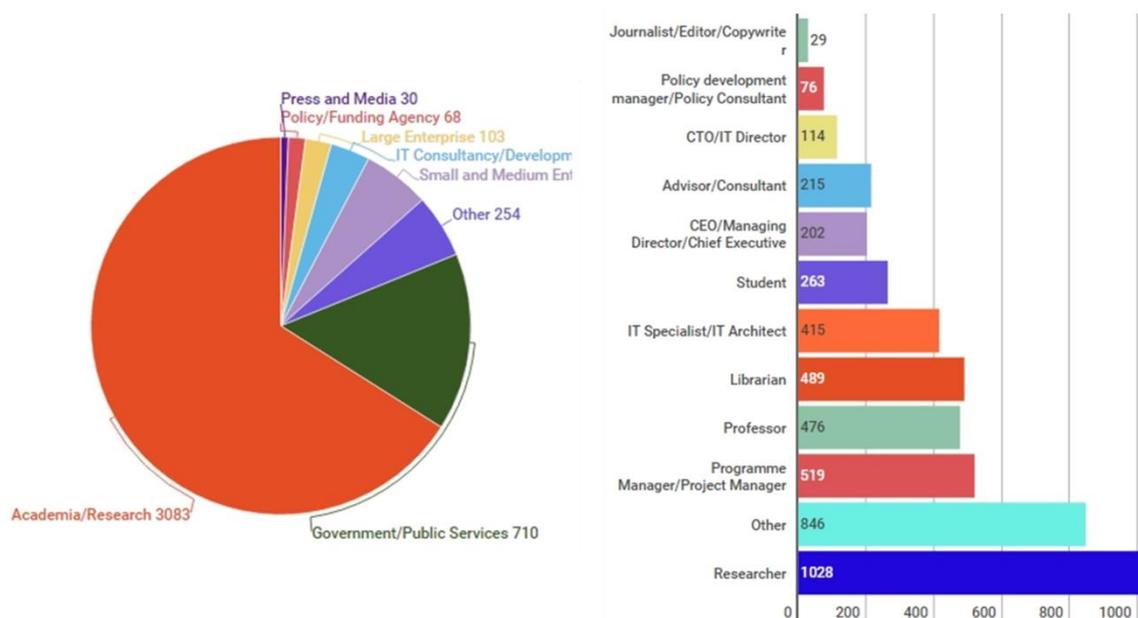
The diagrams below indicate the steady growth of RDA membership to over 4,700 members (as of January 2017) from 117 countries. Many of these members come from Europe and North America, but data practitioner representation from other countries is also increasing. The % representation from Europe and North America is indicative of the consistent and frequent engagement and outreach activities in both regions as a result of the RDA Europe support project and the RDA-US one.



Of course “membership” can imply different levels of engagement. In a light adaptation of the well-known RACI levels we can make the difference between

- “Drivers”: People taking responsibilities, i.e. being engaged as Group chairs and Board members (about 210)
- “Contributors”: People actively contributing, i.e. coming to the plenaries and contributing to the group discussions (about 700)
- “Observers”: People who just want to be informed and follow the discussions and trends.

For the progress in RDA the first two groups are crucial and in 3.1.2 it can be seen how active engagement varies over time. The third group is important for dissemination of RDA activities and outputs, as well as cultural change.



By far most of the members are from academia and research which is the core group of data practitioners RDA wanted to engage with first. The second largest group is from governmental

institutions and public services. A number of participants come from enterprises such as SMEs, consultancy and larger companies. Company challenges were not in the focus of RDA.

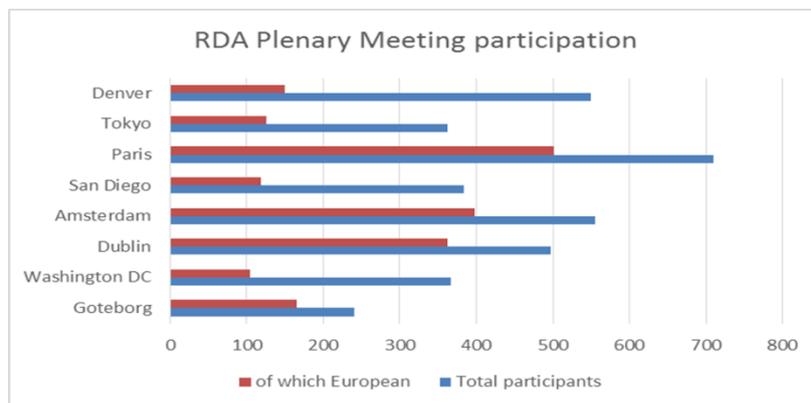
Based on the classification selected upon registration, members choose their role / qualification from a pre-fixed list. The table only gives an indication of the type of experts we can find in the membership, as many choose "Other". We can assume that many people are acting as data managers, data librarians or data scientists in their various institutions.

### Impact

From these statistics we can conclude that

- RDA is increasingly attractive to data practitioners worldwide in particular in the highly developed countries.
- Europe has a strong representation which is due to the intensive outreach work of the RDA Europe projects.
- The membership profile is very broad as intended by RDA, which reflects in the current difficulty to set up proper classifications.

### 3.1.2 RDA Plenary and Group Engagement

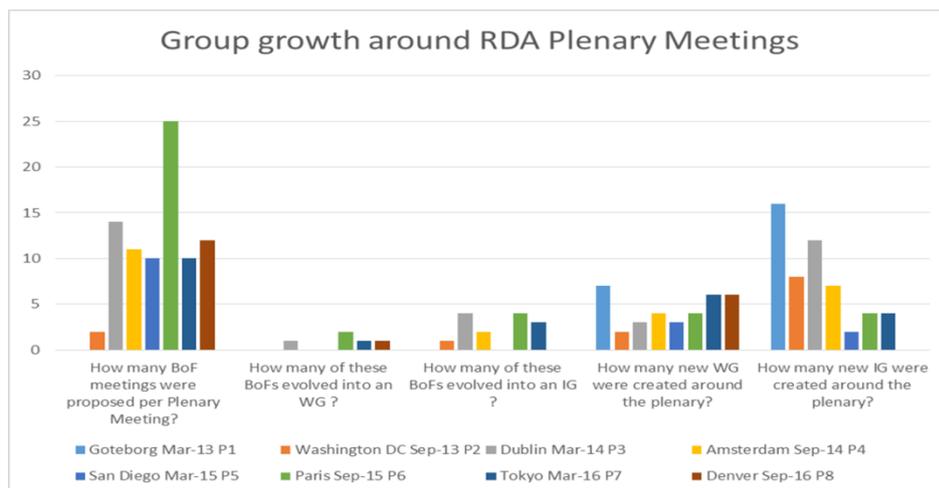


Göteborg:	March 2013
Washington:	September 2013
Dublin:	March 2014
Amsterdam:	September 2014
San Diego:	March 2015
Paris:	September 2015
Tokyo:	March 2016
Denver:	September 2016

The diagram above indicates the participation in plenaries and in RDA group growth. The plenary with highest participation was in Paris with slightly more than 700 participants followed by the Amsterdam meeting (550), indicating high mobilisation of data practitioners in Europe. The participation development for the US meetings (Washington, San Diego and Denver) shows that the mobilisation of US data practitioners grew at a slower rate. The Denver meeting had a comparatively high participation, also due to the fact that it was co-located with SciDataCon2016 (CODATA/WDS). The first RDA plenary meeting outside the Northern hemisphere took place in Tokyo in March 2016 and expectations on the number of European practitioners was lower than normal, but European attendance in Tokyo shows that our extra measures to support travels to Tokyo were successful.

An important indicator for RDA engagement is the development of groups in their 3 forms:

- Birds-of-Feather sessions to indicate new challenges
- Interest Groups to discuss about challenges and eventually produce outputs
- Working Groups to come up with concrete recommendations



From the diagram it can be seen that

- The number of WGs increased continuously after a dip after the initial plenary in Göteborg. In particular we could see that many of the new WGs have been set up by young experts which shows that the mechanisms within RDA are understood.
- Since IGs are organised around broader challenges and last much longer it is expected that the number of new IGs will decrease over time.
- The number of BoFs remains constant if one would ignore the peak for the Paris plenary. It is hard to predict how many of the BoFs will be turned into groups since the rationale behind organising BoFs can be different. They demonstrate the usage of the RDA as a discussion forum.

### Impact

From these statistics we can conclude that

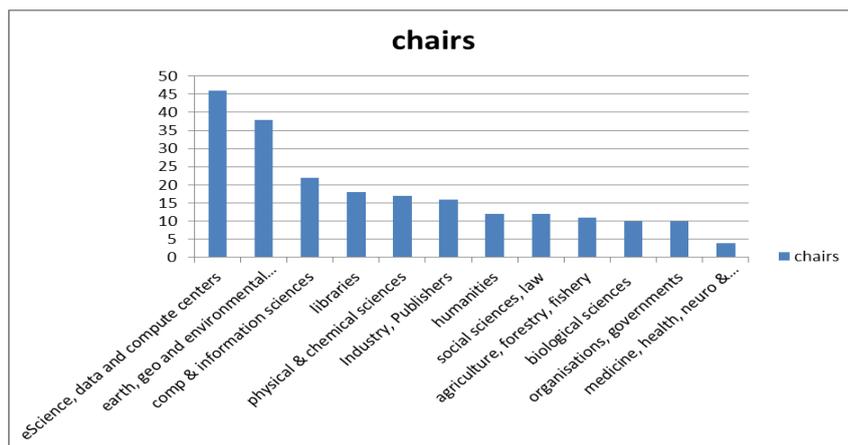
- *The core of RDA work, achieving a dynamic and goal driven interaction between data practitioners, is not only working, but improving over time.*
- *There are sufficient data practitioners motivated to work together in groups to overcome hurdles they meet in their practical work.*
- *Despite the 'energy' and the clear challenges that created the initial WGs and IGs, new topics continue to emerge – the community sees RDA as a forum for addressing an array of challenges.*

*In addition, there are overarching groups that bring a variety of results together to come to larger messages.*

### 3.1.3 Chairs Engagement across disciplines

Another interesting task is to understand which communities are engaged as active chairs in RDA. Here there is no distinction between chairs from Europe and other regions since most of these communities are globally connected and interact about group leadership in many cases. As a global alliance, RDA requires that group chairs include a regional balance.

