The Research Data Alliance – The First Five Years

Francine Berman[1]
2019-2020 Radcliffe Fellow, Harvard University
Edward P. Hamilton Distinguished Professor of Computer Science
Rensselaer Polytechnic Institute

Abstract

The Research Data Alliance (RDA) is an international organization dedicated to promoting the development of global infrastructure for data sharing and data-driven exploration. Launched in 2013, the RDA has grown precipitously from 250 individuals at its first Plenary to over 9000 RDA community members from 135+ countries in 2019. The state and evolution of the RDA provides a case study in the development of an organization from the ground up and the choices made to create both community and impact. This piece provides a perspective on RDA's present, past, and future from one of the original founders of the organization.

Keywords: data, infrastructure, stewardship, preservation, global community, volunteer organization, organizational sustainability

1. WHAT IS THE RESEARCH DATA ALLIANCE?

The Research Data Alliance (RDA) is a community-driven organization dedicated to the development and use of infrastructure that promotes data sharing and data-driven exploration. The RDA supports the development of both technical infrastructure (code, protocols, tools, models, etc.) and social/community infrastructure (common vocabularies, curricula, pre-standards, etc.) that reduce the difficulties researchers encounter when seeking to access, harmonize, and use data to address research and societal problems.

The success of the RDA, as measured by growth of its community and the development of real infrastructure used by many, has been both precipitous and surprising. Precipitous, as RDA has grown from an initial phone call with eight founders in August, 2012 to a community of over 9000 from 135+ countries in 2019 (Figure 1). Surprising, because infrastructure often takes a backseat to research results, even as it enables those results. The success of the RDA demonstrates how critical the development of this infrastructure is for exploration, and the importance of empowering a community focused on its active use.

The story of the RDA is a case study in community building and organizational development. At every stage, the RDA community and leadership sought to adopt or establish best practices and to create an inclusive and "bottom up" environment where diversity – geographical, domain, career level, gender, race, sector, and perspective – was part of the organization's DNA. While the RDA has still has many

[1] This paper was funded in part by the National Science Foundation, Grant 1349002. We are grateful for the support.

organizational challenges, the openness of its culture and its dynamic community are two of its biggest assets.

Figure 1: *Growth of the RDA Community from its beginning. (From [RDA in a Nutshell, 2019])*

The RDA creates impact through both its work and its role as a "town square" for the data infrastructure community. The work of the RDA is conducted through self-organized, vetted Interest Groups and Working Groups that discuss and build infrastructure that solves real problems. Individuals are often attracted to the RDA as a vehicle to help them identify and build infrastructure needed for their own research or in their "day jobs".

To provide the lowest possible barrier to participation for a global community, joining the RDA is free and easy. Interested individuals agree to RDA's Principles – open access and transparent communication, consensus-driven decision making, balanced representation, harmonization across data infrastructure components, community-driven and volunteer efforts, and technology neutrality – and sign up via the web. New members often attend one of RDA's on-site Plenary meetings to participate in a one-time Birds-of-a-Feather session on a topic of interest or join/create an Interest Group or a Working Group focused on the problem they want to solve (e.g. What is a common vocabulary for agricultural data? How do we cite a specific version of a dataset that evolves over time? How can the type of data be registered in a way that is both machine-readable and human-readable?).

1A. RDA'S WORK AND OUTPUTS

Interest Groups and Working Groups drive RDA's infrastructure impact, and the type of RDA group created depends on what RDA members want to do. For example, anthropologists and folklore researchers developed the "Digital Practices in History and Ethnography" Interest Group to focus on identifying and leveraging infrastructure that supports historical and ethnographic research for the Anthropology community, Digital Humanities community, and others. The group has focused on characterizing a number of critical infrastructure components – metadata standards for researcher-created primary data (e.g. field notes and recorded interviews), citation practices, digital exhibition protocols, etc. After considerable discussion and "project shares", the Interest Group spun off a Working Group to define an infrastructure framework (standards, protocols, policies, cultural expectations) useful in making ethnographic and historical data archivable, discoverable and sharable.

The Working Group framework will be used by embedded group members from The Asthma Files (a collaborative ethnographic research project designed to advanced understanding and efforts to address environmental public health challenges) [The Asthma files, n.d.] and The Disaster-STS Research Network (a project that examines the history, organization, and challenges of disaster response) [Disaster-STS Research Network, n.d.], as well as others.

In general, RDA's Interest Groups are long-lived and broadly targeted. Like the Digital Practices in History and Ethnography Interest Group, they come together to discuss infrastructure challenges, develop a sense of best practice and missing tools, standards, policy, protocols, etc., and/or formulate infrastructure that would be helpful for specific needs in the community. Interest Groups often spin off related Working Groups when there is something to build or develop, and Working Groups often reconstitute themselves as Interest Groups when their work is done and they would like to maintain their outputs or conceptualize new ones.

In contrast, RDA Working Groups are focused groups who build infrastructure to solve a specific problem over a 12-18 month time period. Although it is important that RDA Working Groups solve somebody's problem, they need not solve everybody's problem. Few solutions work for every instance and most infrastructure has a lifetime. Early on, RDA conceptualized its role as a pragmatic vehicle to reduce infrastructure roadblocks to innovation. If RDA's infrastructure outputs address one instance of a problem but not another, that's OK. This approach has allowed RDA to remain agile, maximizing its potential for impact.

RDA Working Groups are established by developing a "Case Statement" that describes what infrastructure they want to build, who will be building it, and who will benefit from/adopt it. All RDA Working Group infrastructure must solve a specific problem for specific people/projects. This helps avoid "build it and they will come" scenarios. On receiving a Case Statement from a potential Working Group, RDA's Technical Advisory Board (TAB) makes an assessment of the proposed work, evaluating it for technical strength, relevance to RDA, and the inclusion of adopters of the infrastructure as part of the Working Group. After TAB assessment, community comment, and final approval by RDA's leadership Council, the Working Group is launched with the expectation that the infrastructure will be built and implemented within 12-18 months.

Once the Working Group has completed its work, its Recommendations or Supporting Outputs[2] are checked and endorsed by RDA and deposited in Zenodo and/or the RDA website as appropriate [Working Group Outputs, n.d.]. RDA works proactively to empower further adoption of its outputs for individuals and projects that can use them, with the goal of broadening the impact of useful data infrastructure. Both Working Groups and Interest Groups operate with support from RDA's TAB for technical coordination and from RDA's Organizational Assembly for help with implementation and adoption of RDA's outputs. The current list of RDA Working and Interest Groups is on its website [Groups, n.d.].

