



# Fostering Interdisciplinary Data Cultures through Early Career Development: The RDA/US Data Share Fellowship

INNA KOUPER

LOIS A. SCHEIDT

BETH A. PLALE

*\*Author affiliations can be found in the back matter of this article*

COLLECTION:  
RDA SPECIAL  
COLLECTION

RESEARCH PAPER

][ubiquity press

## ABSTRACT

Openness and interdisciplinarity in research and data are among the challenges that are frequently discussed in the context of changing scientific and scholarly practices. Gradually, the visions of open and widely shared data are being reconciled with complex realities that stem from the disciplinary differences in data cultures. In this paper we discuss interdisciplinarity through data as a way to create research environments that are more flexible and, as a result, more amenable to change. We report our findings from facilitating and evaluating a data-oriented early-career fellowship program that was administered as part of the Research Data Alliance (RDA), a global organization that aims to enable open sharing and re-use of data. We identify ways to foster interdisciplinary data cultures among the future researchers and professionals and propose recommendations for future programs. While the short-term early career programs cannot address the systemic factors that impact openness and interdisciplinarity, such as the systems of reward and recognition or the funding structures, they can introduce mechanisms that support diversity, learning, and leadership and, ultimately, contribute to a culture change.

## CORRESPONDING AUTHOR:

**Inna Kouper**

Department of Informatics,  
School of Informatics,  
Computing, and Engineering,  
Indiana University  
Bloomington, USA

[inkouper@indiana.edu](mailto:inkouper@indiana.edu)

## KEYWORDS:

early career training; data  
culture; open science; data  
stewardship

## TO CITE THIS ARTICLE:

Kouper, I, Scheidt, LA and  
Plale, BA. 2021. Fostering  
Interdisciplinary Data  
Cultures through Early Career  
Development: The RDA/US  
Data Share Fellowship. *Data  
Science Journal*, 20: 2,  
pp. 1–19. DOI: [https://doi.  
org/10.5334/dsj-2021-002](https://doi.org/10.5334/dsj-2021-002)

Openness, transparency, and interdisciplinarity in research are among the challenges that are frequently discussed in the context of changing scientific and scholarly practices. The initial push for data sharing from the funding agencies in the US and Europe decades ago has been met with indifference or resistance from individual researchers, who are reluctant to share data (Campbell et al, 2002; Tenopir et al, 2011). To keep the momentum, the efforts have shifted over time from the calls to share and re-use data to initiatives that focus on tiered data management that accommodates heterogeneous interests and encourages openness and cross-disciplinary interactions while providing necessary protections (Anderson, 2017; Plale et al, 2019).

The visions of open and widely shared data are being gradually reconciled with complex realities that stem from the disciplinary differences and variety of data. As more and more sources and tools are used for data collection and analysis, new forms of expertise are needed to do so. Both in and outside of academia there is a need for individuals who are comfortable and proficient in data-intensive environments and have both the domain expertise and broader skills that allow them to expand their methodological repertoire and use, analyze, and preserve data that comes from multiple sources.

The global and collaborative nature of research is another trend that challenges the existing data practices and training. Research is often done outside of a specific lab, with contributions from such groups as IT, libraries, and data analysts. In addition to being comfortable with data-intensive work, one needs to be able to work in diverse environments. Exposure to international policies and practices and extensive networking can help overcome the isolation some feel within their academic fields and support interdisciplinary work (Porter et al, 2006).

Formal education programs are inevitably slow to respond to the changes in research environments and data practices. While many efforts exist to modify curricula to enable data thinking and action (Cao, 2018; Varvel et al, 2012), creating new effective programs is a long process that includes institutional, social, and technological change. Moreover, as some form of change is necessary in almost any profession, from humanities to astronomy, not everyone will be able to undergo formal training to acquire the ability to work in global data-intensive environments. Shorter programs that work across disciplines and complement core knowledge of the domain with skills in interdisciplinary data work can help fill this gap.

In this paper we discuss paths toward interdisciplinarity, openness and transparency through an early-career fellowship program that was administered as part of the Research Data Alliance, a global organization that aims to enable open sharing and re-use of data. First, we describe our approach to facilitating and evaluating the program that had the goal of exposing students from various backgrounds to the challenges of data stewardship and discuss its processes and outcomes. Second, we use our empirical findings as well as the existing literature to propose recommendations for future programs and identify ways to foster interdisciplinary data cultures among the future researchers and professionals.

## BACKGROUND

### EARLY CAREER DEVELOPMENT

Two concepts are useful in understanding the processes of becoming part of a profession and related early career development: socialization and acculturation. *Socialization* is the process during which people internalize shared understandings that exist in the society and learn the roles, norms, and values necessary for participation in it (Berger and Luckmann, 1990). Professional socialization involves learning the values and behaviors of a particular profession, which happens through a combination of formal and informal, or 'hidden' curricula (Hammer, 2006). The formal curriculum helps students to hone their knowledge and skills. The 'hidden' curriculum is delivered through various forms of mentorship, networking, and communication in which students participate at their institutions, in professional societies, and within peer networks. As students and early career professionals ('early careers' or ECs hereafter) observe others in professional environments, they learn the needed attitudes and behaviors and, presumably, adjust their pre-existing attitudes and behaviors.

*Acculturation* is a process of adaptation to a new cultural context that creates a new sense of identity (Berry, 2003). In the professional context, acculturation means adapting to the cultures of academia and its institutions within a certain locality and disciplinary orientation as well as learning the associated values and norms of conduct. As can be seen from definitions, the terms ‘socialization’ and ‘acculturation’ have overlapping meanings as they both encompass individual and group-level changes and distinguish between using the strategies of learning, adaptation, and integration (Dinmohammadi et al, 2013). In this paper we use the term ‘acculturation’ to emphasize the cultural conception of knowledge production and disrupt its epistemic unity, i.e., the assumption that there is only one type of knowledge, one scientific method that contributes to its production, and one culture to which scholars and scientists belong (Knorr Cetina, 2007, 1999). The cultural conception of knowledge also allows to highlight the differences in working with research data and the need for understanding multiple contexts and overlaps that exist within data-related activities.

Learning the techniques of data collection and analysis as well as adopting the norms of data sharing and archiving is part of professional acculturation, which takes a variety of forms, including early career scholarships and fellowships. Mostly common in medical and clinical fields, a fellowship is usually a relatively short-term yet intense engagement, frequently offered with a financial reward, that allows ECs to further or complete their training and solidify their skills in teaching, research, service, communication, and leadership (Ford Foundation, 2019; University of Birmingham Research and Commercial Services, 2009). It also acculturates the ECs into the field by deepening their connections to practice and helping them adapt to professional values and behaviors and resolve inconsistencies if needed (Handelsman et al, 2005). When a program has specific goals and committed mentors, it can have a strong effect on the ECs’ success in establishing their research agenda, finding external funding, publishing, and securing a job (Burns et al, 2015; Hickey et al, 2014; Shtasel et al, 2015).

Professional acculturation, however, becomes complicated when one is aspiring to thrive in a culture that prioritizes diverse data and technical skills as an addition to primary domain knowledge. Broad data training is different from data science as a disciplinary orientation, adding an interdisciplinary dimension to early career development. Attempts to support ECs in the data-intensive cultures resulted in a number of programs that have emerged in recent years. Some of these programs support diverse data science careers and training (MSDSE, 2019), while others focus on interdisciplinarity or data management and curation skills (DataONE, 2019; Keralis, 2012; Martin and Umberger, 2003). A number of fellowships focus on promoting openness in data, software development, and scientific methods (Mozilla, 2019; School of Data, n.d.).

Most of these programs are still defining their models of supporting open data and interdisciplinarity. This evolving character along with the long-term sustainability problems contributes to the lack of evaluation of these programs in their impact on ECs and data cultures. Exploratory qualitative studies have shown, though, that the challenges of acculturating ECs into open and transdisciplinary data cultures are associated with the rigid academic reward system, ECs’ unfamiliarity with the changing community norms, and lack of training and support (Koch et al, 2018; Nicholas et al, 2019). How can interdisciplinary acculturation help to develop skills to thrive in data-intensive environments, while learning to appreciate epistemic diversity and different types of data? How can interdisciplinary data cultures balance disciplinary and problem-oriented career development? In what follows, we examine a program that was designed with these questions in mind and use its outcomes to inform future programs and research.