The biggest group of chairs is from eScience/data/compute centres since they are in many cases the mediators between the scientists and the data challenges and work on the infrastructure solutions. The same holds for the computer and information scientists and librarians with slightly different foci which follow as third and fourth group. From the scientific disciplines the group of earth, geo and environmental scientists are the most active, since these scientists need to solve huge interoperability problems worldwide.



*This diagram shows from which community the roughly 216 co-chairs of the (84) RDA working and interest groups come. It indicates broad engagement.*

Also a number of other relevant scientific domains are fairly active, but it is obvious that some major domains such as engineering are not engaged yet.

### **Impact**

*From this overview we can conclude that*

- In fact all major discipline clusters are engaged as chairs and thus help overcoming barriers for efficient data use in their domain and others.*
- In particular cross-disciplinary acting sciences and institutions (computer science, libraries, eScience/data/compute centres) are highly active since they are in many cases the mediators to the research disciplines.*

### **3.1.4 Estimation of Personnel Effort in RDA**

The following table created by RDA Secretariat gives an estimate of the personnel effort spent in RDA work. Most of the work is invested by volunteers, only Secretariat work and some TAB and council members are funded by projects supporting RDA (about 10%). This implies that more than 60 FTEs are spent by RDA volunteers in their free time or based on agreements with their institutions.

Most of the work is done by working and interest group chairs who take responsibilities. However, even followers need to read messages and documents, participate in virtual meetings, come to plenaries etc.

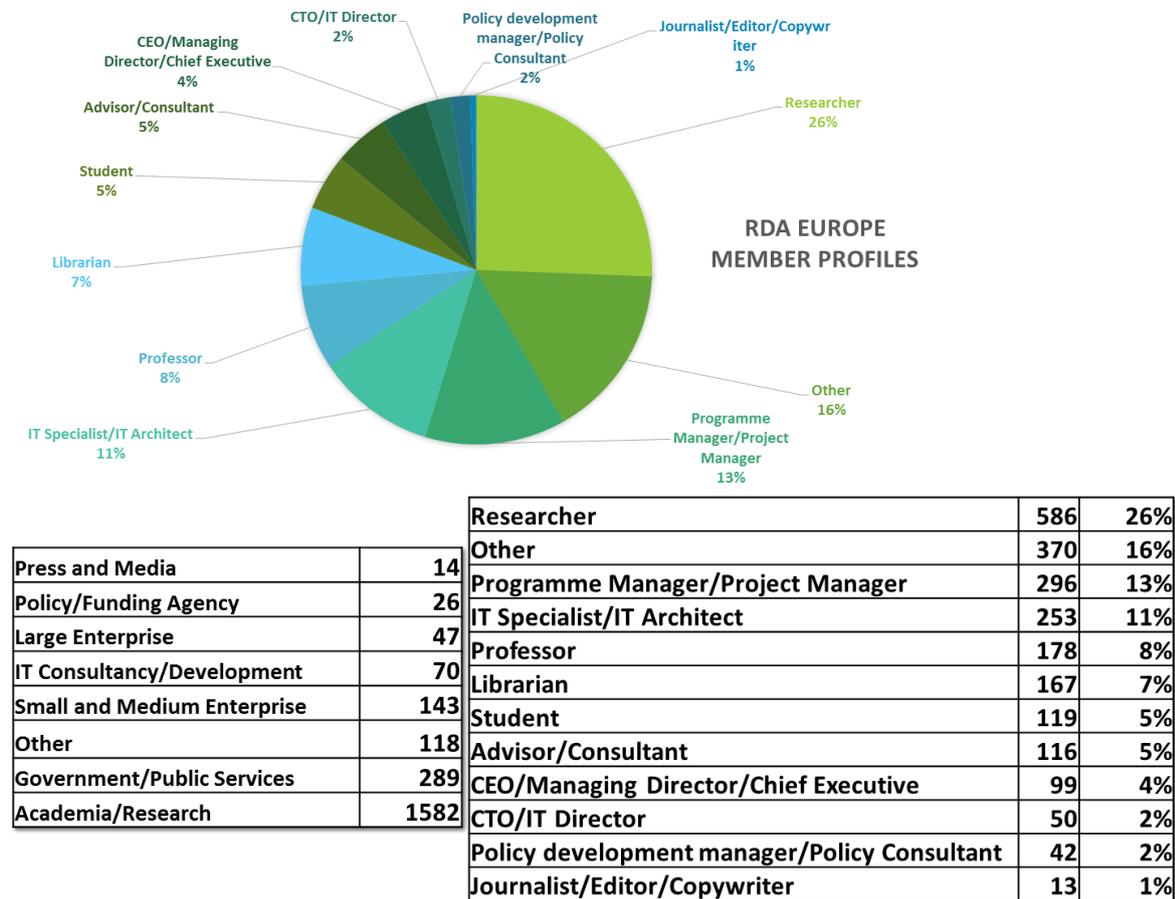
October 2016 Total Personnel (including volunteers)			
	People/Groups	Est. FTE	Formula
Secretariat	9	5.5	actual
TAB	12	1.4	0.2 FTE for chairs, 0.1 for other members
OAB	10	1.4	0.2 FTE for chairs, 0.05 for other members
Council	9	0.9	0.2 FTE for chair, 0.05 for others
WG Chairs	25	12.5	0.5 per group
IG Chairs	42	10.5	0.25 per group
Regular members	4500	43.3	1/2 week per person per year
Total		75.4	

## 3.2 RDA Europe

In this chapter we will discuss statistics on European membership.

### 3.2.1 RDA Europe Membership

The statistics presented here are not so different from those shown in 3.1.1.



Most of the European members are from academia/research and classify themselves in different categories. The second largest group is from governmental and public service institutions followed by a group of experts from different types of companies.



With the exception of Dutch experts most members come from countries that are also engaged as partners in the RDA Europe project indicating the enormous relevance of having ambassadors who

have funds to dedicated effort to creating national interaction platforms. From the tables it can be clearly seen that engagement with the Eastern European countries must be intensified.

### Impact

From this overview we can conclude that

- Those countries that have a member in the RDA Europe consortium have a high number of RDA members. One exception is Netherlands without having a member in the consortium also due to the organisation of a plenary meeting there. In many other countries including Austria, Switzerland, Belgium, Sweden etc. RDA is well-known and people are interested to follow what happens in RDA<sup>15</sup>.
- This recognition of RDA as an interesting exchange platform is known in these countries across the various disciplines as is demonstrated for the example of Germany in 3.4.3 which can be extended to other countries.
- With a few exceptions the eastern countries are not well represented and obviously are not so much interested yet in removing barriers for efficient data work.

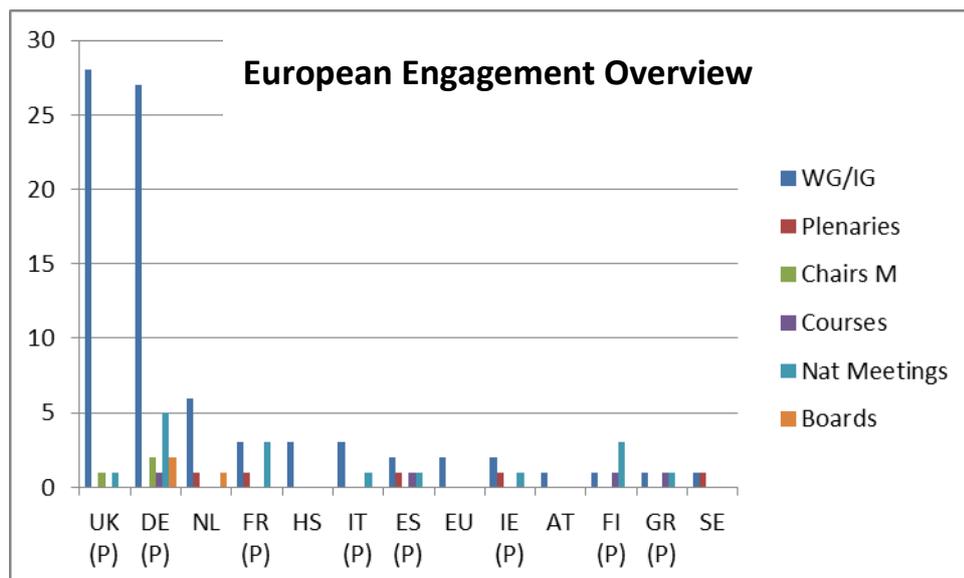
### 3.2.2 RDA Europe Engagement

As indicated in 3.1.1 pure membership does not reflect per se real active engagement in RDA. Therefore it is of interest to use some statistics for Europeans that give a slightly deeper impression about active engagement.