[2] RDA Recommendations are recognized and endorsed outputs from RDA Working Groups. Supporting Outputs are useful solutions from RDA Working Groups and Interest Groups but may be less amenable to adoption. In this paper, we use the general category of "outputs" to refer to both.

1B. IMPACT THROUGH OUTPUT ADOPTION

Beyond the embedded infrastructure adopters in each Working Group, RDA works hard to disseminate its infrastructure as widely as possible. For this reason, some geographical regions of the RDA have created "Adoption Programs" that help potential adopters of RDA infrastructure expend the time and effort it often takes to incorporate new infrastructure within their local environments. For example, one of the first "Adoption Programs" was funded by the John D. and Catherine T. MacArthur Foundation in the U.S. region of RDA. The program offered a small amount of funding to support embedded staff and graduate students in projects who wanted to integrate RDA infrastructure but needed the time and resources to do so. Original infrastructure adoptions from the MacArthur-funded program included

- adoption of the *Data Type Registry Working Group output* by the Vermont Monitoring Cooperative (forestry) and the National Institute for Advanced Industrial Science and Technology in Tokyo and Indiana University in the U.S. (for rice genomics and analysis)
- adoption of the *Dynamic Data Citation Working Group output* by the Vermont Monitoring Cooperative (forestry), the Biological and Chemical Oceanography Data Management Office at Woods Hole Oceanographic Institute (oceanography) and the Center for Biomedical Informatics at Washington University in St. Louis (electronic health records)
- adoption of the *Persistent Identifier Information Types output* by the he National Institute for Advanced Industrial Science and Technology in Tokyo and Indiana University in the U.S. (for rice genomics and analysis)

Adopters integrated these RDA outputs within their local environments and reported on their results at a subsequent RDA Plenary meeting. Additional adoption programs have been funded by the National Science Foundation (NSF) in the U.S., the European Commission in the European Union (E.U.), and others.

The focus on adoption beyond the original Working Group constituents set a tone and a focus for RDA as it evolved. Today, broad adoption is a primary focus of the organization to increase its impact and accelerate its mission. The RDA maintains a current list of outputs and adopters on its website [Adoption of Recommendations, n.d.].

1C. RDA AS A COMMUNITY "TOWN SQUARE"

One of the most important impacts of RDA has been the value its Plenaries bring to RDA members, stakeholders, and other groups. The semi-annual RDA Plenaries (typically 2.5-3 days) are dynamic, high-energy, productive, and inclusive. The Plenaries are primarily working meetings and although each Plenary has high-profile keynote speakers and excellent panels, the majority of the time is spent doing, rather than listening. There are often a dozen "officially scheduled" RDA group meetings going on at the same time and many "unofficial" group meetings, allowing RDA members to make face-to-face progress on their work more rapidly than possible when they interact remotely (often in vastly different time zones).

An additional value of RDA Plenaries for members is whom they attract. Many different stakeholders (Program Officers from a variety of national R&D agencies and non-profits, policy makers, representatives from other organizations, publishers, journalists, and others) attend the RDA Plenary to

collaborate, get the "lay of the land" for a broad scope of problems and disciplines, and interact both with the RDA community and each other. Because RDA provides a neutral and collaborative environment, many groups co-locate meetings with RDA, stretching out the three-day Plenary to a crowded week of synergistic activities.

The evolution of RDA Plenaries as a welcoming community space did not happen by accident. From the beginning, RDA sought to collaborate and coordinate with other organizations whose mission and efforts were complementary and where there would be synergy and value-added in partnership. Survival for community-driven organizations can be tough, and competition is often created out of the need for credit to gain sparse resources, sometimes breeding "world domination" organizational cultures. While dealing with its own (serious) sustainability challenges, RDA has always sought to partner and collaborate with other organizations, seeking synergy and greater impact. This openness and attitude has helped RDA evolve its role as a data infrastructure community town square and brings great value to RDA Plenaries. Groups who have held adjacent or joint meetings with RDA include the World Data System, CODATA, Force 11, DataCite, EarthCube, GO FAIR, the Botswana International University of Science and Technology, E.U. EOSC-related Horizon 2020 Projects, and many others.

2. WHY RDA?

RDA evolved as a specific solution -- *that of creating and elevating the development of infrastructure to help individuals effectively deal with research data* – to a specific problem – *the sparsity and inadequacy of infrastructure to support data sharing and data-driven exploration*.

The problem of insufficient and inadequate data infrastructure is particularly pressing in the academic research community. Infrastructure used by researchers often has a comparatively small "market" of users compared with infrastructure products and services in the private sector or targeted open-source efforts for commonly used programs and systems. Research data infrastructure is often developed and utilized by individual researchers, specific projects, or domain communities as a means to an end (new results, scholarly literature). Responsibility for its maintenance, upgrade, and support may be unclear or inconsistent. Researchers are incentivized to move on when their work evolves to focus on new results, often deprioritizing the maintenance and support of previously useful infrastructure.

Part of the difficulty in creating and maintaining research data infrastructure in academia is recognition and funding. Infrastructure often plays second fiddle to new exploration as a funding priority. It is challenging to obtain funding to maintain or improve infrastructure for the purpose of keeping it going or making it more useful to a larger user base. Moreover, while research results advance the reputation of researchers and their institutions (often leading to greater opportunities and funding), developing or maintaining effective working infrastructure rarely has this outcome. In the academic environment, this generally translates into less recognition and resources for the professionals who enable data-driven research than for their faculty research colleagues who conduct the research.

These problems are exacerbated because what makes research notable and what makes research infrastructure notable is exactly the opposite: Research is noteworthy for its successes; infrastructure development and deployment is noteworthy for its failures. You may not think about the fact that hot water is available in your building until the plumbing breaks or that your commute home is smooth unless there are potholes in the road. Research infrastructure for the access, use, stewardship and

preservation of data is similar. Until data is corrupted, lost, unusable, or becomes inaccessible, support for the infrastructure that delivered it, or what is needed to maintain and support it, may be deprioritized. In some cases, even the loss of data may do little to change priorities.

Part of the "why" of RDA was to address these infrastructure recognition and funding problems and their negative implications for the research data community. By creating a global organization, RDA could provide a venue for the data infrastructure community to come together to build, maintain and deploy infrastructure, as well as to gain recognition for their efforts. The organization has been successful in this, attracting policy makers and funders to its Plenaries and hosting substantive discussions through its "Funders Forum" (a side meeting that is part of every RDA Plenary). The problems of infrastructure recognition and funding also account in part for the attraction and large percentage of RDA membership from research environments in which funding for research infrastructure is tenuous or low priority (e.g. academia), as well as the smaller percentage of RDA members from the private sector (where research infrastructure is considered a strategy for future competitive advantage and is often not in danger of losing support).