## **RDA/US DATA SHARE FELLOWSHIP**

The RDA/US Data Share Fellowship program was established in 2014 to support the engagement of early career professionals (ECs) in the Research Data Alliance (RDA) – an international initiative that builds the social and technical infrastructure to enable data sharing (Berman et al, 2014). The RDA work is carried through the establishment of Working and Interest Groups (WGs and IGs) that develop technical solutions as well as best practices, recommendations, and surveys of community needs. As of August 2019, there were 36 Working Groups and 66 Interest Groups that cover a wide range of topics from legal issues and data interoperability to data science education and bibliometrics. In addition, Birds of a Feather (BoF) groups are convened at the biannual RDA Plenaries to gauge the interest in new topics. With over 8,800 members from

137 countries representing research, data management, publishing, government, and many other communities and stakeholders, RDA as an organization provides many opportunities for discussions, leadership, and expertise (RDA, 2016).

Building upon a pilot program (Kouper, 2014), the fellowship offered financial support, networking, and mentorship opportunities for young scholars and professionals to facilitate the development and advancement of their data-related interests and to strengthen RDA and its US-oriented activities. The RDA/US Data Share Fellowship program differed from other programs in the following:

- Central to the scope of the program was exposure to international data and data sharing norms
- Selection criteria ensured a fit between ECs and RDA vision and projects, therefore prioritizing applicants who emphasized data as the object of their interest
- Fellows' applications were evaluated for having both data and domain expertise

The RDA/US Data Share fellowship was open to graduate students, postdoctoral researchers, and early career professionals at institutions of higher education in the United States without consideration of the applicant's country of origin. In 2015–2018 the program distributed an annual call for applications inviting interested applicants to submit their proposals that described the topics they were interested in and potential connections with RDA WGs and IGs. Program coordinators and directors, in consultation with the Advisory Board, facilitated the review process and provided guidance and training to the fellows. Depending on the complexity of the projects, fellowships ran for either 12 or 18 months. Fellows were expected to develop a portfolio including reports, presentations, data releases, and technological artifacts related to their fellowship experience.

New fellows selected annually attended a cohort-building orientation workshop at the beginning of their fellowship and then worked remotely from their home institutions, while participating in online sessions and attending RDA plenary meetings, bi-annual events that are held in different places around the world and that bring together data managers, librarians, computer scientists, and domain scientists to work on improving data sharing. Plenary attendance was part of the fellowships, and funds were provided for the fellows' travel, lodging, and meals. While in attendance fellows were expected to 1) participate in the activities and deliberations of the WGs and IGs, 2) present a poster related to their project, and 3) take part in RDA/US Data Share fellowship activities at the plenary. Program coordinators helped fellows find WGs and IGs that aligned with their professional and educational activities and encouraged fellows to contextualize their work in the broader RDA membership, institutions, and organizations.

The RDA/US Data Share program complemented the existing fellowship programs by broadening the pool of applicants, by being intentionally flexible in the forms of engagement and dissemination, and by conceiving barriers to openness and data sharing along the socio-technical continuum. It relied on internal and external evaluation to ensure its quality, consistency, and impact. Both evaluations included surveys of attitudes, analysis of progress reports, self-assessments, and observations during plenaries and other events. External evaluation in year 1 and 3 of the project was summative in nature and helped to determine whether the program was positively affecting fellows' careers or research outcomes. The materials from these evaluations serve as data sources for this study.

## METHODOLOGY

To facilitate an in-depth inquiry into the program, its outcomes, and larger impact on data cultures, this study was organized as a case study. The relatively small number of participants and a need to protect their anonymity prompted us to use the whole program as a unit of analysis. Rather than examining individuals and comparing them within and across the cohorts, we collected information about how those individuals perceived the program and its role in their professional development. To go beyond an evaluation of the program, our case study approach was combined with the grounded theory and participant observation methodology in data collection and analysis (Charmaz, 2014). Such combination allowed us to gain a deeper understanding of both the bounded phenomenon of the case, the program itself, and the larger issues of fostering data cultures that such program may raise.

The study, approved by Indiana University Institutional Review Board (protocol # 2007944155), relied on a broad range of data sources, including feedback surveys, interviews, observations from external evaluations, field notes, and fellows’ deliverables and profiles (see [Table 1](#)). The authors of the paper have participated in the program as coordinators and evaluators and collected data at various stages, using parts of it for programmatic purposes and, over time, constructing a rich data set for subsequent analysis.<sup>1</sup> Field notes included notes taken during meetings and plenaries that documented changes and evolution of the program as well as interactions among the fellows and others who came into contact with the program.

DATA SOURCE	INDIVIDUALS PER DATA SOURCE	DESCRIPTION
Field notes	49	Field notes from program coordinators and external evaluators that include observations about fellows and their interactions with the fellowship and RDA.
Fellow CV	49	Resumes and curriculum vitae of accepted applicants.
Fellow application statement	49	Statements submitted by fellows during the application process. Statements included brief descriptions of the applicant’s research interests and relevance of RDA to their career.
Orientation survey	32	Follow-up surveys that were sent to each cohort after the orientation with questions about fellows’ perceptions of the event, the RDA and the fellowship.
Plenary survey	51	Follow-up surveys that were sent to fellow attendees after each plenary with questions about fellows’ perceptions of the event and their participation in it and the RDA.
Fellow interview	35	Structured interviews with the fellows, containing questions about fellows’ perceptions of the program, the organizers, and the benefits to their career. The cohort of 2014 was not included in the evaluation process and therefore was not interviewed.
Program coordinators interview	4	Structured interviews with program coordinators, containing questions about the coordinators’ role in the program and their perception of its activities and impact.
Mentor interview	6	Structured interviews with mentors, containing questions about the mentors’ role in the program and their perception of the fellows’ performance. The formal mentorship program was in place only in 2015.
Fellow final report	17	Reports that were submitted by the fellows at the completion of their fellowship.
Fellow poster	60	Fellows’ posters presented at the plenaries as part of their program (some fellows presented more than one poster).

**Table 1** Sources of data for the case study (see also supplementary files for survey and interview questionnaires).

All the data was organized into a database that grouped documents of the same type and contained brief descriptions of each data source and its contents. Recorded interviews were transcribed and saved as text documents for subsequent analysis. Several interviews were not recorded, and extensive textual notes taken by the interviewer were used for analysis in those cases. Documents within each source type were organized chronologically, by the cohort. Materials quoted in this paper have been anonymized and may include minor editing for clarity.

A sample from each data source was examined independently by two coders who identified the initial themes and discussed overlaps and deviations in interpretations of the themes. The initial coding focused on the processes and activities within the program and on the meanings that participants attributed to those processes. Additionally, the coders used common data management terminology, e.g., terms such as ‘curation,’ ‘preservation,’ ‘sharing,’ ‘metadata,’ and so on as part of the *in vivo* coding that assisted in theme development (Ritchie and Lewis, 2003). The coding proceeded in several iterations, where the coders continued to raise questions about the themes in the data, addressed the gaps and relationships between the themes, and combined the overlapping themes into a reasonable number of categories. After three iterations of going through the data, the coders finalized the themes and generated hypotheses about how the themes were related and what conceptual abstractions they supported. These hypotheses and abstractions are presented below.

<sup>1</sup> Due to the small sample size of the dataset and an increased risk of de-anonymization, the data cannot be shared publicly. The interview and survey protocols are shared as supplementary materials. Researchers interested in examining the data should contact the corresponding author.

ECs in the RDA/US Data Share program went through many acculturation experiences that included both disciplinary and interdisciplinary aspects. Those experiences can be grouped into four larger themes: experiencing diversity, developing professional ties, learning the data landscape, and becoming leaders.

**EXPERIENCING DIVERSITY**

While participating in the program, the fellows have had opportunities to interact with many individuals and groups that varied in their demographic, educational, and other characteristics, both within the program and as part of the RDA community. The program supported four annual cohorts between 2014 and 2017. The first 2014 pilot cohort consisted of two types of participants – the fellows and the interns. Each recruited cohort had individuals of both males and females and had varied characteristics in ethnicity and nationality and career stage categories. There was often almost a 50/50 split between males and females in each cohort (see [Table 2](#)).