The following table and diagram indicate active participation of European countries in RDA activities contributing with own funds (thus not only using funds provided by the RDA Europe project) such as:

- Acting as WG/IG chairs
- Organising RDA plenaries
- Organising chairs meetings
- Organising f2f training courses and datathons (summer/winter schools)
- Organising national/regional meetings
- Being represented in RDA boards without explicit RDA Europe funding support

A “P” is indicating that the country has a partner in the RDA Europe consortium.



*This diagram indicates which European countries are active in taking over roles which are related with own funds. The work funded by RDA Europe is not included here.*

<sup>15</sup> The cases of Netherlands and Ireland show that also plenaries can have an effect on mobilisation.

## Impact

A few conclusions can be drawn:

- Community representatives in UK, DE and NL are active as WG/IG co-chairs and thus driving best practice work. The active people come from different communities (see 3.1.3).
- 4 plenaries have been organised in Europe with another one to come in Barcelona. The one organised in Paris showed RDA's attractiveness to many European practitioners.
- Most activities happened in countries that are represented by a partner in the consortium. At the Bratislava e-IRG meeting in 2016 a number of discussion threads with eastern countries have been started to come to national meetings.

## 3.3 Outreach Activities of RDA Europe

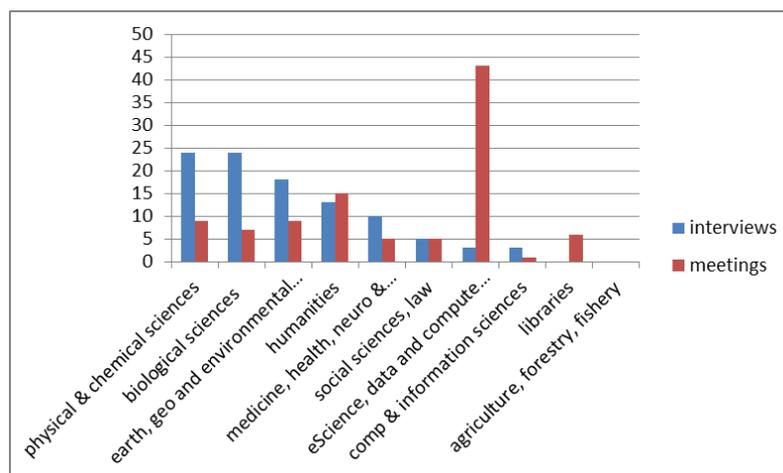
This section illustrates activities where consortium members interacted directly with communities. Events attended by non-consortium parties to present RDA are not included as there are many but to claim direct impact would be incorrect. In the activities listed below RDA EU consortium members were involved either as organisers or active contributors (keynotes, presentations). All meetings have been documented in the respective deliverables.

In all of the tasks, one obvious action is to create trusted, sustainable networks of actors on various levels and fields, with whom to interact on a regular basis. Even though the stakeholder engagement is administratively divided into separate tasks in the RDA Europe projects, it must be noted, that the division into different stakeholder groups, and even within them, can be more or less artificial, as these actors do not operate in vacuums. In the end, the results and impact of RDA activities derive from addressing various stakeholders that also have an effect on each other. Therefore it is important to avoid silos and narrow approaches when addressing different actors, still of course keeping in mind the specific knowledge, expertise and mandate of each actor.

### 3.3.1 Outreach Meetings

#### *RDA EU1 - RDA Europe Analysis Program (November 2014)*

A report, delivered to the EC in Nov 2014, summarising data practices, refers to 24 interviews carried out in RDA Europe, 16 in EUDAT and 9 interviews<sup>16</sup> carried out in a German survey initiative<sup>16</sup>, a total of 119 interactions. The report also refers to 70 community meetings attended by RDA Europe experts to raise and discuss RDA related issues.



#### **RDA EU1 Outreach Overview**

*This diagram shows the distribution of the efforts in the RDA EU1 project to interact with different communities by organising meetings and interviews.*

Statistics related to the interviews shows a proper distribution across all major scientific domains while statistics about meetings showed a strong focus on meetings that had a cross-disciplinary

<sup>16</sup> Both initiatives allowed RDA to make use of their material.

character, i.e. meetings that were organised by the major e-Infrastructures in Europe where e-Infrastructure experts and experts from scientific domains came together.

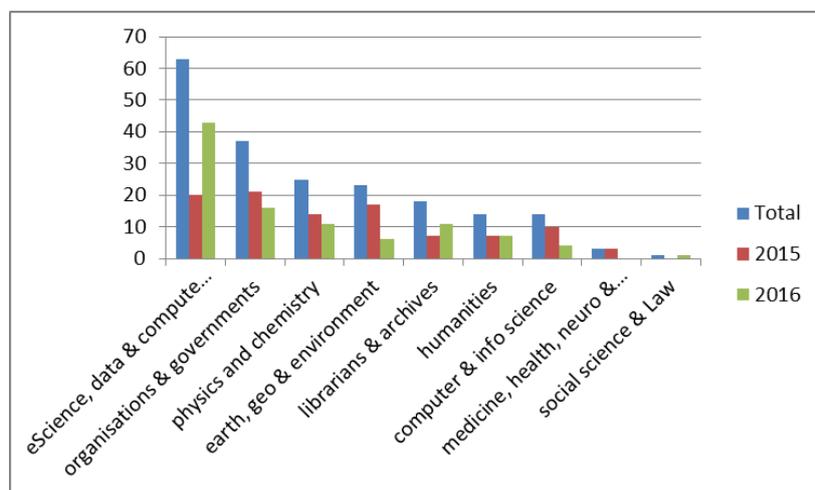
In its observations and recommendations covering a period from Sept. 2012 until Sept. 2014, the report<sup>17</sup> mentions a number of important topics to be addressed such as:

- the need for open access to data, proper metadata, use of persistent identifiers,
- the need to build collections as basic management units,
- the need for policy based management,
- the need to create federations, and
- the relevance of data publishing.

### **Domain researcher outreach in 2015/2016 (RDA EU2/3)**

In 2015 about 29 community meeting participations were documented while in 2016 the total was 70. As a result of the developments around Open Data, Open Science, Big Data a new category for cross-disciplinary meetings should be added, since these topics address experts from various different communities. In particular in 2016 the number of cross-disciplinary events (co-) organised by RDA EU, which are included in the "eScience/data/compute centre" cluster, increased by more than a factor of 2.

One example of such a cross-disciplinary event is the e-Infrastructures & RDA Workshop on data intensive science held on the 22 September 2015 co-located with the 6<sup>th</sup> Plenary where experts from the Human Brain Project, environmental infrastructures, the global linguistic community, the Big Data & High Performance Computing projects and experts engaged in open scholarship discussed in 5 streams about the common goal of enhancing collaboration on a global level in support of research data sharing. This meeting attracted over 400 experts from different scientific disciplines and countries. Despite these different backgrounds agreed outcomes in terms of action points that can be clustered into five main areas were achieved: Collaboration, Integration, Openness, Sustainability and Trust.



### **RDA EU2/3 Outreach Overview**

*This diagram shows the distribution of the efforts in the RDA EU2/3 projects to interact with different communities by organising meetings.*

### **Policy Level Outreach in 2015/2016 (RDA EU2/3)**

Meetings with policy stakeholders increased over this period as a direct result of new action lines such as Digital Single Market, EU copyright regulation, Open Science and European Open Science Cloud. Some are highlighted below:

<sup>17</sup> Data Practices Analysis: <http://hdl.handle.net/11304/6e1424cc-8927-11e4-ac7e-860aa0063d1f>

A major activity in RDA EU2 was the "[Open science, open innovation, open scholarship and open infrastructures – insights to the European Digital Single Market](#)" event, organised in collaboration with EC DG CNECT and hosted by MEP Henna Virkkunen (FIN) and MEP Catherine Stihler (UK), in the European Parliament (EP) 10 November 2015. 40 participants, including a member of the Cabinet of Commissioner Günther Oettinger, from the EC, EP, Committee of Regions, European e-infrastructure and library actors, national and other stakeholders, joined and called for a set of actions and interventions from the European Parliament in order to support more jobs for Europe, improve democracy and citizen power and exploit innovation. It was agreed that open science and open innovation IS and MUST remain a free and global market, and in terms of regulation, it is wiser to proceed in small steps

To raise the awareness of the current policy environment, a High Level European policy meeting on funding research data to support open innovation, "[Weaving the Internet of Data](#)", was organized in Amsterdam on 6 April 2016. Over 60 funders and other experts in data creation, curation and re-use from all over Europe attended this meeting to discuss and address issues regarding how research data in general, and the RDA activities in particular, are and should be funded in the future to ensure a convergent and coherent approach.

In this phase also the interactions with e-IRG and ESFRI were intensified to collaborate on strategic data policy issues such as interoperable/reusable data services across communities and the EOSC, and sense about prioritisations. Amongst others this led to

- an agreement to closely synchronise activities between RDA and eIRG regularly;
- to set up the Group of European Data Experts (GEDE) from large infrastructures;
- to add the chairs of eIRG and ESFRI as observers to the GEDE process
- to set up participation of European RDA members in e-IRG Task Forces relevant to their expertise.