As the need for enabling data infrastructure is particularly pressing in academia, it is fitting that the RDA began with the recognition from astute public sector R&D leaders that the data community needed help, acknowledgement, and support to more effectively build and deploy the data infrastructure on which modern research depends.

2A. RDA's ORIGINS STORY

In 2011 and 2012, many discussions focused on the need for more and better infrastructure to support data-driven research efforts around the world. Both the funding and research communities recognized that the growing gap between research exploration and enabling data infrastructure would slow the advancement of innovation if not better addressed. The concept of establishing an initiative to promote the exchange of data across international boundaries was presented at the "Global Research Data Infrastructures: The Big Data Challenges Conference" in Brussels. Harmonization of data access and interoperability solutions was also the focus of ICRI (International Conference on Research Infrastructures) held in Copenhagen in 2012, sponsored by an emerging group called the DAITF (Data Access and Interoperability Task Force).

Discussions on these themes continued at the Open Infrastructures for Open Science meeting in Rome, where funders Chris Greer from the National Institute of Standards and Technology (NIST) in the U.S., Alan Blatecky from the NSF in the U.S., Carlos Morais Pires and Konstantinos Glinos from the European Commission in the E.U., and Clare McLaughlin from the Australian Government strategized on how they could appropriately empower the community to come together in a cohesive way to focus on building and increasing effective research infrastructure world-wide.

In the U.S., Blatecky and Greer developed a concept paper with these ideas for a community group they called the Data Web Forum (DWF). The DWF concept paper [Concept Paper – Data Web Forum, 2012] envisioned a bottom-up non-governmental organization focused on data infrastructure that would be structured along the lines of the Internet Engineering Task Force [IETF, n.d.]. They conceptualized a DWF that would be community-based, action-oriented, and driven by data practitioners with an analogue of the IETF's "rough consensus and running code" approach. Blatecky had worked with a number of middleware and infrastructure projects at NSF and elsewhere and Greer had long been

involved in the data world and was currently overseeing the Smart Grid Interoperability Panel for NIST. Blatecky and Greer conceptualized the DWF as an organization that would be pragmatic and results-oriented. They approached colleague Fran Berman at RPI to provide feedback on the document and during the discussions, all three became compelled by the potential of such an organization.

Concurrently in the E.U., the enabling role of research infrastructure was recognized in the influential reports such as "Riding the Wave" [Riding the Wave, 2010] from the High Level Expert Group on Scientific Data Infrastructures," Beyond Sharing and Re-Using: Toward Global Data Networking" [Beyond Sharing, 2011], and by the creation of DAITF. It also led to the creation of the European project iCordi, later known as RDA Europe, in September 2012. iCordi's aim was to enable convergence of emerging global infrastructures via a "coordination platform" that could reduce barriers to the interoperability of data architectures by fostering discussion between data practitioners in a bottom-up process.

Similarly in Australia, the importance of data was described in the influential "Towards an Australian Data Commons" report and served as the focus of a wide-ranging National Collaborative Research Infrastructure Strategy [NCRIS, n.d.]. The development and use of infrastructure was a primary focus of the Australian National Data Service [ANDS, n.d.] (now subsumed by the Australian Research Data Commons) who focused on the support of Australian research data through a broad set of efforts that included policy, skills, national data services, and coordination of institutional services. ANDS saw data as a research asset, rather than a researcher problem, and recognized that much of the data and data infrastructure activity needed to happen internationally and not just nationally to optimize effectiveness.

These world-wide initiatives were advanced by discussions within the G8+5 group on Research Infrastructures, who established a working group on Data in July 2011 to look at incentives and standardization for interoperability, access policies, certification, legal frameworks, and international governance. The G8+5 group produced a draft report in October 2011 [G8+O5 Global Research Infrastructure, 2011] and a white paper in May 2012 [Blatecky, Bicarregui, Morais Pires, 2013]. Their efforts led to the G8 ministerial communique in June 2013 [G8 Science Ministers Statement, 2013] which promoted openness in scientific research data, and embraced the perspective that publicly funded scientific research data should be discoverable, accessible, assessable, intelligible, useable, and wherever possible, interoperable.

Throughout the world, it was clear that the opportunity to collaborate, exchange, and share best practices would "raise the level of the ocean" for all involved. In August 2012, at the encouragement of Blatecky, Greer, Pires, Glinos and McLaughlin, a call was arranged between what were to become the RDA's founders: Berman and Beth Plale from the United States, Leif Laaksonen from Finland, John Wood and Juan Bicarregui from the U.K., Ross Wilkinson and Andrew Treloar from Australia, and Peter Wittenburg from Germany.

The group discussed the possibility of a DWF-type organization and agreed to organize a planning meeting to explore its creation. The resulting Global Data Meeting was held in October 2012 in Washington, D.C. with about 100 key community members from the DAITF, major infrastructure projects in the U.S., E.U., Canada and Australia, and other interested parties. In arguably the first instance of what has consistently been the community-driven "bottom-up" culture of RDA, the Global Data Meeting was organized as a working meeting where the organizational structures proposed in the

DWF Concept Paper were discussed and, importantly, the first RDA Working Groups self-organized to begin to discuss building infrastructure that would enable better data sharing and data-driven research for specific challenges.

With the Global Data Meeting, the concept of RDA began to pick up steam. Plans were made for the first RDA Plenary in March of 2013 in Gothenburg, Sweden. Between October 2012 and March 2013, the Global Data Meeting group continued to work towards the development of a real organization. A charter, principles, vision, mission, and organizational structure were drafted. Funders from the U.S. (NSF), the E.U. (the European Commission), and the Australian Government provided support for RDA development in their home regions. The Data Web Forum was renamed the Research Data Alliance. The nascent Working Groups who came together at the Global Data Meeting continued their formulation of the work they wanted to do under the auspices of the RDA. The core of RDA's first leadership Council was appointed by the original funders and consisted of Fran Berman (co-Chair), John Wood (co-Chair) and Ross Wilkinson. The founders and others began to introduce RDA to potential new members and stakeholders, and the organization was underway.

Plenary 1 in Sweden [First Plenary Meeting, 2013] became the official launch and first meeting of the RDA and was tremendously exciting. Attended by 250 new RDA members from 45 countries, Plenary 1 provided a working meeting for 17 pre-Working Groups and a venue to continue to conceptualize the organization. Although the community was still building the RDA train as it was going down the tracks, RDA founders and new members were committed to building an open, diverse, active community that got things done.