	2014 INTERN/FELLOW	2015 FELLOW	2016 FELLOW	2017 FELLOW	TOTAL
Female	6	3	7	9	25
Male	8	5	7	4	24
<b>Total</b>	<b>14</b>	<b>8</b>	<b>14</b>	<b>13</b>	<b>49</b>

**Table 2** Gender representation across cohorts.

While the majority (about 73%) of fellows came from the US, close to one third (27%) were citizens of countries in South, East and West Asia, and the Middle East. This breakdown is comparable to the graduate enrollment statistics in the U.S. with international students constituting 18.5% in Fall 2017 (Okahana and Zhou, 2018).

In addition to inherent diversity, the program had a noticeable degree of acquired diversity, which included experiences and education. The RDA/US Data Share program defined ‘early career’ as being a graduate student or having received a graduate degree within 3 years at the time of application. As a result, each cohort had a mix of students, postdoctoral researchers, and early career professionals, with graduate students representing about half of the fellows (see [Table 3](#)).

TITLE CATEGORY	FREQUENCY	PERCENT
Academic (e.g., assistant professor, lecturer, other faculty)	8	16%
Professional (e.g., data manager, librarian, software developer)	8	16%
Graduate student (doctoral or master’s)	25	51%
Postdoctoral fellow	8	16%

**Table 3** Fellows’ position title categories.

In their feedback on program orientation and plenaries most of the fellows commented that meeting others and learning about their varied backgrounds was one of the most positive experiences of the fellowship. The combination of structured and unstructured interactions increased the perceived benefits of mutual sharing and learning:

‘I liked getting to know the other fellows and meeting the fellowship team face to face. And having bonding moments!’

‘Getting to know the other fellows and their interests. ... This actually occurred mostly during the downtime, but the organized interactions helped with getting to know the less talkative.’

'I especially enjoyed the exercise of speculating on possible collaborations as it pushed us to consider additional collaboration even as we were beginning a new fellowship project.'

The fellows often discussed their differences and similarities during orientation and other events, trying to find common ground and ways to connect. During our observation we documented occasions where fellows have discussed the issues of gender and childcare in academia, stereotyping and racism, and methodological purism in some research areas. Most of these discussions took place during breaks from formal programming and informal meetups organized by the fellows. In their interviews fellows also acknowledged the positive effect of meeting people who come from other disciplines or backgrounds. The mixture of career stages and academic positions within each cohort helped ECs to learn about the challenges of moving from one career stage to another, in and out of academia, or between disciplines.

Despite the efforts to attract applicants from various disciplines, many applications came from the areas that are related to information processing and technologies. The topics of research data and the activities of RDA as an organization have attracted many applicants from the library and information science area. As a result, almost half of the program participants were from that disciplinary area (47%, see [Table 4](#)). At the same time, the rest came from a variety of disciplines that cut across computer science and engineering, life and social sciences and some interdisciplinary areas such as public health or biostatistics.

DEGREE AREA	FREQUENCY	PERCENT
Computer science/Engineering	8	16%
Life sciences	5	10%
Library and information science	23	47%
Physical sciences	2	4%
Social sciences	4	8%
Other (e.g., statistics, public health, IT)	7	14%

**Table 4** Fellows' degree areas.

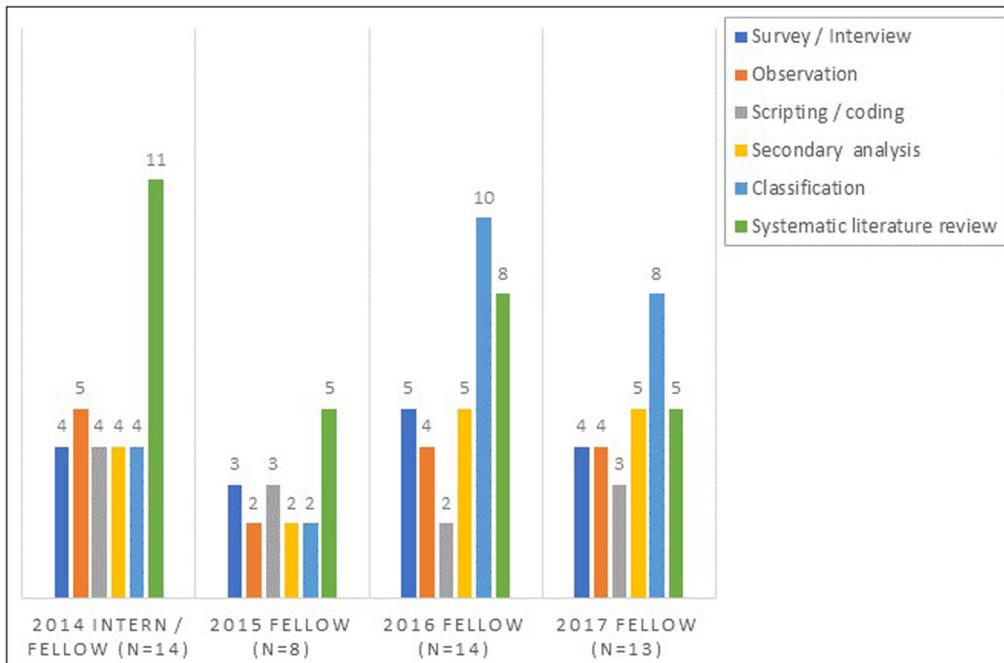
This prevalence of library and information science orientation was noticeable to some fellows who commented that this could have a negative effect on diversity of experiences and connections and called for more outreach to other disciplinary areas:

'So if there were future RDA fellows, if there was another round of co-funding, that there needs to be expanded outreach into other fields. I found that, it wasn't in my cohort, it was very much library sciences folks, and that's great, but there needs to be more diversity.'

Besides demographics and education, the program encouraged diversity of project designs. Fellows proposed and implemented projects that ranged from examining metadata standards in specific domains to studying data needs and behaviors to designing discovery platforms and workflows, their methods included surveys and interviews, writing code, and performing secondary analysis (see [Figure 1](#)).

Many fellows used more than one method in their research, and none of the cohorts had projects that relied on statistical modeling, quantitative methodologies, or experiments. The first cohort had more projects that relied on more than one method with an average of 3 methods per project. All other cohorts averaged around two methods per project. Many fellows, especially in 2014, 2015, and 2016 cohorts, included systematic literature review into their projects (11, 5, and 8 fellows correspondingly). The method of classification used by a small number of fellows in each cohort refers to the analysis of existing entities, such as repositories, metadata schemas, user profiles, or databases and grouping them into classes according to the criteria developed by the fellow. Fellows also used surveys, interviews, and observations as well as programming techniques to work on the topics of their projects.

Even though some fellows have used scripting or coding in their projects, the overall methodological orientation of the projects can be described as qualitative and exploratory. The limited duration of the program required the coordinators to select projects that were smaller



**Figure 1** Use of methods in fellows' projects (each fellow could use more than one method).

in scope, achievable within one year. This created an environment that could be described as less engaging for fellows who were interested in quantitative approaches or had ambitious plans to implement a full-scale research project, which could have decreased the theoretical and methodological diversity of the program. Over time, several fellows who had interest in quantitative and experimental research designs began to withdraw from the program activities. One such fellow, for example, when asked to contribute to a collaborative project within the program, refused, arguing that it was unclear who would benefit from such an effort.

The analysis of program applications and subsequent contributions to the program demonstrated that ECs' levels of activity and patterns of participation in RDA corresponded to their awareness of diversity. Thus, several participants within each cohort mentioned in their applications the increasingly diverse nature of data, practices, infrastructures, and research personnel. Those fellows appreciated the active efforts of the program to provide diverse experiences and noted that they learned a lot from such a diversity:

'The spirit of RDA and the energy and initiative of its members, in particular the organizers of the fellowship program made participating in the community an absolute pleasure. The diversity of interests in the community is fascinating and makes meetings both educational and enjoyable.'