### **Impact 3.3.1**

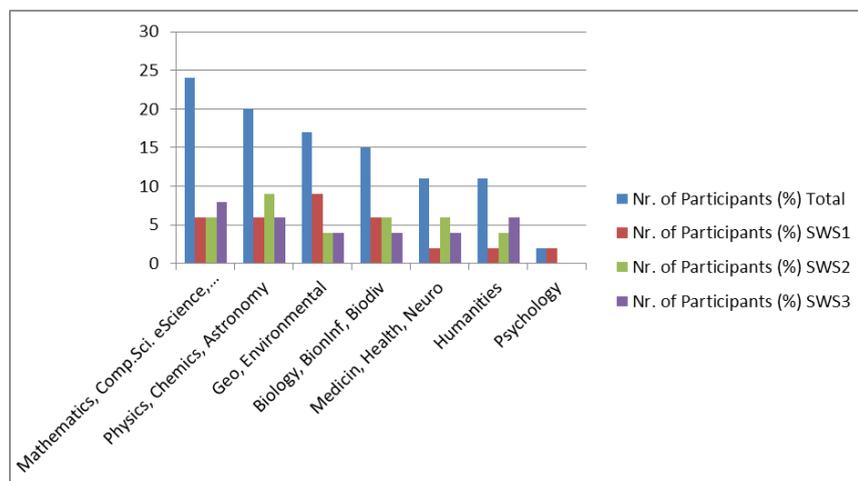
*Based on this report, we can draw the following conclusions:*

- *RDA started with in average about 47 community meetings in the first two years, had a dip in 2015 with about 29 meetings and an increase again with 70 meetings in 2016, i.e. in average RDA organised or participated in about 50 meetings per year presenting RDA or RDA related matters.*
- *RDA Europe created a well-respected report on data practices in the various communities at the end of the first phase after many intensive bilateral and meeting interactions. RDA activities are focusing around the major topics that came out as focus areas in this report.*
- *Many meetings focused on generic topics with a cross-disciplinary audience, but in fact interactions with all major scientific domains were carried out. The fact that there were almost no specific meetings with "biological sciences" in 2015 and 2016 is due to the fact that these communities are now very active with respect to tackling data issues.*
- *eScience, data and compute centres as well as libraries and archives play the most important role as mediators.*
- *The policy level meetings changed their content with the focus on the new challenges such as Digital Single Market, Open Science, Open Data and European Open Science Cloud which dominated the policy level interactions in 2016 and how RDA can support these areas.*
- *RDA Europe performed broad outreach to most of the scientific communities with the goal to make the experts aware of new trends and new results in the data domain and to act as an interaction platform to exchange opinions and experiences. This impact is high, after 4 years of hard work in fact almost all communities agree with the basic principles and models. The German council for information infrastructures recently held a meeting together with experts from UK and NL where one of the most important conclusions was that Germany needs to support platforms for a much more intensive interaction about all aspects of data. At the*

European level and also in some countries RDA Europe stimulated exactly such intensive dialogues.

### 3.3.2 Science Workshops

RDA Europe organised 3 Science Workshops in collaboration with relevant research organisations: 2014 at MPS in Munich, 2015 at CERN in Geneva and 2016 at CNRS in Paris. Together with invited guests there were 54 participating scientists from different scientific areas.



#### **RDA EU Science WS**

*This diagram shows the distribution of participation in the Science WS to which leading scientists are invited.*

The largest group of participants came from the fields of mathematics, computer science, eScience/compute centres and libraries (24%) – thus cross-disciplinary operating experts. The second largest group was from classical fields of physics and chemistry (20%). The third largest group from geo and environmental sciences (17%), followed by researchers from biology, bio-informatics and bio-diversity (15%), then experts from the areas of medicine, health and neuro science (11%) and finally humanities (11%). Fields such as social science (2%), engineering and agro-sciences have not been well represented until now.

#### **Impact**

- *RDA EU organised small, focused workshops with (~15) selected key scientists from various fields enabling an intensive dialogue. This strategy needs to be assessed for future phases.*
- *These selected number of researchers have heard about RDA, took notice about trends and results to improve data management, access and reuse, and provided strategic input and feedback to RDA. These are issues this group of scholars normally delegate to their co-workers.*
- *Changes in the various disciplines with respect to data management during the last few years is obvious and the Science Workshops facilitated this development.*

### 3.3.3 National/Regional Meetings (Overview)

The organisation of national/regional meetings have proved to be strong mechanisms to reach out, transmit the RDA results and listen to the relevant problems. The following aspects are of relevance:

- All data practitioners are not in a position to attend RDA plenaries, therefore it is important to organise interactions also at national level to ensure broad outreach and awareness.
- For cultural and language reasons national meetings make it much easier to conduct a fruitful interaction and to interchange essential messages. Due to the higher density of the interactions it is easier to make use of each other's knowledge and expertise.
- National meetings ensure the integration of national funders and representatives of national research organisations which is important for future sustainability.

As can be seen below (chapter 3.4.3) national meetings also allow engaging different communities such as engineering and education that are not well represented at plenaries or in RDA groups. This is often mediated by personal contacts and by the recognition of active people as experts.

### 3.3.4 Industry Meetings

The main focus of RDA in the first years has been the academic and research fields, however private sector engagement has, to a much lesser degree, been a focus too. Engaging big IT companies will prove difficult since they have a primary interest to impress their best practices globally to achieve a high return of investments and they will only invest in new solutions if they can improve their market presence. Yet RDA results are not of that sort that industrial practice can take profit.

Therefore, the initial private sector activities have focused on close interactions with SMEs that related with communities active in RDA or that are operating in the area of data analytics. Both groups could take profit from improved interoperability suggestions emerging in the scientific communities when re-using scientific data for their businesses. Therefore, one network has been established by Barcelona Supercomputing Centre engaging with SMEs working in the domain of climate and weather research where data is being used for generating services. The other network has been established by CAP Digital with French SMEs active in the area of data analytics where scientific data is being used for various services.

Recently, in the US and in Europe discussions were started with big industry in the area of Internet of Things (also Industry 4.0). Companies discovered the relevance of data for various purposes such as optimising industrial processes and a large amount of cloud solutions have been established tailored to the particular needs of the companies. In all these cases one of the basic cloud services (Amazon, Microsoft, etc.) has been taken and amended by a virtualisation layer where metadata is being used to implement a useful data organisation. Now companies are at the step to discover

- the relevance of improved data solutions since the many devices for example embedded in a car are generating unforeseen amounts of data to be dealt with and
- the relevance of cross-cloud solutions, i.e. to interlink the various cloud systems, leading to similar interoperability solutions as in science.

Interaction with big industry have directly let to the organisation of two events:

- a pre-plenary event in April 2017, Barcelona to start interacting with each other and to understand the challenges and the potential of solutions.
- a workshop at the IoT Week in June in Geneva to get into a deep interaction about solutions.

#### **Impact**

- *SME networks create awareness about data issues and can be seen as a platform to exchange knowledge and experience which is crucial for SMEs if they want to succeed in offering data services. Currently, it is still difficult to find useful open data that can be combined to develop new business models. However, RDA Europe helped to establish first communities and is currently developing specialised training courses that could even more facilitate this work.*
- *Even big companies are now at the point where they realise that if they want to remain competitive they need to link data that is in cloud silos, i.e. they realise that they run into the same difficult problems to be solved as we find them in science. Intense interactions and coming joint workshops show the important role of RDA as catalyser.*

### 3.3.5 Outreach beyond Europe

Many scientific communities are organised globally such as geo and environmental, health and agricultural sciences and hence are dependent on data from all regions of the world. RDA has been accepted by many as a global and neutral place for interaction, thus it makes sense to engage data

practitioners from all over the globe and here in particular from Africa. RDA Europe co-funded therefore a few activities:

- a joint summer-school together with CODATA about transferring knowledge in state-of-the-art data practices
- a joint workshop together with DIRISA (South Africa) bringing together about 50 researchers from South-Sahel countries.

Most of these researchers come from scientific areas that are of particular interest for developing countries: environmental, geo, health and agriculture. Computer experts, IT specialists and librarians are also being engaged as mediators.

In addition RDA Europe

- contributed to the CODATA-UNESCO meeting on data sharing in Nairobi (2015) as a preparation meeting for the workshop in South-Africa,
- was and is active in interacting with Chinese policy makers and practitioners with the effect that many Chinese practitioners now come to plenaries and that probably one of the next plenaries will be organised in China
- had interactions with experts from South American countries to engage them in RDA.

#### **Impact**

- *It is important for RDA to engage researchers and data practitioners from all over the world in finding solutions to make data work more efficient and to improve data sharing.*
- *The differences in speed in the various regions need to be respected.*
- *RDA Europe is a major driver in RDA's international outreach strategy.*

#### **3.3.6 RDA Ambassadors**

The liaison between the RDA and specific communities, in particular but not only scientific communities, can be ensured by the participation of data practitioners from that community in the RDA activities. Even people playing an observer role in the RDA (Section 3.1.1) can efficiently disseminate knowledge of the RDA activities and outputs into their professional circles. But people more engaged can play a more developed Ambassador role in their communities. The establishment of RDA Ambassadors has been recognised in the RDA Global Future Direction Plan as an efficient way to engage communities with the RDA and is currently being put in place. RDA Europe is currently working at defining the RDA Ambassador role, in liaison with the RDA Council Engagement and Communication Subcommittee, by identifying the Europeans who are currently operate as unofficial Ambassadors. Two examples can already be documented through RDA Europe activities, and details are provided for two communities at different stages with respect to data sharing below: digital humanities and astronomy.