At Plenary 1, the interests of the community were already clear. The RDA organization should be developed based on best practices observed from other groups and agile enough to iteratively improve when something wasn't working. RDA should be inclusive. The community wanted all leadership groups to be geographically balanced, gender balanced, and to represent a balance of data professions. The Working Group vetting procedures should ensure that infrastructure that was developed was used and that time was not spent re-inventing the wheel. A new kind of group – the Interest Group – was discussed as a longer-term vehicle for discussion of data infrastructure issues. It was exhilarating for the community to work together to invent something from the ground up to fill a need that, although often under-valued, was so critical to progress.

After Plenary 1, the new RDA community continued to work. The first Technical Advisory Board – the leadership RDA group that vets the technical quality of new groups and their outputs – was elected at Plenary 1 with Andrew Treloar and Francoise Genova serving as its first co-Chairs. RDA's founders and initial Council developed RDA's governance documents, organizational components, and began to tackle what remains one of RDA's thorniest problems today – organizational sustainability. Between Plenary 1 (March 2013) and Plenary 2 (September 2013), RDA began to add to its leadership Council and to develop an additional group – the Organizational Assembly – as a way for organizations to interact with and provide advice and support to the RDA[3].

It was also clear that RDA needed an administrative arm. The RDA Secretariat was developed from RDA staff paid through national funding vehicles. Secretariat members remained in their geographical

[3] The first co-Chairs of the Organizational Assembly were Juan Biccaregui and Leif Laaksonen. As the Organizational Assembly grew, it created an Organizational Advisory Board (OAB) as its leadership subgroup.

regions but worked together virtually to ensure the operation of the global organization. The first RDA Secretariat leveraged staff from the U.S., E.U. and Australia to manage the RDA website, group discussions, leadership activities, communications, and Plenary planning. Although these staff were employed in different regions, their purpose was to create an administrative structure for "RDA international" (or "RDA/Earth" as it was sometimes referred to internally …). After two Plenaries, it became clear that RDA needed a full-time lead and that the Secretariat would benefit from a more formal management structure. RDA's Council developed a job description, and more importantly an almost adequate funding stream, and advertised for a Secretary General for the new organization.

RDA's first Secretary General was Mark Parsons, previously the Managing Director of the U.S. region of RDA. Parsons focused on building and coalescing the diverse and fast-growing RDA community. With the addition of a Secretary General, the RDA community had a public point of contact, important both internally and externally. Initially and more so as it evolved, the Secretary General's job was highly multi-dimensional and involved a lot of juggling of important priorities – community building, interactions with stakeholders, conflict resolution, management of the distributed Secretariat, partnership with RDA's Council, TAB, Organizational Assembly (and later Organizational Advisory Board (OAB)), and partner organizations, and focus on sustainability. Parsons also emphasized RDA's pragmatic focus on solutions for real and specific problems, rather than creating a central technical architecture. RDA was growing quickly, in terms of community, partnerships, and outputs, yet its administrative and financial support stayed constant at best. These challenges were also challenges for the Secretary General.

Along with the focus on staffing for RDA international, it was clear that RDA needed a bank account for revenue that was not associated with the growing regional communities in RDA. In 2013, facilitated by John Wood and Juan Bicarregui, RDA created the RDA Foundation (RDAF). RDAF was established as a U.K. Charity, whose Articles of Association designated its Board of Directors to be whomever was an active RDA Council member at the time. RDAF enabled RDA to have a bank account for any incoming revenue for the international RDA that could be used to pay the expenses of RDA. The choice of a U.K. Charity for RDAF was based on the ability to accept funding from governments and organizations within RDA's 3 major revenue-providing regions (the E.U., the U.S. and Australia) and because support for the development of RDAF could be generously provided by RDA's European colleagues[4]. The RDA Foundation is the employer of RDA's current Secretary General, Hilary Hanahoe.

After the first year of RDA Plenaries, community building and organizational development, RDA continued to grow and its Plenaries began to evolve as a data community town square. A matrix of RDA Plenaries, their attendance, international distribution, and the number of Working Groups and Interest Groups in the organization is shown in Figure 2.

Plenary 1 / 2013 (spring) *Gothenburg, Sweden*	**Plenary 2 / 2013 (fall)** *Washington, DC, U.S.*	**Plenary 3 / 2014 (spring)** *Dublin, Ireland*	**Plenary 4 / 2014 (fall)** *Amsterdam, The Netherlands*	**Plenary 5 / 2015 (spring)** *San Diego, CA, US*

[4] Note that this was pre-Brexit.

250 attendees from 45 countries 17 (pre) Working Groups	380 attendees from 22 countries 22 Working Groups and Interest Groups	497 attendees from 32 countries 34 Working Groups and Interest Groups	550 attendees from 40 countries 40 Working Groups and Interest Groups	383 attendees from 30 countries 55 Working Groups and Interest Groups
Plenary 6 / 2015 (fall) *Paris, France* 603 attendees from 34 countries 58 Working Groups and Interest Groups	**Plenary 7 / 2016 (spring)** *Tokyo, Japan* 373 attendees from 33 countries 73 Working Groups and Interest Groups	**Plenary 8 / 2016 (fall)** *Denver, CO, US* 814 attendees from 33 countries for the combined *International Data Week* co-sponsored by RDA, CODATA and WDS	**Plenary 9 / 2017 (spring)** *Barcelona, Spain* 600 attendees from 44 countries 85 Working Groups and Interest Groups	**Plenary 10 / 2017 (fall)** *Montreal, Canada* 437 attendees from 29 countries 88 Working Groups and Interest Groups
Plenary 11 / 2018 (spring) *Berlin, Germany* 661 attendees from 41 countries 91 Working Groups and Interest Groups	**Plenary 12 / 2018 (fall)** *Gaborone, Botswana* 850 attendees from 66 countries for the combined *International Data Week* co-sponsored by RDA, CODATA and WDS	**Plenary 13 / 2019 (spring)** *Philadelphia, PA, US* 450 attendees from 34 countries 102 Working Groups and Interest Groups	**Plenary 14 / 2019 (fall)** *Helsinki, Finland* [Planned]	**Plenary 15 / 2020 (spring)** [TBD]

Figure 2. *RDA Plenaries, dates, attendance, number of Working Groups and Interest Groups.*

2B. STRATEGIC PLANNING AND PROFESSIONALIZATION

As RDA evolved, its community grew and the organization began to mature. In 2015 at Plenary 5, RDA's leadership held a strategic planning meeting to assess the state of RDA and identify areas for organizational improvement over the next three years. Three themes emerged as areas for RDA to address growing forward. These were *Communications* (to better inform and facilitate interaction within RDA's membership), *Engagement* (with the broader community), and *Coordination* (better interconnection for RDA's constituent organizational groups) [Berman, Collins, Stewart, Wilkinson, 2015]. As the RDA community had grown from 250 in 2013 to 3000+ in 2015, the strategies that had worked to keep the community cohesive were now straining due to scale and lack of growth of RDA's Secretariat and leadership groups. Over a dozen tasks were created to address these issues, many of which were successful and some of which "died on the vine" due to lack of the human resources needed to implement the tasks.