Although this needs to be tested in a larger sample, our analysis shows that fellows' attitudes toward diversity and interdisciplinarity had an effect on how engaged they were and how they formed relationships with other fellows. Those who focused narrowly on their own research, disciplinary practices and career achievements tended to be less willing to participate in discussions or collaborative activities. Those who already had some previous exposure to or understanding of diversity were more open to new opportunities and further benefitted from mutual sharing and learning:

'I would also point to the fact that the two of the other fellows and I have continued to kind of work on the side projects that are both related to RDA and not related to RDA... we continue to have conversations offline, we continue to talk about and share our work with one another, so I think those are also tangible outcomes, right. They're pointing to the development of a cohort of colleagues'.

## DEVELOPING PROFESSIONAL TIES

Even though the program was positioned as part of the global organization that encourages interdisciplinarity and cross-pollination, fellows within the program sought their respective disciplinary communities first as they wanted to form closer ties with members of those communities. This was evident in the discussions and activities around mentorship and the

acknowledgement that professional obligations have a significant impact on fellows' behavior. Thus, making connections that would allow the fellows to collaborate, submit grant proposals and publish within their disciplines was among the primary goals for young researchers. The program helped ECs to gain confidence within their discipline and branch out into their own work apart from advisors and mentors:

'I personally got a lot out of attending [a plenary]. ... I met [name], the chief editor of a journal that I am planning to submit an article for publication. I was able to describe to him the idea I had for the paper and get helpful tips on how to improve the paper which will hopefully increase the likelihood of its acceptance.'

Networking within the organization was one of the largest components of how the fellows developed ties to their professional communities and beyond. Despite their disciplinary commitment, most fellows recognized the importance of wider and more diverse ties and often formed such ties, either on their own or with the help of program coordinators, who extended their own networks and introduced fellows to people from a variety of backgrounds. Professional ties extended into diverse transdisciplinary ties when domain-oriented ECs paired with computer scientists and vice versa or researchers engaged in the discussions with librarians and data managers. As some fellows pointed out, meeting people from academic and practice-oriented communities as well as from businesses and governments provided them with new perspectives on their research and its impact and on ways to collaborate.

Fellows also networked across geographical regions beyond the US. The international reach of RDA was a noticeable component of the fellowship, although its long-term impact is not clear. One of the common benefits of international exposure was learning about other regions' data challenges and ways to address them. Similar to other interactions, in international networking fellows looked for disciplinary connections as the quote below illustrates:

'My research topic that I did for RDA was very-very particular in that not many researchers are doing research in the area so being connected with a group of people who are doing work in the area was really-really helpful. And the network was international so I was also connected with some international researchers in [regions] which leads me to initiate another collaborative project so that was really helpful.'

Several fellows established international collaborations as part of the working and interest groups, although such collaborations stopped or slowed down due to the more immediate professional obligations or lack of funding. Lack of funding and time was mentioned several times as one of the main obstacles in engaging in new cross-disciplinary and cross-regional collaborations.

Another challenge in forming interdisciplinary professional relationships and learning to navigate transdisciplinarity was the topical breadth of RDA as a diverse and global organization. For individuals at the early stages of their career such breadth creates difficulties in joining conversations and making connections. Fellows sometimes felt frustrated with the need to communicate across multiple disciplinary languages, especially in large virtual collaborations. They had to learn to overcome those challenges through utilizing multiple forms of communication and asking for mediated introductions. Those who learned to communicate early and be persistent were able to better align their work within the organization and receive help:

'I believe that communicating with a working/interest group from early on is essential. I should have started the discussion with the group before I made my initial project plan. To stress the point, once I was in continuous contact with one of the group chairs it became easier to manage my work. It also lent credibility to my efforts and helped me communicate better with my collaborator.'

Our program began with a formal mentorship model, where in addition to reviewing and selecting applicants, we also reviewed ongoing RDA projects and matched fellows to them. This resulted in several ineffective or incompatible matches, so in subsequent years the program implemented informal mentoring system through the position of the program ambassador. The main ambassador's role was to regularly communicate with the fellows, learn about their needs, and introduce fellows to RDA members who might be helpful to them.

The ambassador position also helped to address the challenges of mentorship across disciplines. It is difficult to match individuals for interdisciplinary mentoring; if the mentor and mentee come from different disciplinary backgrounds, the relationship cannot be built on the shared domain knowledge or techniques. Rather, mentors and mentees have to seek common ground in such areas as potential career paths, methodological diversity, or data-to-domain bridges. Although there was some frustration with how busy and not always available the mentors were and some fellows voiced a desire for an assigned RDA mentor, whose interests were closely linked to their own, many fellows appreciated informal interdisciplinary mentorship and the breadth it provided:

‘And what I liked about RDA is that not only are they giving me career advice but they were also like in fields or tangential fields of what I’m doing. So it’s not just like my major professor who knows me and gives me good career advice. When someone doesn’t know as much about you to give you that advice is more valuable I guess. ... that’s the biggest thing I kind of walked out with.’

Despite some dissatisfaction with matching and the amount of mentoring within the program, all fellows recognized the importance of mentorship and many expressed willingness to become mentors in the future. The flexibility of our approach and mediation of the ambassador allowed for mentorship to take many forms and rather than focusing on matching and choosing the right individuals, to support the variety of ECs needs, including professional development, emotional support, a sense of community, role modeling, and creating a space for discussions.

## LEARNING THE DATA LANDSCAPE

The fellowship has been advertised as a program that focuses on data issues, particularly, the issues of openness and sharing. The applicants, especially those who were subsequently selected to become fellows, have demonstrated awareness of the data issues in their statements. They discussed their experiences of working with data, lack of sharing in their disciplines, and the challenges of publishing data or finding data sets for re-use. Many expressed dissatisfaction with the existing state of knowledge stewardship in their domains and asked to join RDA to improve the existing practices.

Applicants approached their statements from the traditional proposal writing perspective. They identified a gap in the literature or in practice, described its significance, proposed steps to address this gap, and developed a timeline for their projects. In their statements they communicated confidence in their skills to complete the projects. Very few of them acknowledged the need to learn more about data issues even though the challenges of data stewardship are still not commonly taught in academic programs. Rather than focusing on learning, the applicants considered the fellowship as a way to build their professional networks and leadership and research skills and advance their careers:

‘In the next five years, with opportunities provided by RDA/US, the support of a strong professional network, and my leadership skills, I can play an important role as a researcher and educator in providing better data access and distribution infrastructure...’

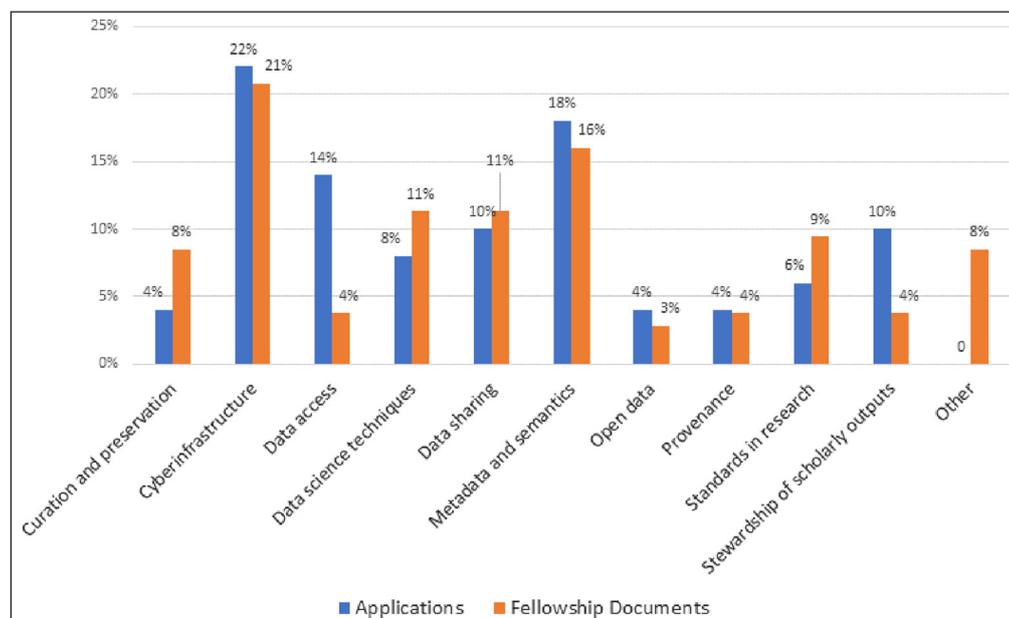
Over the course of their fellowship, however, many fellows expanded their perspectives through the exposure to the global and diverse networks of RDA. The Research Data Alliance is an organization that focuses on improving research and development through improving data practices. For ECs it meant that in addition to learning their disciplinary practices, they were learning about open science and transparency movements, the practices of data stewardship, data organizations, both RDA and others, and about data science techniques: ‘Seeing the data challenges that different communities face and proposed solutions to these problems, from various perspectives, was extraordinarily helpful.’