##### **Astronomy Community**

One example of the dissemination of information about the RDA in the astronomy community which has been at the forefront for data sharing and reuse and is already well organised, with annual meetings which gather the world wide data provider community (ADASS - the Astronomy Data Analysis and Software Systems Conference Series, which had their 26<sup>th</sup> edition in 2016) and a global organisation in charge of defining the disciplinary facet of interoperability standards (the so-called Virtual Observatory framework), the International Virtual Observatory Alliance (IVOA), which meets twice a year. In this case the Ambassador role has global, European and national aspects. The first RDA presentation in an IVOA meeting was by RDA Europe representatives, including the person who now plays the ambassador role, in Heidelberg in May 2013, just after RDA creation. RDA continues to be a standing item in the 6-monthly meetings of the IVOA, both in Board and technical discussions. It was also presented in the two last ADASS meetings, in meetings of the European-funded projects which support European participation in the Virtual Observatory (leading for instance to participation

of astronomy representatives in the Munich PID Workshop in 2016), and in national meetings in France, from which the RDA Europe “disciplinary ambassador” comes from. The lead of the VAMDC Collaboration Projects selected in 2015, who is also participating in the Virtual Observatory, is another efficient ambassador of the RDA Output he is implementing. The work of the IVOA on its Provenance Data Model is also for instance discussed in the relevant RDA IG. Astronomy does not feel the need to set up a disciplinary RDA Group, which would duplicate the IVOA, but people from that community regularly participate in the RDA Plenaries and would like to be proactive, with representatives from other disciplines, to create an Interest Group gathering the different disciplines present in the RDA.

### **Digital Humanities Community**

Another example is the effort to more fully engage with the Digital Humanities community in Europe. Unlike in the Astronomy community, where data sharing and reuse is well organised, Digital Humanities scholars do not have well developed practices in this area. In many cases, scholars come from traditional humanities departments, and are not even consciously aware that they produce data worth stewarding for dissemination and reuse. However, the community is very large, and the annual Digital Humanities conference attracts nearly 1,000 participants, so it provides an excellent forum for reaching Digital Humanities scholars. Awareness of data sharing is growing, partly because Digital Humanities scholars engage with archival data in digital repositories, and are increasingly working with large volumes of data, which requires developing best practices in data management. This past July 2016, RDA Europe organised two events for DH2016 in Krakow. First, was a panel of experts from a range of sectors who spoke to best practices in Digital Humanities research data management. The panel included the director of an HSS repository, a member of the Digital Curation Centre, a librarian who works with research data and a digital humanities scholar who is advancing the practice of digital archiving. The second event was organised in conjunction with RDA US, via the RDA/ADHO liaison. This was a hands-on workshop, focused on paving the way for adoption of two RDA outputs: Data Types Registry, and Persistent Identifier Types. RDA Europe provided two scholarships for postgraduate students working with these outputs to attend.

#### **Impact**

- *RDA and RDA Europe started to implement the ambassador program and it can be seen in a few communities how important this role is in improving awareness of the challenges and possible solutions worked out by RDA and others. Important is that the ambassadors are anchored in the community interactions and work.*

### **3.3.7 Technology Working Meetings**

Two technology working meetings were held until now with the intention to bring together computer scientists, RDA experts and where applicable industry experts with the goal to assess the work of RDA groups, where possible to achieve improvements and to also identify new areas to be tackled within RDA.

#### **Impact**

- *These Working Meetings improved the engagement of networks of computer scientists and made them aware of RDA activities.*
- *It is too early to finally conclude if they will lead to new activities and more engagement<sup>18</sup>.*

### **3.3.8 Webinars**

RDA Europe uses webinars to disseminate RDA results and other related aspects on data. After some testing it seemed to be feasible to organise two webinars per months. A few observations can be made:

---

<sup>18</sup> It should be noted here that some statistics reported in this document show that computer scientists are actively involved already in some activities - for example as chairs of RDA groups.

- The number of registrations is much higher than the number of participants and the scientific background of participants is widely un-known and not predictable.
- Webinars are not suitable to create communities of usage, but they are useful in disseminating facts and results in particular also to individuals.
- Their impact is difficult if not impossible to quantify.'

**Impact:**

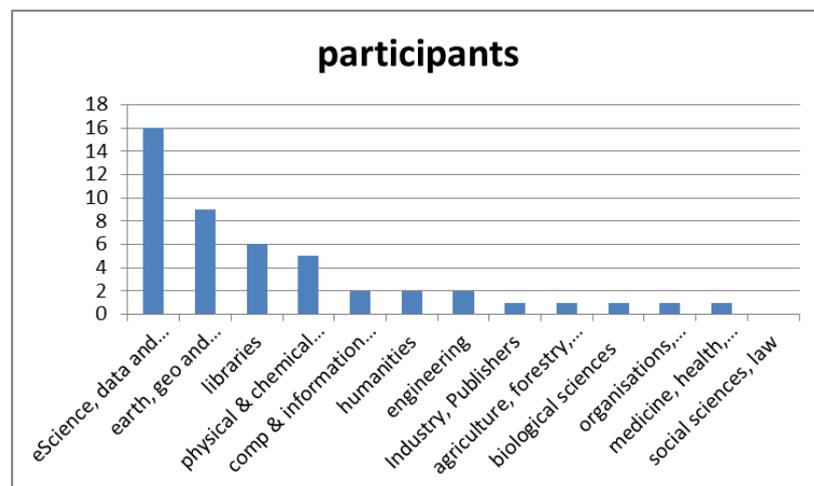
- *Despite its unclear impact RDA Europe will continue to organise the webinars to reach out to data practitioners which are hard to contact otherwise.*

**3.3.9 F2f Training Meetings (training, SS/WS/Datathons)**

In contrast to webinars, face-to-face meetings better channels to come to an intensive exchange of ideas and to transmit details in particular when the training is combined with hands-on sessions. The effort to organise such meetings is much higher and thus limits their frequency.

RDA Europe 3 organised 2 events in 2016, is planning 3 further meetings in 2017 and helped organising training courses associated with national and community meetings. The following events can be mentioned:

- A 2 day datathon was organised in Greece in close collaboration with the agriculture groups in RDA attracting about 40 participants mostly from that community. The way it was organised was successful and got the attraction also from international organisations such as FAO. It is in discussion whether this course can be repeated in another place or the idea can be reused for other communities/disciplines.
- In Germany every year special 2-day training courses including hands-on opportunities and extensive training sessions at annual conferences are organised reflecting the interests of the participating communities (an overview about participating communities can be seen in chapter 3.4.3). Experts giving the courses come from Europe and in specific cases also from the US.



**RDA EU PID Meeting in Munich/Garching**

*This diagram shows the distribution of participation in the PID training and workshop held in September 2017 in Garching.*

- One special training course on Usage of PID Systems was organised in Munich in September 2016 and combined with a workshop about the same topic motivated by the discussions at the Paris Science Workshop. The following communities were represented at this training course and workshop: computer science, climate modelling, neuroscience, eScience/data/compute centres, humanities, medicine/health, astronomy, environment, oceanography, engineering, libraries.
- A 2 day training course will be organised in Spain in April 2017 addressing the needs and interests of the climate and weather scientists.

- Together with ENVRI+ RDA Europe is preparing a 5 day datathon which will take place in June 2017 in Helsinki. The course will be set up using data from environmental observations. The modules are designed in a way that they can be repeated with other communities. Preliminary discussions are ongoing with the biodiversity and humanities communities.

#### **Impact**

- *The impact of these meetings is extremely high since deep discussions about challenges and solutions were facilitated which have a clear effect on the community discussions.*
  - *Example1: The datathon in Athens got the attention of FAO and will be repeated.*
  - *Example2: The PID meetings in Munich had as result a wide agreement on the use of PIDs and PID systems (Handles, DOIs) which will foster solution finding in the participating communities and beyond.*
- *The preparation of future events in collaboration between RDA and the chosen community requires very close interaction with the communities. Environmental experts will learn how RDA results can change practices, can be combined and can be applied to their concrete data.*
- *It is still a separate and cost intensive step to develop reusable training modules that can be used in different contexts and communities easily to repeat the training. Efforts will be made to achieve this goal which would then have much broader effects.*

#### **3.3.10 Collaboration Projects**

RDA Europe 3 accepted 7 collaboration projects in its first round which started working in 2016. The following communities are involved:

- The EURO-ARGO team (oceanography) is adapting its data workflows to adopt a common citation model as recommended by RDA:
- The agriculture community is implementing the recommendations on metadata interoperability which were made by RDA's Wheat Working Group.
- The Digital Curation centre (eScience centre) is implementing a direct connection between its DMPOnline tool and the metadata standards directory as specified by the RDA group with the same name.
- The CLARIN and the Perseus project colleagues (humanities) want to better integrate collections by adopting systematically the results of the Data Foundation & Terminology Group in RDA and thus create a basic level of interoperability.
- The VAMDC (Virtual Atomic and Molecular Data Centre) community want to adapt their database such that it is compliant with proper citation recommendations as suggested by the Data Citation WG in RDA.
- The OpenAIRE community (library/eScience) wants to make its Data-Publication Linking Solution more professional by fully adopting the results of the Publishing Data Services Working Group.
- The Armenian eScience centre wants to adopt the Data Foundation & Terminology group results by establishing a national infrastructure that allows researchers to assign Persistent Identifiers to all digital objects.