The scale of the RDA community in comparison to its relatively small Secretariat and loose community of volunteers was becoming a serious problem. Evolution of agency priorities and program officers also changed the funding landscape for the organization over time. The lack of growth in both staffing and funding remain serious problems for RDA to this day.

Most recently, RDA updated its strategic planning priorities. During its 2017 off-plenary meeting, RDA's Council developed task forces to assess RDA's current status and needs to further professionalize all aspects of the organization. Four pressing areas were identified:

- *Growth management* -- the challenge of serving and responding to the needs of RDA's fast-growing community
- *Value proposition* -- the challenge of creating viable value propositions for RDA's existing and potential stakeholders and members
- *Regional engagement* -- the challenge of formalizing the important role of RDA's regions
- *Financial sustainability* – the challenge of providing consistent and adequate funding to operate RDA, accomplish its mission, and serve its community and stakeholders

These challenges are critical to RDA's potential for impact and ultimately to its useful existence. At Plenary 13 in spring 2019, it was clear that RDA had greatly matured as an organization and that the efforts to keep the community open, pragmatic, focused on impact, and collaborative had helped it deliver on its promises to build effective infrastructure and elevate the broad and important community of data professionals. The community can be justifiably proud of what RDA has achieved and its outstanding trajectory of success.

3. WHO IS RDA

RDA is comprised of a highly diverse and broad community. It has grown and evolved based on the committed efforts of its leadership groups – TAB, OAB, and Council, as well as its Secretariat. In

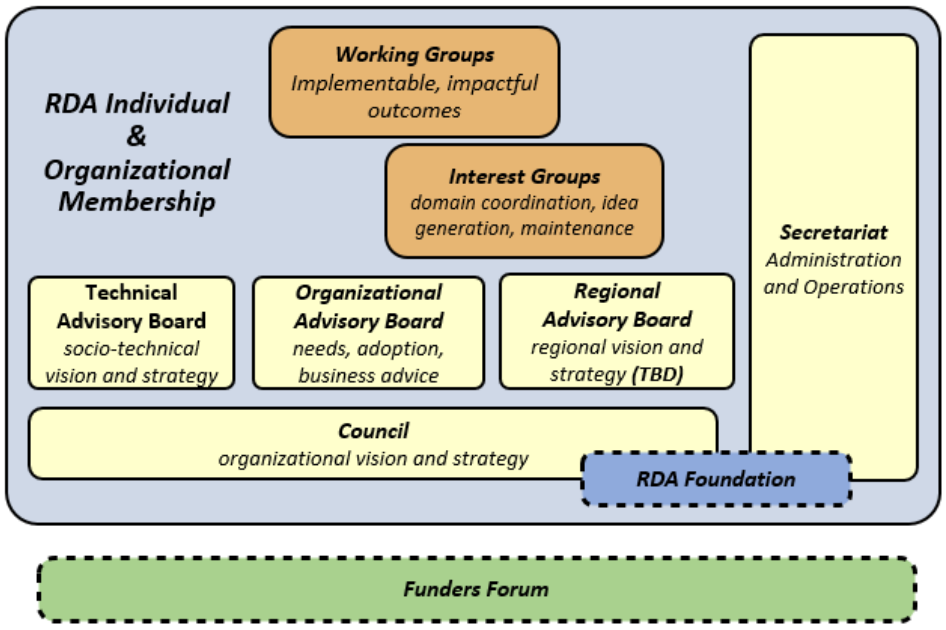

Figure 3. *RDA Organigram with current and proposed groups (courtesy of Hilary Hanahoe).*

addition, international R&D agencies and non-profits with an interest in RDA who meet in a loosely affiliated RDA Funder's Forum have been important fellow-travelers for both the organization and the community. In this section, we describe all of these groups, their roles, and influences.

RDA's "organigram" is shown in Figure 3. It includes a Regional Advisory Board, currently proposed to be included in the organizational structure, whose co-Chairs will likely be new non-voting members of Council (similar to the co-Chairs of TAB and OAB).

3A. RDA Members

RDA membership is global. Due to the strong regional funding base in Europe and the U.S., most RDA members have been from these regions, but RDA has continued to enjoy steady growth in all of its regions. At every Plenary, the RDA's global "map" includes new jurisdictions. The most recent RDA membership map (fall 2019) is given in Figure 4. Note that for many of the reasons described in Section 2, most RDA members come from academia.

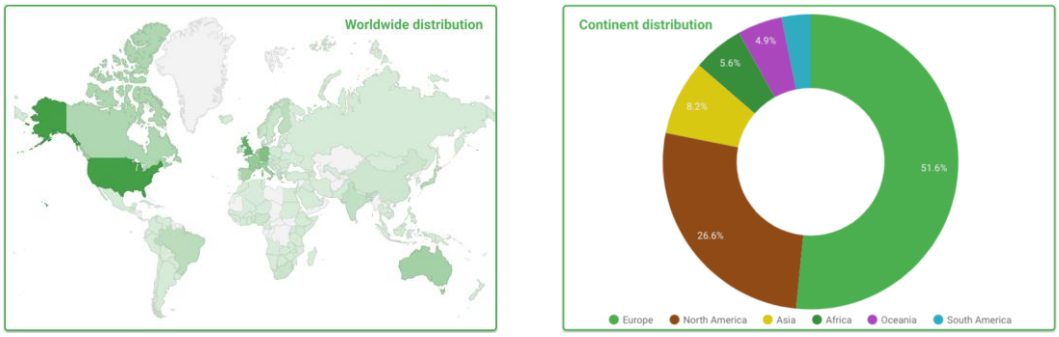

Figure 4. *RDA membership map as of fall 2019. (Graphs from [RDA in a Nutshell, 2019])*

To maximize impact, RDA has solicited additional resources to support the work of two particular cohorts of RDA members – Working and Interest Group Co-Chairs, and Student and Early Career Professionals. It was clear early on, and especially because of the small number of Working Groups, that coordination between developed outputs, when possible, would increase the effectiveness of RDA infrastructure. Funding from the U.S. (NIST) and from the E.U. supported off-Plenary workshops twice yearly for Working Group co-Chairs to discuss coordinating their work. The RDA "Chairs" meetings have been held for 5 years, alternating between the E.U. and the U.S. As the number of groups has grown, the coordination focus of the Chairs meetings has been particularly important to foster synergy and to maximize interaction between related groups.