Fellows also broadened their view of the data world:

‘It helped me get a sense of the broader data landscape and that was definitely valuable. I didn’t do as much so that was a big help. I did get involved with groups that were not necessarily part of my fellowship project, so that was helpful.’

‘And so I just think that, that kind of opening my eyes as an early career researcher, to the vast landscape that is out there, that I can be a part of. It was really cool.’

Fellows learned through both discussions and practical work. They engaged with WGs and IGs and collaborated along the many aspects of the data landscape, sharing and acquiring expertise in the areas of programming and data science techniques and data management and curation. *Figure 2* below illustrates changes in the themes that ECs discussed at the time of their application to the fellowship and the themes they addressed while being in the program as evidenced in their plenary surveys, interviews, posters and final feedback.



**Figure 2** Themes in ECs applications (N = 49) and documents during the fellowship (N = 112).

As can be seen from the image above, over the course of the fellowship ECs’ awareness and attention remained almost the same with regard to some issues and shifted with regard to others. Thus, attention to cyberinfrastructure, or creating tools, platforms, and other technical support to enable data sharing remained relatively high in both application statements and other documents (22% and 21% correspondingly). Fellows made these issues a prominent part of their applications and later they acknowledged that they learned about many cutting-edge developments in data repositories, platforms and programming languages, and distributed management systems. As one fellow put it, ‘RDA is still the best venue I have found for learning about large scale and global data sharing infrastructure development and issues.’ Other themes that were stable and relatively high in the levels of attention included metadata and semantics (18% and 16% correspondingly) and data sharing (10% and 11%).

Awareness of the issues of curation and preservation as well as the issues of standards in research have increased. While at the beginning, curation and preservation was mentioned only in a couple of applications of those who specialized in these areas, more fellows have later realized the importance of taking care of research data early on to avoid poor quality, shallow or incorrect insights, and loss of data. In their posters and plenary observations fellows discussed the needs and challenges of curating and preserving geoscience and climate data, data from specific populations such as indigenous communities and the need to educate others on these topics. With regard to the standards in research, fellows realized how much needs to be done in establishing the standards for publishing data and software, for policies around data, and for domain-specific standards that encourage data exchanges.

The topics that decreased over time included data access and stewardship of scholarly outputs. In application statements these were the topics that were the most broad and non-specific and included calls for improving access to data or for the need to improve how we share the results of research and scholarship. Over time, through their experiences at RDA fellows deepened their knowledge of these topics and discussed them in more concrete terms, thus transforming them into the issues of curation, sharing, data science techniques, and other topics. Several topics that were not addressed in the application statements became visible in other documents.

Such topics included data citation, interoperability, ethics, and data governance. All these topics are regularly addressed with the RDA events and discussions.

Learning the data landscape also included understanding how data organizations work as they do not necessarily organize their events and meeting the same way as discipline-oriented conferences or meetings of other professional societies do. Thus, when asked about what worked for them in the program, fellows commented about their confusion about the RDA plenaries at first and the time needed to understand the purpose of the organization,

‘For the most part these seemed to lack purpose, at least compared with other conferences I have been to that are more individual project focused. But I think I am alright with that now that I better understand the purpose of plenaries.’

From the fellows’ perspective, data organizations such as RDA provide a venue for interdisciplinary international interactions that bridge gaps between practice and research. Data professionals come from a variety of backgrounds and there is cooperation between the government, business and academic sectors. Those who are confronting data challenges in their work meet those who know how to address those challenges. The interactive style of plenary sessions and the format of working groups was perceived as more purposeful and engaging than the regular ‘talk-at-you style of conference’, although those who were interested in advancing their research found RDA plenaries somewhat difficult to contribute to:

‘...one of my challenges that the more I’ve learned about how the organization works the more I’ve realized that my project was perhaps a little bit too theoretical and not quite focused on bridging the gap between research and practice enough...’

Nevertheless, many fellows mentioned the importance of such organizations as places where change in the practices of data and knowledge production could happen.

## BECOMING LEADERS

The change from a novice to a knowledgeable professional and, moreover, a leader in the field, was hard to observe in a short timeframe of the fellowship, but the way ECs talked about their experiences sometimes indicated that they were making this transition. Fellows mentioned how they learned the useful methodological skills, e.g., interviewing, coding, and so on, and the skills that are part of the changing academic landscape, e.g., teamwork, negotiating interdisciplinary boundaries, and appreciating the disciplinary diversity. As they discussed the issues of data, fellows shifted from the ‘learning’ mode to the ‘problem solving’ mode. The quote below, for example, shows ECs as active participants in the emerging changes and uses a collective ‘we’ do describe contributions (*emphasis added*):

‘I think that most early career fellows **want to see their disciplines advance data infrastructure**, and it was great having a space where **we could share ideas about how to move this forward**.’

In addition to learning the elements of both the traditional academic cultures and the data landscape, the fellows used their experiences to advance their respective fields, considering their accomplishments within the program as part of their personal and professional growth. ECs recognized their role of becoming the leaders in their professions in the future as well as becoming potential agents of change in academia and data-related professions. They were well aware of the discussions about openness and accountability in science as well as the need to encourage data sharing. They also understood that the existing fragmentation of scientific research and its division into technical and non-technical domains creates gaps in building cyberinfrastructure and wanted to address these gaps by organizing joint interdisciplinary sessions,

‘I ran a session called [session name]. I was very excited about this session because I think the discussion was very engaging and that it helped build connections between the more technical WG chairs and those in the [other] domain. I was asked to send the findings that I presented in this session to TAB [Technical Advisory Board], so I felt as though it had potential to make a real difference in RDA.’

Fellows' growth into leadership happened in several different ways. Through their application proposals and subsequent work on projects they learned to manage their time and adjust to the changing circumstances of the program that were beyond their control. They learned to manage their projects alongside their coursework or other professional responsibilities and think about their work through the lens of interdisciplinary community that values openness and responsible handling of data. They also tried some ambitious ideas and learned to accept and manage both the successful and unsuccessful parts of such ideas.

Through their networking and communication experiences, fellows reached out to external entities, both organizations and people, and established connections in search for funding, partnerships, and so on. They found collaborators and submitted papers for publication, conference presentations, and grant proposals. As ECs worked through their collaborations, they began to reflect on this topic in a broader sense as something that needs cyberinfrastructure support:

'How do we create and produce tools to seamlessly connect related research projects in complimentary, organic ways that compel collaborations in both the physical and virtual worlds where our collaborative projects are developed and executed?'

Several fellows in each cohort went on to take leadership positions within RDA. They organized plenary sessions, chaired working and interest groups, and even participated in RDA governance. Despite often going back to their respective professional communities and sometimes prioritizing those communications, fellows were eager to explore connections within RDA and build diverse collaborations. Even though some of them were rather hesitant to initiate conversations as they felt like 'fish out of water' among the more experienced or more connected RDA members, many appreciated opportunities to learn about others' backgrounds, share ideas, and plan future work together. Through their interactions with professionals from many fields, they also expanded their view of possible careers in data and became ready to seek and craft diverse career paths.

The level of initiative varied among the fellows. To a certain extent it depended on the amount and scope of ongoing activities within RDA, but also on the career stage of the fellows and how comfortable they were speaking up and proposing ideas in front of a group. Thus, fellows who became leaders within interest and working groups or led specific initiatives had their research topics defined and closely connected to RDA group activities. This alignment was rather hard to predict at the application stage of the program because RDA groups are dynamic, and their work changes based on group members availability and contributions. The fellows were able to either make a connection between their work and the group or carve out a piece and make it fit with the group work. As one fellow put it, emphasizing one's own initiative: '... there's a lot of people who are very willing to help but especially help those people who are willing to help themselves...'