A second call has been launched and RDA Europe received 28 applications from a wide variety of communities. Approximately 8 proposals can be funded with the budget available. We can summarise that the following communities received small collaboration grants or have shown their interest in participating in the second round:

	1 <sup>st</sup> Round	2 <sup>nd</sup> Round (not decided)
eScience, data & compute centres	2	3
Libraries	1	2
Computer & information science		2
Earth, geo & environmental science	1	5
Physical and chemical sciences	1	3

Biological sciences		4
Agricultural sciences	1	2
Medicine, health, neuro & cognitive sciences		2
Humanities	1	2
Social sciences, law		3

In the first round we could see many questions about intentions and practicalities of the call. In the second round we could see a broad interest which basically covers all areas of sciences. The RDA call was advertised in all scientific communities and responses show there is significant interest in testing or adopting RDA solutions.

### Impact

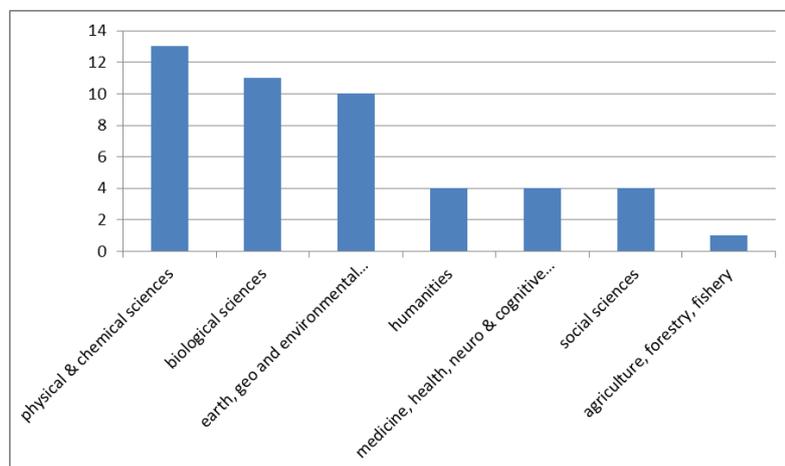
- *The impact can be compared with those of summer schools in so far as selected communities can actively deploy RDA results and can see the advantages of the emerging solutions.*
- *The disadvantage is that only selected projects with few communities can be funded and that only small amounts of funds are available not allowing developing larger testbeds with combining different results.*
- *The adoptions can be turned into success stories to be disseminated to help other communities.*

### 3.3.11 GEDE Group (Group of European Data Experts)

RDA EU needs mechanisms to reach out to and interact with the experts building the infrastructures in Europe, since we cannot expect everyone coming to plenaries. After discussions with e-IRG, ESFRI and research infrastructures the idea grew to create an interaction platform moderated by RDA Europe. In April/May 2016 about 60 infrastructure projects (based on interactions with the EC) were contacted by email. From these 45 reacted within the following weeks positively and 2 more joined after the Washington RDA WG/IG chairs meeting.

At this time these 47 large European infrastructure projects (mostly ESFRI projects) have declared to participate in this information platform, the so-called GEDE (Group of European Data Experts) and have nominated delegates. These are ACTRIS, AGRO, AnaEE, ASTERICS, BBMRI, CESSDA, CLARIN, DARIAH, DiSSCo, EATRIS, ECRIN, EISCAT, ELI, ELIXIR, eLTER, EMBRC, EMFL, EMPHASIS, ENES, ENVRI Plus, EPOS, E-RIHS, ESRF, ESS, ESS – Social, EST, EUCALL, EU-OPENSREEN, EURO-ARGO, EURO-BIOIMAGING, FAIR, HBP, IAGOS, ICOS, INSTRUCT, KM3NeT, LifeWatch, METROFOOD, MIRRI, MYRRHA, NIDI, PARTHENOS, SCK-CEN, SIOS, SKA, VAMDC.

These GEDE members can be categorised as shown in the diagram (for details see appendix A).



### GEDE Membership Overview

*This diagram shows the distribution of interest in participating in the GEDE platform.*

The diagram indicates that

- all major scientific domains are represented in GEDE
- the participation is correlated with the number of funded projects in the respective scientific areas

### Impact

- *Already now GEDE has shown that it is an excellent cross-disciplinary platform for intensive interactions between the practitioners active in the different infrastructures and thus fills an essential gap in the European landscape.*
- *It is widely complementary to the plenaries as all practitioners are not able to go to plenaries, nevertheless want to have an interaction platform, in addition better focused on infrastructure practices and needs.*
- *Not all communities are equally interested in such platforms since they have active global interactions within their domains.*
- *Interaction platforms such as GEDE may not be eternal, but must react in agile ways, i.e. sense how the focus of the major challenges will evolve over time and look for appropriate ways to meet the needs. RDA Europe can and should take the moderating role.*
- *The national meetings in several countries also have a similar and utterly important role.*

### 3.3.12 European Organisational RDA Members

Part of community outreach can also be identified by looking at the organisational members that so far joined RDA. 20 from the total 43 organisational members come from Europe. The 20 members come from various sectors.



### European Organisation Members

*This diagram shows RDA's current organisational members coming from Europe.*

In the coming months, an active interaction will be started to convince more initiatives and potentially interested companies. Only now the role of the organisational members as possible incubators and adopters has been identified clearly enough allowing a much more pro-active attitude from RDA Europe.

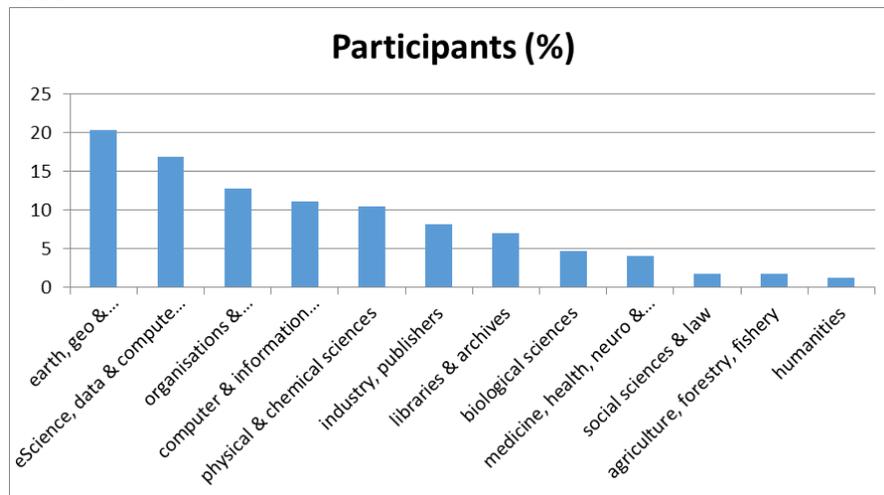
### Impact

- *Now that the role of organisational members in RDA has been defined much more clearly as possible incubators and adopters, we can state that their activities will have a high impact on shaping the landscape, driving RDA and disseminating RDA results into the respective communities.*

### 3.3.13 European Participation in Denver Week

For the International Data Week in Denver including the RDA plenary we made a re-categorisation of the European participants knowing that there was some fluctuation between the two meetings. From the 172 European participants most came from the earth/geo/environmental community and

the eScience/data/compute centres. Many members of organisations and governmental/public services came to join the first part resulting in a comparatively high participation rate. Communities with comparably low budgets such as social science and humanities were not as well represented as usual.



### **European Participation in the Denver Week**

*This diagram shows the distribution of participants that came to the International Data Week in Denver from which most also participated in the RDA plenary.*

### **Impact**

- *Despite the large effort in cost and time it can be seen from the distribution that all scientific domains are represented at such plenaries. For the European plenaries we expect to find a more even distribution.*
- *Plenaries are indeed the cross-disciplinary interaction platform that RDA wanted to achieve and RDA Europe succeeds to mobilise the data professionals to participate.*

## **3.4 National/Regional RDA Meetings**

### **3.4.1 Finland/Nordic**

The national approach is based on the idea to have at least one national RDA event per year, and preferable after an RDA Plenary, to inform and encourage the Finnish research community to engage. From the beginning, it has also been clear that a stronger regional (Nordic) effort is needed.

The Nordic Council of Ministers (NCM) is the official inter-governmental body for co-operation in the Nordic Region. Finland has been the chairing country for 2016. One of the Finnish themes for 2016 has been the Open Science with the Nordic Open Science and Research Forum 2016, organised on 21-23 November. The Open Science and Research Forum 2016 had two events closely co-organised by RDA / RDA Europe. “The RDA and the Nordics – viewpoints on open science” was organised on the 21 November (<http://openscience.fi/rdanordic-workshop>) with a strong emphasis on strengthening the Nordic RDA dimension. This was also a follow-up on the many RDA Finland events already organised through the RDA Europe projects. The event on the 23 November “Nordic Workshop on Data Citation Policies and Practices: How to Make it Happen?” was co-organised with the Finnish Committee on Research Data and ICSU CODATA (<http://openscience.fi/fcrdcodata-workshop>).

The goal of the “Nordic Workshop on Data Citation Policies and Practices: How to Make it Happen?” was the following:

- Increase awareness about the concept of data citation and the emerging ecosystems supporting it.
  - The event had clearly reached its goals as it had attracted a broad range of people from the academia, publishers, and ministries as well as from the funding agencies.