RDA members have also successfully solicited funding to support student and early career professionals to engage with the organization. A strong student and early career population in the RDA is a win-win. Students and early career professionals benefit from RDA's diverse and far-reaching network, gain practical experience working with infrastructure, and are exposed to a broad perspectives to help with their own work. The community benefits from the opportunity to incorporate next generation leadership, new perspectives within RDA, and to integrate the community "vertically" (in professional age) as well as horizontally (across sectors and disciplines). Funding from the Alfred P. Sloan Foundation and the NSF in the U.S. has helped support student engagement within the RDA. Similar programs have been funded in the E.U. by the European Commission. The RDA Early Career and Engagement Interest

Group, endorsed in 2017, offers events, networking, mentoring and professional development opportunities for RDA students and early career professionals.

3B. RDA Secretariat

The RDA Secretariat is unarguably the hardest working group in the RDA and RDA's only paid members. Currently led by RDA Secretary General Hilary Hanahoe, the Secretariat consists of 9 individuals who spend between 2.5% and 70% of their time (and generally more) working on the operational and administrative side of RDA. (Note that the sum of all the fractional efforts of the Secretariat totals 2.5 full-time equivalent positions, excluding the [full-time] Secretary General.) RDA simply could not exist without its Secretariat and it is a testament to the commitment and dedication of this group that RDA manages to put on two international Plenaries a year, oversee over 100+ Working and Interest Groups, maintain a website that facilitates its work, provide communications that are distributed worldwide, maintain appropriate accounting and legal structures, and many other functions.

3C. RDA Funder's Forum

Research funders have been important allies for RDA since the beginning and use each Plenary as a venue to both interact with RDA and each other. Most recently, this group initiated an open Interest Group -- the Research Funders and Stakeholders on Open Research and Data Management Policies and Practices Interest Group [Research Funders and Stakeholders Interest Group, n.d.], whose charter is to create "*… a forum for research funders to initiate or further data-related discussions could result in lasting benefits for the research community (increased coordination, standardization and support) and the funders themselves in terms of avoiding duplication of efforts and building on existing expertise and resources.*" Both the Funder's Forum and the related Interest Group are great examples of how groups use RDA's organizational structure to increase both engagement and effectiveness.

3D. RDA's Leadership Groups: Technical Advisory Board, Organizational Advisory Board, Council

Whereas the community does the "work" of RDA, RDA's leadership groups and its Secretariat conduct the "business" of RDA. RDA members are typically most familiar with TAB, which vets new Working and Interest Groups and helps guide them on a successful trajectory. TAB evaluates every case statement (proposal) of every Working Group and every charter of every Interest Group, facilitates useful interaction within and between Groups, troubleshoots Group problems, etc. Because of their heavy (volunteer) workload, the size of TAB has grown from 12 to 15 over time. Members are elected by community votes that serve as inputs to a "balancing algorithm" [RDA TAB Responsibilities and Processes, n.d.] that ensures that TAB members cover key cohorts of the data community and are balanced by geographical region. The co-Chairs of TAB serve as non-voting members on RDA's Council.

The OAB represents RDA's Organizational Assembly, a unique group within RDA that represents its organizational members. OAB members provide an organizational perspective that is both helpful with respect to RDA's own organizational structure, and helpful in terms of providing an organizational assessment of Working Group outputs and potential for adoption. The OAB review of outputs is important as impact is maximized when entire organizations are able to incorporate RDA's infrastructure into their local environments. The OAB helps RDA maintain a diverse perspective and

incorporate the needs of environments beyond academia. The co-Chairs of OAB serve as non-voting members on RDA's Council.

RDA's Council serves as its Board of Directors. Charged with the strategic direction and the health and well-being of RDA, Council provides final approval on Working Group Case Statements and Interest Group charters (generally on recommendation from TAB), and focuses on organizational strategy, sustainability and impact. RDA's Council consists of nine elected "statespeople" whose role is to focus on the international organization, rather than to represent their own jurisdiction or data profession, and includes the Secretary General (non-voting) and non-voting co-Chairs from TAB and OAB. Council oversees partnerships and collaborative activities with other organizations and works closely with stakeholders, especially the Funder's Forum. The co-Chairs of Council and the Secretary General form an unofficial Executive Committee for RDA and maintain frequent contact to ensure that the organization is on track. Voting members of Council serve as the Board of Directors of the RDA Foundation, the legal entity that hosts RDA.

As of this writing, there are plans to include additional non-voting members on Council, the Co-Chairs of the emerging Regional Advisory Board, an important and currently unofficial group in RDA. The important role of RDA's regions is described below.

3E. RDA Regions

For current purposes, RDA defines "region" to be a "national level" geographic entity or a consortium of "national level" entities. [RDA Regions, n.d.]. This means that RDA/Europe can be a region but RDA/Chemistry cannot. (RDA/"region" is the usual designation of an RDA region.)

From the outset, RDA regions have had a complementary and critical relationship with RDA international. RDA regions bring their local challenges and priorities to RDA; RDA provides useful international perspectives and additional expertise to address these challenges. Regional funding supports the international Secretariat (often "in-region"); regional leaders are often among the leadership of the international RDA. Regional success builds RDA; RDA's international character provides value added to regional efforts and the professional trajectories of regional community members.

Since RDA's inception, funding for the Secretariat and Secretary General has come largely through RDA regions. This arose in part because many funding programs focus on national constituencies and it has been difficult to solicit funding directly for RDAF. RDA's first 3 Council members – Berman, Wood, and Wilkinson – were also leaders in their respective regions - the U.S., E.U., and Australia -- creating leadership alignment. As RDA grew and additional regions emerged, regional leadership and international leadership has become less coupled. It is challenging if regions are accountable for providing RDA funding but not responsible for its management. This is a problem the RDA Council wanted to address during its 2017 strategic planning process.

As a result, a Council subgroup was tasked with creating a framework for engaging regional leadership and the community to formalize the role of RDA regions as an RDA leadership group. Regional representatives from a diverse set of regions and Council met at Plenary 11 to launch an effort to create a Regional Assembly (and its leadership subgroup, the Regional Advisory Board). At the meeting, interested parties formed a "Regional Engagement Task Force" to tackle the issues and engage the

community with a proposal at Plenary 12. As in most things, the "devil is in the details" to get regional engagement right. Key questions include:

- What should RDA international expect from the regions and what do the regions get by being formal members of the Regional Assembly?
- How do we avoid competition for funding between RDA regions and RDA? How should regional funding for RDA international be controlled and managed?
- How can we harmonize an individual's membership in RDA and one or more regions (i.e. an individual might be a member of RDA, RDA/Europe, and RDA/Germany)?
- What are the mission, roles and responsibilities of the Regional Advisory Board?