While a few fellows were driven by the need to enable changes in academia and the practices of research and data production from the beginning, for many this became a prominent and a more nuanced issue through their activities within RDA. At the beginning of their fellowship, the ECs were confident in making promises 'to improve the global flow of research data,' 'to influence a multitude of disciplines,' or 'to make widespread changes to organizational and individual practice.' As they learned about various challenges in the data landscape, ECs recognized that much work has to be done in small increments and called for more consultations and collaborations in attempts to enable change, again using a collective 'we' and establishing themselves as part of the solution:

'... if we can kind of provide more best practices and governing documentation and work with stakeholders from funding agencies through publishers to faculty members and students I think that a lot of these issues we're facing as individual labs can really be overcome by kind of having more of a community driven approach.'

One aspect that was lacking in the transition of ECs from learning to mastery and leadership was their own open data sharing. While the fellows were not explicitly asked about their data sharing practices, they were asked to report all outcomes and achievements at the end of their program. Program coordinators also collected citations of all publicly available products where the fellows were listed as authors. Several ECs mentioned creating datasets among

their achievements, others reported analyzing data, writing scripts, publishing the results, or learning new skills. Only one fellow said that they published their dataset and another one was ‘still considering how to publish my data’. This lack of references to publishing datasets in reports and public citations was interpreted as no commitment to open data sharing.

In their applications, interviews, and final reports fellows talked at length about how they would enable others to share data, code, and research findings, but almost none of them provided examples of their own sharing of data or mentioned themselves as examples of commitments to open science, even though several of them shared their publications openly online. This difference between values and actual behavior in science has already been documented by others (Anderson et al, 2007).

## DISCUSSION

The findings from our analysis of the RDA/US Data Share Fellowship program demonstrate that the ‘hidden’ curriculum of early career programs helps to foster interdisciplinarity and encourage an active position toward change in data cultures through the conversations around data. Such curriculum is especially helpful when it has an explicit emphasis on openness, diversity, and learning. At the beginning, the main objectives of the program were rather broad – to improve career options of the fellows. Over time, as program coordinators adjusted the program based on the feedback from the fellows, program evaluators, and the advisory board, interdisciplinarity emerged as a strong factor in how the program was perceived and practiced. Recognizing the difficulties of developing a formal curriculum for such a diverse group of ECs, the program coordinators drew on the breadth of expertise of RDA members and focused on creating the networking and informal learning opportunities.

Even though all ECs gravitated toward their own professions, mandatory gatherings of all fellows and broader peer-to-peer and RDA member interactions allowed the fellows to learn about the expanding system of data objects that is moving toward more computation and integration of various forms of evidence in support of one’s knowledge claims (Bechhofer et al, 2010; Chen, 2018; Kahn and Wilensky, 2006). In our program we created an environment conducive for cross-disciplinary interactions through the following efforts:

- Developing the review and admittance criteria that promoted cohort diversity
- Leveraging RDA structure and programming to expose ECs to multiple data discussions and problems
- Creating support for joint mentoring, supervision, and support for peer-to-peer interactions

While the short-term early career programs cannot address the systemic factors that impact openness and interdisciplinarity, including institutional support, disciplinary specialization, or funding structures (Bruce et al, 2004; McGraw and Biesecker, 2014), they can introduce mechanisms that support diversity, learning, and leadership and, hopefully, lead to a culture change.

Fostering interdisciplinarity through data acculturation, i.e., through exposing ECs to other data cultures and guiding them to learn and appreciate other forms of evidence without forcing them to assume certain values or actions is a fruitful avenue for action and research. While reshaping academic training from disciplinary programs into interdisciplinary clusters may be a long-term solution (Golde and Gallagher, 1999), examining and creating conditions that favor acculturation and epistemic fusion within the short-term fellowships and programs is a necessary intermediary step.

To promote such an intermediary step and further stimulate the discussion on how to develop and evaluate fellowship programs that foster interdisciplinary and open data cultures, we combine our findings with the framework for interdisciplinary research programs developed by Carr et al. (2018). This framework, which emphasizes learning and sustained interaction as a way to develop a shared understanding of research, aligns well with our findings of experiencing diversity, developing professional ties, and learning about data.

Our examination of the ECs interactions showed that the overall exposure to diverse research fields through data provided ample opportunities to learn about interdisciplinarity and build

the practices of openness and transparency, so we center our recommendations around that. [Table 5](#) below recommends several programming approaches that promote interdisciplinarity and openness that are based on our findings. It also describes the outcomes that implementing those recommendations will bring. Each outcome can be measured quantitatively by collecting relevant statistics or qualitatively via interviews and observations.

RECOMMENDATION	POSSIBLE OUTCOMES
Reach out to <b>multiple disciplinary and cross-disciplinary areas</b> to ensure a diverse representation of applicants and participants	Increased opportunities for participation in the program along multiple criteria of diversity (demographic, educational, social, and so on) Changes in one's own attitudes and behavior
Use <b>mutual sharing among participants, guided discussions, and collaborative projects</b> to help ECs learn about data practices in other research fields and in other geographical regions	Respect for diverse methodologies and epistemological standpoints Improved communication across diverse groups Broader professional networks
Incorporate both <b>formal and informal training</b> that covers a broad range of competencies in professional conduct	Skills that are applicable to all stages of the research and data lifecycle (European Commission Working Group on Education and Skills, 2017) Awareness of multiple stakeholders that are involved in and contributed to data cultures Ability to continue learning and participate in diverse communities
Engage <b>speakers and mentors from both academic and professional fields</b> to bridge gaps between practice and research	Interpretations of openness and interdisciplinarity as concepts that are changing and can sometimes be contested, expanded or transformed (Barry et al, 2008; Fecher and Friesike, 2014) Problem-oriented mindsets Broader professional networks
Provide <b>interdisciplinary mentorship</b> and discuss alternative career paths	An expanded repertoire of possible careers Diversification of norms, expectations and decisions in academic careers (Laudel et al, 2019)
Incorporate themes and <b>activities that focus on action, leadership and change</b> , e.g., include an explicit open science pledge for ECs (Farnham et al, 2017)	Understanding of the gaps between norms and behaviors in science and research (Anderson et al, 2007; Bray and von Storch, 2017; Nosek et al, 2015) Changes in one's own attitudes and behavior

**Table 5** Recommendations and outcomes that promote interdisciplinary and open data cultures.

This study provides many insights into how to use early career training programs to foster openness and interdisciplinarity through integrative data practices. At the same time, it has several limitations. First, the follow-up activities and post-fellowship interactions were not a funded part of the program. Even though the coordinators interacted with the fellows, encouraged plenary attendance from the previous fellows and provided continuity by supporting the early career interest group within RDA, these efforts were rather informal and did not have a sustained focus. As such, there is a danger that the effects of the program are short-term. Long-term effects need to be nurtured and further studied.

Second, the findings may be difficult to generalize to other early career programs, especially the programs that exist within the specific professional associations. Even if the program and our recommendations are implemented in another cross-disciplinary or global organization, it is not guaranteed that the effects of mentorship, curriculum development and other activities can be replicated. Almost half of our participants came from the library and information science domain, which could introduce bias in data collection and interpretation. Nevertheless, as one of the first in-depth looks into the role global interdisciplinary organizations into the careers in academia and emerging data-related professions, our study is useful in generating and further testing the hypotheses and outcomes related to this area.

Third, the program mostly supported informal interactions and learning through practice rather than offering formal curriculum. As several recent reports point out, many skills that support open science and interdisciplinarity can be developed through formal curricular, including the skills around databases, publication processes, ethics, and others (European Commission Working Group on Education and Skills, 2017). Both interdisciplinary and disciplinary-oriented organizations, particularly in the European Union, offer relevant training (EOSC Secretariat, 2020). Our findings and recommendations can be expanded in the future by developing and studying early programs that combine both formal and informal training.

Finally, what has also been omitted from this study is the role of administrative support in organizing and sustaining a fellowship program. Staff personnel time was needed to coordinate and manage the fellows, that is, to process applications, stipends, and travel and to navigate the policies of multiple fellows' home institutions as well as the program host institution and even RDA. While fostering interdisciplinary data cultures through global organizations such as RDA can be invaluable, these kinds of EC programs require administrative overhead and can face uncertainty due to changing funding cycles and priorities. Both the overhead and associated uncertainty may impact the decision-making within the program as well as its quality.