- According to the lively discussion there is a clear need to have an open discussion on how research data can or should be counted for when defining the research indicators for open science in the university funding models.
- Discuss the concrete actions needed to make data citations a part of everyday research practice in Finland and the other Nordic countries.
  - Examples were presented on what the research indicators could look like for the funding models.
- Support university and funding agency policy makers considering the use of data citation metrics as quality indicators.
  - Results from the RDA Working Group on Data Citation were presented in combination with approaches and challenges in Managing permanent identifiers (PIDs) both on a global as well as national context.

The national impact comes through the close collaboration with the major Finnish Open Science effort (openscience.fi), European e-Infrastructure service provider (EUDAT) and the ESFRI RIs having a strong anchor in Finland either through the coordination or being a major partner. This has created a fruitful collaboration landscape between RDA/(RDA Europe) – European e-Infrastructure Service providers – ESFRI RIs – Finnish Universities.

On the regional (Nordic) level the challenge has been to find a vehicle to promote the re-use of research data. The Open Science effort and the Finnish effort during the 2016 has provided a window of opportunity, which will be used fully over the coming years to strengthen the Nordic impact both in active participation in RDA Working and Interest Group activities as well as in implementing RDA output and recommendations.

### 3.4.2 France

The following activities were performed before RDA EU3:

- four RDA France meetings organised in Paris on 28 April 2014, 25 November 2014, 6 February 2015, 8 June 2015 with around 40 participants, both for people involved at the policy level at the Ministry and in Research organisations and for practitioners. The focus was at the beginning to explain what RDA is and to encourage people to participate in the activities, and progressively included presentations by French RDA members about their activities in RDA.
- A plenary talk and splinter hands-on workshops were presented at the National meeting of research and higher education software developers in Bordeaux, 30 June-1 July 2015, on invitation from the organisers. At this meeting about 500 developers from many different scientific communities participate. This time this was a meeting co-organised by RDA EU, i.e. detailed information about the RDA results were presented and discussed.

In addition, RDA has been discussed with data practitioners in astronomy and earth sciences in their own topical meetings, plus with research organisations, including CNRS, INRIA, and the French Geological Survey (BRGM). The latter led for instance to 6 BoF proposals for the Paris Sixth Plenary in September 2015, to some clarification of possible practical interface between OGC and RDA, and to the successful proposal of a Global Water Integration IG.

For some communities at the national level, RDA is part of their life. The information they got from meetings, and for some of them participation of key people in P6 have been the triggering factor. The Figure below is a slide shown at the last Workshop of the national Earth Science Data Poles in November 2016, in a talk about their recommendations on DOIs. Even if you cannot read French, you will recognize 'RDA' and 'query store' from the dynamic data citation WG (BP means best practices here). DOI-SUFF-04 is that the meaning of DOI fragments should be kept and maintained into query stores as recommended by RDA.

**BP relatives au suffixe**

DOI-SUFF-03 [Recommandé] : les variantes d'un jeu de données doivent être indiquées via des fragments et ne doivent pas donner lieu à la production de nouveaux DOI

DOI-SUFF-04 [Recommandé] : les fragments doivent être concis et non significatifs. Leur réelle signification doit être conservée et maintenue au sein de query stores.

**RDA**

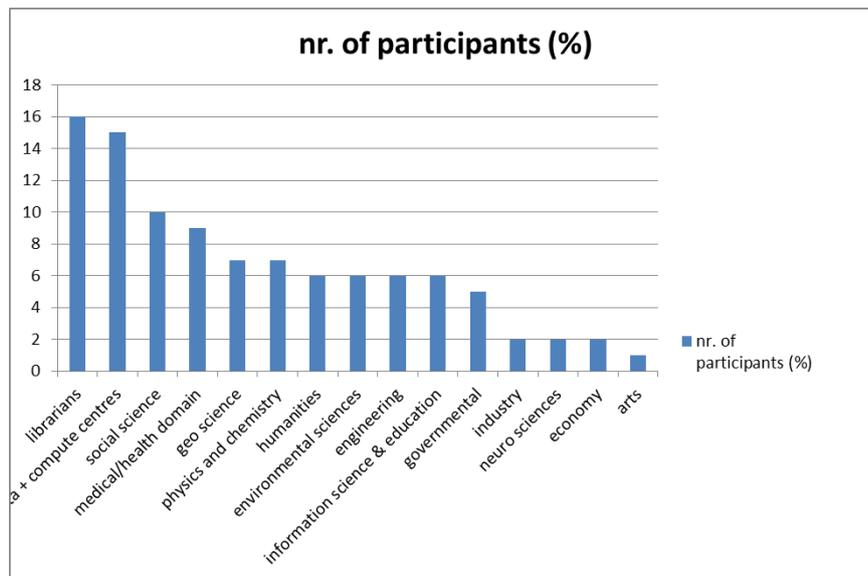
**AERIS** JOURNAUX ET SERVICES POUR L'ATMOSPHÈRE

www.aeris-data.fr | 7

### 3.4.3 Germany

From 2014 two events were organised per year: one training course with about 40 participants and one annual conference with an increasing number of participants over the years:

- 2014: 121
- 2015: 140
- 2016: 165



**Participation in the RDA Germany plenary**  
*This diagram shows the distribution of participants of the RDA Germany meeting in November 2016 in Potsdam.*

We analysed the distribution across disciplines only for the 165 participants of the annual meeting in 2016 assuming that the distribution was stable across the years.

The largest group of participants came from facilitating and mediating institutions such as libraries, eScience, data and compute centres including the large German HPC centres. Also very active in Germany are the social science and medical/health communities facing huge interoperability challenges. In Germany some of the relevant engineering institutions are engaged and delegates from industry are participating.

### 3.4.4 Greece

The ATHENA Research Centre organised two events, one in 2015 and one in 2016. The first event was a regional dissemination event taking advantage of the RDA Secretary General presence in Greece. The purpose of the event was to present the RDA initiative, to investigate the further engagement of South Eastern Europe, Mediterranean and Caucasus regions in RDA and brainstorm about the development of national and/or regional RDA initiatives. Although the vast majority of attendees were from Greece as it was expected, there was participation from other countries in the region,

even without financial support<sup>19</sup>. Experts from 7 countries participated: Greece (56), Armenia, Cyprus, Georgia, Israel, Turkey and US (all 1). Of these participants the vast majority (48) was coming from academia/research, but there was also participation from government and public services(6), SMEs (6) and large enterprises (1), plus 1 from a consultant company and 2 from other sectors. Although detailed discipline statistics were not kept, the majority of the participants were from the ICT domain, i.e. computer scientists and computer science researchers (more than half), but there were also representatives from Agro & Food, Bio & Health, Environment, Humanities and Engineering sectors.

The main outcome of the meeting was the decision to create a regional initiative.<sup>20</sup> The regional mailing list is now composed of around 150 members. The following additional points also arose from the meeting:

External support is needed for several countries in the region; in kind support may not be enough.

The second event was organised in 2016 and as already stated it was a Datathon around the Wheat community. The participation exceeded expectation by far as we expected 15 participants and got 41 registered participants over the two days. The first day comprised of presentations and the second day allowed participants to come with ideas and implement them. The number of participants was limited to around 20 people on the second day, who took part in the hands-on session and were researchers, students and data specialists from the first day. Around 80% of the participants had an IT background, while the rest were students in the agricultural university of Athens. In terms of status, participants were researchers, students, independent data specialists from the computer science, the agriculture and other communities. In addition, there were a few project managers. The winner of the first prize along with the main organiser of the Datathon visited the Food and Agriculture Organisation (FAO) headquarters in Rome. Reports from the Datathon and the visit to FAO are available in the regional webpage<sup>21</sup>.

### 3.4.5 Ireland

The third Plenary of RDA was held in Ireland, and Irish members of RDA number quite highly compared to the population size, suggesting that this kind of exposure helps to grow interest in RDA. Several events have been organised recently in Ireland:

- An RDA national workshop was held at the Royal Irish Academy, with 30 participants including data scientists, repository managers, information professionals and funders. Mark Parsons, Hilary Hanahoe and Peter Fox attended as representatives of RDA. June 2015
- RIA hosted a two day workshop on Open Science with keynote by Barend Mons, Chair of the EOSC, at the Royal Irish Academy. Sandra Collins presented on 'RDA in a Nutshell,' The event was co-organised by Natalie Harrower (Digital Repository of Ireland) and the Health Research Board in Ireland, drawing widely from across health care, Humanities and Social Sciences. It was supported by the ALLEA e-Humanities Working Group, making initial forays into that large community. June 2016
- "Infrastructural support for research information management" panel at HEAnet national conference 2016, with RDA presentation by Rebecca Grant. November 2016
- FAIR Data Stewardship for Open Science workshop, with RDA presentation by Sandra Collins. Jan 2017

---

<sup>19</sup> An exception to this was the participant from Armenia, who received travel funding support. In addition, a small number of the registered participants did not show up.