The Regional Engagement Task Force is addressing these issues and developing the requisite governance documents. When these are completed, RDA will welcome the Regional Assembly and its Regional Advisory Board as one of its formal leadership groups (with its co-Chairs being a formal non-voting part of the RDA Council).

3F. REGIONAL CASE STUDY: RDA/US

Each RDA region is different. The political landscape, national priorities, path to leadership, and relationships with and expectations of national stakeholders all vary dramatically around the world. In the U.S., RDA has been relatively independent from its funders and stakeholders, allowing it to evolve with more freedom from specific expectations than other regions, but also with less commitment for funding.

As described earlier, RDA/US evolved from initial discussions between Fran Berman, Alan Blatecky and Chris Greer on Greer and Blatecky's Data Web Forum concept paper. Berman signed on to lead RDA/US with partner Beth Plale from Indiana University. During 2012, Berman and Plale worked to develop funding and create a core group for RDA/US that included Larry Lannom, Rebecca Koskela, Jane Greenberg and others who had attended the Global Data Meeting. By Plenary 1, RDA/US had established an "office" at Rensselaer Polytechnic Institute with an administrative coordinator (Jamie Petta) and had hired a Managing Director (Mark Parsons). RDA/US grew steadily after that. In 2014, the RDA/US office recruited Yolanda Meleco as Communications Manager and to be part of the international Secretariat. Like other regions, the focus of the RDA/US office was to both support the international RDA as part of the Secretariat and to build a regional community within the U.S.

Over the years, the RDA/US office expanded to include a Community Development Director and additional Secretariat member (Lynn Yarmey, coming in 2016), Managing Director (Kathy Fontaine from 2014-2015 succeeding Mark Parsons), and an RDA student "resident" (Candice Lanius in 2014-2015) with occasional help from other part-time student interns. Foundational support was provided for the RDA/US office and RDA/US programmatic funding by the NSF. Additional programmatic funding for the RDA/US community was provided by the Alfred P. Sloan Foundation, NIST, and the MacArthur Foundation. RDA/US is indebted to them all.

The RDA/US office focused on outreach to U.S. stakeholders and the broad U.S. data community, support for the international RDA Secretariat, hosting of RDA Plenaries (Plenaries 2, 5, 8, 13), and participant support for U.S.-based RDA leadership (Working and Interest Group co-Chairs, TAB

members, OAB co-Chairs, Council members, Secretariat members) to attend international RDA conferences. This participant support was critical to help U.S.-based RDA leadership fully participate in RDA. RDA/US also held annual leadership meetings with this group. Note that leadership in RDA/US was self-initiated: U.S. members who had the effort and commitment to lead something in RDA were automatically included in the RDA/US leadership group for as long as they held their roles. This kept the U.S. leadership group dynamic active, and current.

In 2017, RDA/US began to plan for years 5-10 and "RDA/US 2.0". The RDA/US office developed the position of an "Executive Director" who would create a formal vehicle to move RDA/US forward and develop sustainable funding for RDA/US' next generation. Leslie McIntosh was hired as RDA/US Executive Director and worked with a pro bono law firm to develop a 501c3 that could support a broad spectrum of efforts and revenue sources. RDA-US, Inc. is in the process of being established with a focus on community, adoption, and sustainability. Borrelli recruited Anthony Jeuhne to join the RDA/US office with a focus on adoption and support of RDA/US 2.0.

RDA/US 2.0 is in its early stages, but early on, McIntosh began to strengthen important relationships with U.S. neighbors in North America and to reach out to neighbors in South America. On McIntosh's watch, the relationship between the U.S. and Canada was particularly strengthened, leveraging the momentum built from RDA Plenary 10 in Montreal. The partnership is helping both the U.S. and Canada build strong communities and create the leverage so characteristic of RDA.

4. WHAT NEXT? – RDA's Next Five Years

Since its founders' call in 2012, RDA has grown from 8 to 9000+ and created over 100 self-organized Working Groups and Interest Groups. The RDA community has generated many outputs used all over the world by individuals, projects, organizations, communities, and policy makers to advance data sharing and data-driven exploration. Moreover, RDA has become a community town hall, bringing together groups to develop data standards and practices, technical specifications, and common frameworks for communities of practice. By any measure, RDA has created the impact hoped for by its original sponsors and founders and attracted a vibrant, recognized community of infrastructure creators, developers, and users.

As an organization, RDA is maturing and also at a crossroads. How does the organization stay effective? What needs to evolve, grow, or be revised for RDA to remain useful and relevant? How can RDA seize current and future opportunities for greater impact? What are lessons learned from RDA's evolution from 2012 until now and how can they be incorporated into future strategies to promote continued success in the future?

All of these questions are the focus of continuing discussion for RDA's community and leadership. To create a foundation for these discussions, three challenges seem particularly important to tackle as RDA focuses on its next five years.

LEADERSHIP ALIGNMENT

As RDA has grown, the internal connections between organizational responsibility, accountability, and authority have loosened. Regions have been accountable for revenue but are not responsible for how it is managed. The Secretary General has the tough job of managing a dedicated and distributed staff, all

of whom report regionally to other supervisors. Council is tasked with both leadership of the organization and serving as its Board. Strategies that work for some parts of the community and some countries do not work for others. As RDA has scaled, the number of individuals playing multiple regional, international and strategic roles (creating informal alignment) has lessened and it would be useful for RDA to revisit its leadership structure to update it for greater effectiveness. More resources (which could support, for example, a less fragmented Secretariat) would help, but greater attention to roles, responsibilities, authority, and the relationship between RDA's constituent leadership groups will set the organization up to deal with further scale and create broader impact.

Externally, this is also a good time for the RDA to work collaboratively with other high-profile groups (e.g. CODATA, GO FAIR, WDS) to provide clarity on how the missions and efforts of these groups are differentiated, and to seize opportunities for synergistic coordination. This is both strategic and important as community members and stakeholders need to understand where RDA sits in the landscape, as well as how to value its role, for RDA to be a success. RDA's partnerships with external groups should focus on what is really important – useful data infrastructure that supports the community – in a way that leverages the strengths of each.

Internal and external leadership alignment are both opportunities and challenges for RDA. Evolving an updated version of roles and responsibilities for different groups and players with an eye towards 2023 (RDA's 10th anniversary) will help create the internal and external organizational infrastructure and partnerships needed for future impact and success.

SUSTAINABILITY

From its outset, RDA has been consistently challenged by the stability of its revenue sources and its accelerated growth. Funding for the organization is at most minimal with rarely more than 1-2 years of committed support for the Secretary General. It's hard to run an organization on a shoestring, and dealing with the urgent present makes it hard to plan strategically for the future.