## CONCLUSION

In this paper we reported the findings from the four years of the RDA/US Data Share program and its successes and challenges of fostering interdisciplinary data cultures. We began with the assumptions that epistemic diversity is beneficial to the production of scientific knowledge and research and that such diversity can be promoted by focusing on data practices and connecting those who practice diverse approaches to data. We posited that early career programs such as RDA/US Data Share contribute to the development of cultures of interdisciplinarity and can be successful in expanding the future researchers views on what it means to work with data, generate new knowledge, and be a successful researcher in the context of openness and interdisciplinarity.

The findings from our study support those assumptions and confirm the benefits of organizing an early career program within an organization that focuses on data, such as the Research Data Alliance. These benefits, which can be largely described as acculturation of ECs to the practices of interdisciplinarity and openness, accrue through diversity and leadership experiences and through the exposure to broad professional networks and the landscape of data practices.

Based on the analysis of the rich data that has been collected over the course of the program, we recommended a combination of formal and informal training, activities that focus on mutual sharing and collaboration, an interdisciplinary mentorship, and activities that support leadership and behavioral modifications. Future research in this area can evaluate the effectiveness of implementing these recommendations and examine the relationships between ECs values, attitudes, and behaviors with regard to openness and interdisciplinarity.

## ADDITIONAL FILES

The additional files for this article can be found as follows:

- **Supplementary Material 1.** Interview Protocols. DOI: <https://doi.org/10.5334/dsj-2021-002.s1>
- **Supplementary Material 2.** Survey Protocols – Survey Protocols. DOI: <https://doi.org/10.5334/dsj-2021-002.s2>

## ACKNOWLEDGEMENTS

This research was funded in part by a grant from the Alfred P. Sloan Foundation (#G-2014-13746) and a grant from the National Science Foundation (#1349002).

The authors would like to thank all the RDA Data Share Fellowship program participants for their enthusiastic contributions and commitment to the program.

The publication of this paper was supported by the RDA Europe 4.0 project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777388.

## COMPETING INTERESTS

The authors have no competing interests to declare.

Kouper and Plale equally contributed to funding acquisition, conceptualization, and methodology of the program and the study. Scheidt conducted the interviews and transcribed and coded the data. Kouper and Scheidt equally contributed to the analysis and interpretation of data. Kouper is responsible for final data analysis, data curation, visualizations, and the preparation of the manuscript. Plale is responsible for project administration. All three authors reviewed the manuscript, provided final approval, and agreed to be on the author list.

## AUTHOR AFFILIATIONS

**Inna Kouper**  [orcid.org/0000-0001-9801-2277](https://orcid.org/0000-0001-9801-2277)

Department of Informatics, School of Informatics, Computing, and Engineering, Indiana University Bloomington, USA

**Lois A. Scheidt**  [orcid.org/0000-0002-9824-5916](https://orcid.org/0000-0002-9824-5916)

School of Information Studies, University of Wisconsin – Milwaukee, USA

**Beth A. Plale**  [orcid.org/0000-0003-2164-8132](https://orcid.org/0000-0003-2164-8132)

Department of Intelligent Systems and Engineering, School of Informatics, Computing, and Engineering, Indiana University Bloomington, USA

## REFERENCES

- Anderson, MS, Martinson, BC and De Vries, R.** 2007. Normative Dissonance in Science: Results from a National Survey of U.S. Scientists. *Journal of Empirical Research on Human Research Ethics*, 2(4): 3–14. DOI: <https://doi.org/10.1525/jer.2007.2.4.3>
- Anderson, WP.** 2017. *Data management: A global coalition to sustain core data*. Nature Publishing Group. DOI: <https://doi.org/10.1038/543179a>
- Barry, A, Born, G and Weszkalnys, G.** 2008. Logics of interdisciplinarity. *Economy and Society*, 37(1): 20–49. DOI: <https://doi.org/10.1080/03085140701760841>
- Bechhofer, S, De Roue, D, Gamble, M, Goble, C and Buchan, I.** 2010. Research Objects: Towards Exchange and Reuse of Digital Knowledge. *Nature Precedings*, (ERIM Project Document erim1rep091103ab12). Available at <http://eprints.soton.ac.uk/268555/>. DOI: <https://doi.org/10.1038/npre.2010.4626.1>
- Berger, PL and Luckmann, T.** 1990. *The social construction of reality: A treatise in the sociology of knowledge*. New York: Anchor Books.
- Berman, F, Wilkinson, R and Wood, J.** 2014. Building Global Infrastructure for Data Sharing and Exchange through the Research Data Alliance. *D-Lib Magazine*, 20(1/2). DOI: <https://doi.org/10.1045/january2014-berman>
- Berry, JW.** 2003. Conceptual Approaches to Acculturation. In: Chun, KM, Balls Organista, P and Marin, G (eds.), *Acculturation: Advances in Theory, Measurement, and Applied Research*. Washington, DC: American Psychological Association. pp. 17–37. DOI: <https://doi.org/10.1037/10472-004>
- Bray, D and von Storch, H.** 2017. The Normative Orientations of Climate Scientists. *Science and Engineering Ethics*, 23(5): 1351–1367. DOI: <https://doi.org/10.1007/s11948-014-9605-1>
- Bruce, A, Lyall, C, Tait, J and Williams, R.** 2004. Interdisciplinary integration in Europe: the case of the Fifth Framework programme. *Futures*, 36(4): 457–470. DOI: <https://doi.org/10.1016/j.futures.2003.10.003>
- Burns, LJ, Clayton, CP, George, JN, Mitchell, BS and Gitlin, SD.** 2015. The Effect of an Intense Mentoring Program on Junior Investigators' Preparation for a Patient-Oriented Clinical Research Career. *Academic Medicine*, 90(8): 1061–1066. DOI: <https://doi.org/10.1097/ACM.0000000000000742>
- Campbell, EG, Clarridge, BR, Gokhale, M, Birenbaum, L, Hilgartner, S, Holtzman, NA and Blumenthal, D.** 2002. Data Withholding in Academic Genetics. *JAMA*, 287(4): 473. DOI: <https://doi.org/10.1001/jama.287.4.473>
- Cao, L.** 2018. Data science education. In: *Data Science Thinking*. Cham: Springer. pp. 329–348. DOI: [https://doi.org/10.1007/978-3-319-95092-1\\_11](https://doi.org/10.1007/978-3-319-95092-1_11)
- Carr, G, Loucks, DP and Blöschl, G.** 2018. Gaining insight into interdisciplinary research and education programmes: A framework for evaluation. *Research Policy*, 47(1): 35–48. DOI: <https://doi.org/10.1016/j.respol.2017.09.010>
- Charmaz, K.** 2014. *Constructing grounded theory*. Sage.
- Chen, S.** 2018. *Big data in computational social science and humanities*. Cham: Springer. DOI: <https://doi.org/10.1007/978-3-319-95465-3>
- DataONE.** 2019. *Internships – DataONE*. Available at <https://www.dataone.org/internships> [Last accessed 24 May 2019].