<sup>20</sup> A full report is available under <http://europe.rd-alliance.org/plenaries-events/events/engagement-in-rda-southern-eastern-europe-mediterranean-and-caucasus-region>

<sup>21</sup> <https://www.rd-alliance.org/group/rda-south-eastern-europe/event/innovating-wheat-community-through-rda-services-and-outputs>

### 3.4.6 Italy

An Italian national RDA workshop on “FAIR Data Management: Best Practices And Open Issues” was organised and held on 14-15 November 2016 in Florence. According to the Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data – so called FAIR principles – launched by the Data FAIRport initiative and recently adopted by the European Commission Guidelines on FAIR Data Management in Horizon 2020 the aim of this workshop was to bring new lymph to the political, legal, technical and technological debate around re-use and reproducibility of research data in Italy and to also address on going RDA activities related to FAIR. This workshop called to action the coordinators of research infrastructures, individual researchers, funding agencies, research institutions, data scientists, data librarians / curators and computing scientists. The workshop audience (more than 100 participants) included researchers, heads of research infrastructures and of computing infrastructures, pro-rectors for research, university research and open access departments, repository managers, librarians experts in metadata and data curation. The large majority of them were not yet part of RDA community. This event largely increased awareness among researchers and stakeholders of major Italian Universities on the different aspects that characterize a FAIR data management and on the solutions developed in the RDA and in other national and international contexts that can be exploited, either by each single institution or thought a coordinated effort, to implement it.

### 3.4.7 UK

48 people participated at the RDA/UK meeting on 2 November 2017. In particular the breakout topics (Trust and Certification, Data Citation, Metadata Standards and Publishing Data) attracted experts from various scientific communities and the list of speakers also indicates the variety of communities that were interested and contributing.

<https://researchdata.jiscinvolve.org/wp/2016/11/28/rda-uk-workshop/>

#### **Impact of National Meetings**

- *It is obvious that in particular for the non-native English speaking countries national meetings have a great impact.*
  - *They typically attract additional communities and national projects which are not represented well at plenaries.*
  - *They allow a much denser interaction and can focus on the special challenges of the respective countries.*
  - *In Europe at least the potential of adoptions fostered by national meetings is very high, since people more easily rely on "national ambassadors" who come up with useful showcases and since policies and principles can be transformed such that they can be applied in the countries.*
- *In all countries which have an active national community we can see that there is a higher percentage of people active in RDA groups and we also see that there is much more activity in adopting RDA results.*

## 3.5 Not-well represented communities

In this chapter we will present some thoughts about communities where we indicated a certain distance still after 4 years of activity. RDA finally is a bottom-up initiative, i.e. a distance of a certain community is an indicator that RDA is not seen as a useful platform to discuss and overcome inefficiencies in data practices. RDA Europe as a support action until now did not take specific measures to improve interaction with distant communities. It focussed on efforts towards those who are interested to make use of RDA. It is obvious that the list below cannot be comprehensive.

### 3.5.1 Computer Scientists

Within the BoD of RDA Europe partly open discussions are being triggered on how to better include the computer scientists. As can be seen from various statistics RDA and RDA Europe are engaging

many computer scientists and IT experts. However, it seems that there are existing networks of computer scientists which are not participating in RDA yet. The following aspects are mentioned in the discussions:

- RDA engagement does not offer sufficient incentives for reward.
- RDA Europe should support a PhD program that would make it interesting for computer scientists to become engaged,
- for many it was not obvious why there should be special measures for this community in particular and whether this would that really make a difference.

To build bridges RDA Europe 2 & 3 have been actively supporting a program for Technology Working Meetings (see chapter 3.3.7) which is meant to bring together in particular computer scientists with RDA experts and experts from interested communities and industry. The expectation is that due to the involvement of these experts the resulting highly quality reports will lead to

- an improvement of the RDA results
- initiatives for new working/interest groups
- an increased engagement of computer scientists

Yet it is too early to draw final conclusions although the first technical meeting on data citation led to a confirmation of the basic RDA results and to making them more precise.

### 3.5.2 Large European Entities

There is a perception that some of the large European entities such as CERN, ESO and EBI are not as active in RDA as it would be expected. It should be noted here that

- the EIROforum IT working group (including CERN and the other large organisation from the EIROforum) is an RDA Organisational member
- the CERN library is deeply engaged and in leading roles in the publishing groups of RDA
- members of the CERN infrastructure group are leading two RDA groups (archiving, federated identity management) and were elected to be in TAB in the first period
- the ELIXIR research infrastructure in which EBI plays an important role is highly active in various functions in RDA and RDA Europe
- ESO is informed through the EIROforum and RDA dissemination activities towards astronomy.

These large European Entities have an important role in Europe and beyond to establish standards within their domains, i.e. there is certainly an overlap in interests with RDA and yet it is not clear how they can use RDA as platforms to improve global and cross-disciplinary interactions.

Convincing these large European Entities to act as adopters is of great interest. Yet not all are active organisational members of RDA which obviously requires extended interactions.

### 3.5.3 Engineering

Despite intensive interactions at national level (see 3.4.3) institutions involved in research in engineering are not yet active in RDA. The reasons for this absence need to be analysed. It could be that RDAs outreach to industry is yet limited and that engineering departments are traditionally very close to industrial processes.

#### **3.5.4 High Energy Physics Community**

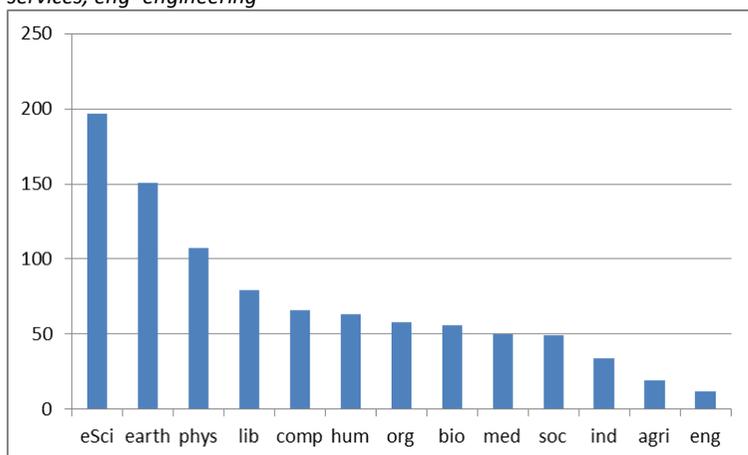
Despite highly active participation from physics communities it is obvious that the researchers from the specific High Energy Physics community are not represented in RDA. This probably has to do with the specific solutions found in this community and the high degree of international standardisation.

### 3.6 Summary Engagement Matrix

This summary matrix includes all quantifiable measures which can be associated with scientific communities. Since completely different kind of activities are mixed this matrix can only give indications. The big numbers from plenary participation overrule the participation in collaboration projects which have a higher impact factor.

Even type	lib	comp	eSci	earth	phys	bio	agri	med	hum	soc	ind	org	eng
Chairs	18	22	46	38	17	10	11	4	12	12	16	10	
RDA EU1 Interactions	4	2	30	13	15	14		7	15	5			
RDAEU2/3	10	6	36	9	12			1	7	1		17	
Science WS		3	10	9	11	8		6	6	1			
Collaboration projects running	1		2	1	1		1		1				
Collaboration projects 2 <sup>nd</sup> round	2	2	3	5	3	4	2	2	2	3			
GEDE	1			10	13	11	1	4	4	4			
RDA DE Meeting	26		25	22	12			17	11	19	2	8	10
Denver Week	12	19	29	35	18	8	3	7	2	3	14	22	
RDA EU PID Workshop	6	2	25	9	5	1	1	1	2		1	1	2
Total	80	66	197	151	107	56	19	50	63	49	34	58	12

lib=libraries/archives, comp=computer/information science, eSci=eScience/data/compute centres, earth=earth/geo/environmental sciences, phys=physical/chemical sciences, bio=biological sciences, agri=agriculture/forestry/fishery, med=medicine/health/neuro/cognitive sciences, hum=humanities, soc=social sciences/economics/law, ind=industry/publishers, org=organisations/governments/public services, eng=engineering



#### Activity Indication Diagram

This diagram gives an indication about active engagement in RDA and RDA EU activities. Most active are the members of the **eScience/data/compute centres** followed by the very active **earth/geo & environmental** and **physical & chemical** communities. Also the **library & archive** communities are fairly active followed by **computer & information scientists**, experts from **humanities, organisations & governmental** institutions, the **biological, social science & law** and **medical/health/neuroscience & cognitive science** communities. **Industry** engagement mostly comes from **publishers**, but in some cases also people interested in data analytics. Slightly less represented in total are the people from **agricultural/forestry & fishery** and the **engineering** communities, although the agricultural community is using RDA to drive standardisation. Engineering engagement comes from the German activities where an intensive interaction started.

## Appendix A: GEDE Group Classification

AnaEE	environmental	earth
EISCAT	atmosphere	earth
ENES	climate	earth
ENVRI Plus	environmental	earth
EURO-ARGO	marine-environmental	earth
IAGOS	earth observation	earth
ICOS	environmental	earth
SIOS	earth observation	earth
EPOS	earth observation	earth
CLARIN	humanities	hum
DARIAH	humanities	hum
E-RIHS	humanities	hum
PARTHENOS	humanities	hum
BBMRI	biomed	med
EATRIS	medicin	med
ECRIN	medicin	med
HBP	neuro science	med
Asterics	astronomy	phys
ELI	physics	phys
EMFL	physics	phys
ESRF	physics	phys
ESS	physics	phys
EST	astronomy	phys
EUCALL	physics	phys
FAIR	physics	phys
KM3NeT	astronomy	phys
MYRRHA	physics	phys
SCK•CEN	physics	phys
SKA	astronomy	phys
VAMDC	physics	phys
CESSDA	social science	soc
ESS - Social	social science	soc
NIDI	social science	soc
SERISS	social science	soc

The GEDE group has not yet decided to include the eInfrastructures in the initiative. Therefore, typical eScience/data/compute centres are not yet represented. It is assumed that in 2017 the GEDE group will open up in this respect.