Part of the problem is that commitment to RDA's support in the public sector has changed over time. Original investments in RDA have evolved to next generations of funding and shifts in funding priorities. Moreover, it is often easier to attract funding for the creation of a new organization than it is to maintain funding for an existing organization, no matter how successful. RDA struggles with an inadequate business model that does not work for an organization of its current size and maturity. Membership is free by design and Plenary registrations are modest to ensure a low barrier to access for everyone, but this also means that it is hard to make them revenue sources. RDA does not own intellectual property and gives its infrastructure away at no cost. This essentially means that RDA limits its revenue options to increase its impact. Ironically, there is not enough revenue to recruit professional business development assistance to carry out any new plan to increase revenue.

RDA's Council has been exploring new possibilities for funding including strategically expanding the set of funders and national stakeholders, fee-based workshops, regional "dues", etc. The expertise of the Funder's Forum is a particularly important resource that brings a broad perspective on fundraising to RDA and can help define new strategies for sustainability.

NEXT GENERATION MISSION/VISION

Finally, RDA's mission to build and deploy infrastructure to support data sharing and data-driven exploration remains at the core of the organization, but what does this mean as data becomes more ubiquitous and integral to virtually all exploration? How do we share data and do reproducible research in the highly decentralized Internet of Things environment we are building? How will today's standards and infrastructure hold up as data becomes more multi-purpose and its context, policy considerations (e.g. privacy), and integrity need to be protected in highly heterogeneous environments?

It may take many areas of research some time to catch up to the dynamism we currently experience in the "big data" area and especially with Artificial Intelligence. Nevertheless, it is important for RDA to be prepared to create and deploy the data infrastructure needed for both current environments and the future research scenarios we are likely to encounter.

RDA is not alone in these challenges and many community-driven organizations grapple with similar issues. If the past is a predictor of the future, RDA's passionate community will tackle today's and tomorrow's challenges with energy, foresight, strategy, pragmatism, and agility. Given the trajectory of RDA's first 5 years, the next 5 years are likely to be even more successful for the organization, and well worth looking forward to.

Acknowledgements

Many thanks to the outstanding and committed RDA members who helped fill in or vet the information in this document. They include Yolanda Meleco, Lindsay Poirier, Alan Blatecky, Chris Greer, Hilary Hanahoe, Ross Wilkinson, Juan Bicarregui, Ingrid Dillo, Josh Greenberg, Lynn Yarmey, Bridget Walker, and Mark Parsons.

I am grateful to many fellow travelers throughout the RDA experience including RDA's founders, the amazing RDA/US office, and Ross Wilkinson and my colleagues on Council. Special thanks to Chris Greer and Alan Blatecky for involving me in RDA at its inception, as well as to the tremendously supportive Program Officers from NSF, NIST, Sloan and MacArthur who have funded RDA/US during its first 5 years. Your faith in us has helped RDA build a strong, active, and impactful community, both within the U.S. and internationally.

REFERENCES

[**Adoption of Recommendations, n.d.**] Adoption of Recommendations. Retrieved from https://rd-alliance.org/recommendations-outputs/adoption-recommendations

[**ANDS, n.d.**] Australian National Data Service website. Retrieved from https://www.ands.org.au/

[**Berman, Collins, Stewart, Wilkinson, 2015**] (2015, October) Future Directions for the Research Data Alliance. Retrieved from https://rd-alliance.org/system/files/documents/Future%20Directions%20-%20Final%20Report.pdf

[**Beyond Sharing, 2011**] Friend, F., Guedon, J.-C., and Van de Sompel, H. (2011) Beyond Sharing and Re-using: Toward Global Data Networking. White Paper produced in the context of the EC FP7 call for scientific data infrastructure, https://www.webcitation.org/6GHNLQoiR

[**Blatecky, Bicarregui, Morais Pires, 2013**] (2013, December) Blatecky, A., Bicarregui, J., Morais Pires, C. White Paper: 5 Principles for an Open Data Infrastructure. Retrieved from https://epubs.stfc.ac.uk/work/12236702

[**Concept Paper – Data Web Forum, 2012**] (2012) Concept Paper – Data Web Forum, Coalition for Networked Information website. Retrieved from https://www.cni.org/wp-content/uploads/2012/06/DataWebForum_Concept_Paper.pdf

[**Disaster-STS Research Network, n.d.**] The Disaster-STS Research Network website. Retrieved from http://disaster-sts-network.org/

[**First Plenary Meeting, 2013**] (2013, March) RDA First Plenary Meeting, Gothenburg, Sweden. Retrieved from https://rd-alliance.org/plenaries/rda-first-plenary-meeting-gothenburg-sweden

[**G8+O5 Global Research Infrastructure, 2011**] (2011, October) G8+O5 Global Research Infrastructure - Sub Group on Data Draft Report. Retrieved from https://epubs.stfc.ac.uk/work/24111652

[**G8 Science Ministers Statement, 2013**] (2013, June) G8 Science Ministers Statement. Retrieved from https://www.gov.uk/government/news/g8-science-ministers-statement

[**Groups, n.d.**] RDA Working and Interest Groups. Retrieved from https://www.rd-alliance.org/groups

[**IETF, n.d.**] Internet Engineering Task Force website. Retrieved from https://www.ietf.org/

[**NCRIS, n.d.**] National Collaborative Research Infrastructure Strategy. Retrieved from https://www.education.gov.au/national-collaborative-research-infrastructure-strategy-ncris

[**RDA in a Nutshell, 2019**] (2019, April) RDA in a Nutshell [RDA website]. Retrieved from https://rd-alliance.org/about-us/communication-kit/rda-nutshell

[**RDA Regions, n.d.**] RDA Regions. Retrieved from https://www.rd-alliance.org/groups/rda-regions

[**RDA TAB Responsibilities and Processes, n.d.**] RDA TAB Responsibilities and Processes. Retrieved from https://zenodo.org/record/1313432#.XMihnKQpA2w

[**Research Funders and Stakeholders Interest Group, n.d.**] Research Funders and Stakeholders on Open Research and Data Management and Policies Interest Group. Retrieved from https://rd-

alliance.org/groups/research-funders-and-stakeholders-open-research-and-data-management-policies-and-practices-ig

[**Riding the Wave, 2010**] (2010, October) High Level Expert Group on Scientific Data Infrastructures, Riding the Wave: How Europe can gain from the rising tide of scientific data. Retrieved from http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=707

[**The Asthma Files, n.d.**] The Asthma Files website. Retrieved from http://theasthmafiles.org/

[**Working Group Outputs, n.d.**] Working Group Outputs. Retrieved from https://www.rd-alliance.org/groups/creating-and-managing-rda-groups/working-group-outputs.html