- Dinmohammadi, M, Peyrovi, H and Mehrdad, N.** 2013. Concept Analysis of Professional Socialization in Nursing. *Nursing Forum*, 48(1). Available at <https://onlinelibrary.wiley.com/doi/pdf/10.1111/nuf.12006>. DOI: <https://doi.org/10.1111/nuf.12006>
- EOSC Secretariat.** 2020. EOSC Skills & Training Working Group Outputs. Available at <https://www.eoscsecretariat.eu/working-groups/skills-training-working-group/eosc-skills-training-outputs> [Last accessed 21 August 2020].
- European Commission Working Group on Education and Skills.** 2017. Providing researchers with the skills and competencies they need to practise Open Science. Available at <http://dx.publications.europa.eu/10.2777/121253> [Last accessed 21 August 2020].
- Farnham, A, Kurz, C, Öztürk, MA, Solbiati, M, Myllyntaus, O, Meekes, J, Pham, TM, Paz, C, Langiewicz, M, Andrews, S, Kanninen, L, Agbemabiese, C, Guler, AT, Durieux, J, Jasim, S, Viessmann, O, Frattini, S, Yembergenova, D, Benito, CM, Porte, M, Grangeray-Vilmint, A, Curiel, RP, Rehnrona, C, Malas, T, Esposito, F and Hettne, K.** 2017. Early career researchers want Open Science. *Genome Biology*, 18(1): 221. DOI: <https://doi.org/10.1186/s13059-017-1351-7>
- Fecher, B and Friesike, S.** 2014. Open Science: One Term, Five Schools of Thought. In: Bartling, S and Friesike, S (eds.), *Opening Science*. Springer. pp. 17–47. DOI: [https://doi.org/10.1007/978-3-319-00026-8\\_2](https://doi.org/10.1007/978-3-319-00026-8_2)
- Ford Foundation.** 2019. Ford Foundation Fellowship Programs. *The National Academies of Sciences, Engineering, and Medicine*. Available at <http://sites.nationalacademies.org/pga/fordfellowships/index.htm>
- Golde, CM and Gallagher, HA.** 1999. *The challenges of conducting interdisciplinary research in traditional doctoral programs*. DOI: <https://doi.org/10.1007/s100219900076>
- Hammer, D.** 2006. Improving student professionalism during experiential learning. *American Journal of Pharmaceutical Education*, 70(3): 59. DOI: <https://doi.org/10.5688/aj700359>
- Handelsman, MM, Gottlieb, MC and Knapp, S.** 2005. Training ethical psychologists: An acculturation model. *Professional Psychology: Research and Practice*, 36(1): 59–65. DOI: <https://doi.org/10.1037/0735-7028.36.1.59>
- Hickey, KT, Hodges, EA, Thomas, TL, Coffman, MJ, Taylor-Piliae, RE, Johnson-Mallard, VM, Goodman, JH, Jones, RA, Kuntz, S, Galik, E, Gates, MG and Casida, JM.** 2014. Initial evaluation of the Robert Wood Johnson Foundation Nurse Faculty Scholars program. *Nursing Outlook*, 62(6): 394–401. DOI: <https://doi.org/10.1016/j.outlook.2014.06.004>
- Kahn, R and Wilensky, R.** 2006. A framework for distributed digital object services. *International Journal on Digital Libraries*, 6(2): 115–123. DOI: <https://doi.org/10.1007/s00799-005-0128-x>
- Keralis, SDC.** 2012. Data Curation Education: A Snapshot. Council on Library and Information Resources (CLIR). Available at [https://digitalcommons.bucknell.edu/fac\\_pubs](https://digitalcommons.bucknell.edu/fac_pubs)
- Knorr Cetina, K.** 2007. Culture in global knowledge societies: Knowledge cultures and epistemic cultures. *Interdisciplinary Science Reviews*, 32(4): 361–375. DOI: <https://doi.org/10.1179/030801807X163571>
- Knorr Cetina, K.** 1999. *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.
- Koch, A, Glover, KC, Zambri, B, Thomas, EK, Benito, X and Yang, JZ.** 2018. Open-data practices and challenges among early-career paleo-researchers. *Past Global Changes Magazine*, 26(2): 54–55. DOI: <https://doi.org/10.22498/pages.26.2.54>
- Kouper, I.** 2014. 4th RDA Plenary – A breakout session on engagement. Available at <https://inkouper.blogspot.com/2014/11/4th-rda-plenary-breakout-session-on.html>
- Laudel, G, Bielick, J and Gläser, J.** 2019. ‘Ultimately the question always is: “What do I have to do to do it right?”’ Scripts as explanatory factors of career decisions. *Human Relations*, 72(5): 932–961. DOI: <https://doi.org/10.1177/0018726718786550>
- Martin, PE and Umberger, BR.** 2003. Trends in Interdisciplinary and Integrative Graduate Training: An NSF IGERT Example. *Quest*, 55(1): 86–94. DOI: <https://doi.org/10.1080/00336297.2003.10491792>
- McGraw, DK and Biesecker, AG.** 2014. Tribes, boundaries, and intellectual silos. In: *ETHICS’14 Proceedings of the IEEE 2014 International Symposium on Ethics in Engineering, Science, and Technology*. Available at <https://dl.acm.org/citation.cfm?id=2960673>
- Mozilla.** 2019. *Fellowships*. Available at <https://foundation.mozilla.org/en/fellowships/>
- MSDSE.** 2019. Moore-Sloan Data Science Environments – About. Available at <http://msdse.org/about/> [Last accessed 24 May 2019].
- Nicholas, D, Watkinson, A, Boukacem-Zeghmouri, C, Rodríguez-Bravo, B, Xu, J, Abrizah, A, Świgoń, M, Clark, D and Herman, E.** 2019. So, are early career researchers the harbingers of change? *Learned Publishing*, 32(3): 237–247. DOI: <https://doi.org/10.1002/leap.1232>
- Nosek, BA, Alter, G, Banks, GC, Borsboom, D, Bowman, SD, Breckler, SJ, Buck, S, Chambers, CD, Chin, G, Christensen, G, Contestabile, M, Dafoe, A, Eich, E, Freese, J, Glennerster, R, Goroff, D, Green, DP, Hesse, B, Humphreys, M, Ishiyama, J, Karlan, D, Kraut, A, Lupia, A, Mabry, P, Madon, TA, Malhotra, N, Mayo-Wilson, E, McNutt, M, Miguel, E, Paluck, EL, Simonsohn, U, Soderberg, C, Spellman, BA,**

- Turitto, J, VandenBos, G, Vazire, S, Wagenmakers, EJ, Wilson, R and Yarkoni, T.** 2015. *Promoting an open research culture*. American Association for the Advancement of Science. DOI: <https://doi.org/10.1126/science.aab2374>
- Okahana, H and Zhou, E.** 2018. Graduate Enrollment and Degrees: 2007 to 2017. Available at [https://cgsnet.org/ckfinder/userfiles/files/CGS\\_GED17\\_Report.pdf](https://cgsnet.org/ckfinder/userfiles/files/CGS_GED17_Report.pdf)
- Plale, BA, Dickson, E, Kouper, I, Liyanage, S, Ma, Y, McDonald, RH, Walsh, JA and Withana, S.** 2019. Safe open science for restricted data. *Data and Information Management*, 3(1): 50–60. DOI: <https://doi.org/10.2478/dim-2019-0005>
- Porter, AL, Roessner, JD, Cohen, AS and Perreault, M.** 2006. Interdisciplinary research: Meaning, metrics and nurture. *Research Evaluation*, 15(3): 187–195. DOI: <https://doi.org/10.3152/147154406781775841>
- RDA.** 2016. About RDA. RDA. Available at <https://www.rd-alliance.org/about-rda> [Last accessed 18 November 2019].
- Ritchie, J and Lewis, J.** 2003. *Qualitative research practice: A guide for social science students and researchers*. SAGE.
- School of Data** n.d. Fellowship Programme | School of Data. Available at <https://schoolofdata.org/fellowship-programme/> [Last accessed 13 November 2019].
- Shtasel, D, Hobbs-Knutson, K, Tolpin, H, Weinstein, D and Gottlieb, GL.** 2015. Developing a Pipeline for the Community-Based Primary Care Workforce and Its Leadership. *Academic Medicine*, 90(9): 1272–1277. DOI: <https://doi.org/10.1097/ACM.0000000000000806>
- Tenopir, C, Allard, S, Douglass, K, Aydinoglu, AU, Wu, L, Read, E, Manoff, M and Frame, M.** 2011. Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6(6). DOI: <https://doi.org/10.1371/journal.pone.0021101>
- University of Birmingham Research and Commercial Services.** 2009. Fellowships. Available at <https://www.birmingham.ac.uk/Documents/college-mds/percat/GrantsandFellowships/fellowship-fact-sheet.pdf>
- Varvel, VE, Bammerlin, EJ and Palmer, CL.** 2012. Education for data professionals. In: *Proceedings of the 2012 IConference*. ACM Press. pp. 527–529. DOI: <https://doi.org/10.1145/2132176.2132275>

**TO CITE THIS ARTICLE:**

Kouper, I, Scheidt, LA and Plale, BA. 2021. Fostering Interdisciplinary Data Cultures through Early Career Development: The RDA/US Data Share Fellowship. *Data Science Journal*, 20: 2, pp. 1–19. DOI: <https://doi.org/10.5334/dsj-2021-002>

Submitted: 10 December 2019

Accepted: 28 September 2020

Published: 25 January 2021

**COPYRIGHT:**

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.