



A School *and* a Network: CODATA-RDA Data Science Summer Schools Alumni Survey

SPECIAL
COLLECTION: RDA

RESEARCH PAPER

LOUISE BEZUIDENHOUT

SARA DRUMMOND-CURTIS

BRIDGET WALKER

HUGH SHANAHAN

MARCELA ALFARO-CÓRDOBA

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

]u[ubiquity press

ABSTRACT

The CODATA-RDA Schools for Research Data Science (SRDS) is a network of schools originating in the RDA in 2016. In 2019 it was recognized as an RDA output. To date, over 400 students from 40 countries have been trained in 10 schools. The majority of these students were postgraduates from low/middle-income countries (LMICs). In contrast to many other data science training approaches, the SRDS schools are designed to be 2-week, disciplinarily-agnostic, residential events where students are introduced to a broad range of tools requisite for efficient and responsible data-centric research. This paper presents the results of a survey carried out on alumni from schools held between 2016 and 2019 (45% response). The results of the survey strongly support the SRDS's long-term goals of facilitating data science training/capacity building within LMICs, and to foster communities of early career researchers (ECRs) conducting responsible and open data science research. The survey results demonstrated that 90% of respondent alumni continued to conduct research and make use of the skills acquired at the SRDS. Modules on open and responsible research and research data management were rated as important for future research. 79% of respondents confirmed that they maintained contact with peers, and 31% had set up academic collaborations with peers and/or instructors. Many had gone on to present content from the schools in their home institutions. The survey results clearly demonstrate the impact of the SRDS, and the value of an expanding network of schools supported by the RDA and CODATA.

CORRESPONDING AUTHOR:

Louise Bezuidenhout

University of Oxford, GB

louise.bezuidenhout@insis.ox.ac.uk

KEYWORDS:

RDA; CODATA; data science; school; alumni; survey

TO CITE THIS ARTICLE:

Bezuidenhout, L, Drummond-Curtis, S, Walker, B, Shanahan, H and Alfaro-Córdoba, M. 2021. A School *and* a Network: CODATA-RDA Data Science Summer Schools Alumni Survey. *Data Science Journal*, 20: 10, pp. 1–12. DOI: <https://doi.org/10.5334/dsj-2021-010>

The CODATA-RDA Schools for Research Data Science (SRDS) is a network of schools that have run since 2016. The SRDS has the long-term goal of creating communities of early career researchers (ECRs) that possess the skills to make the most of the Data Revolution in modern research. In response to the lack of data science training and capacity within low/middle-income countries (LMICs), the SRDS has a strong commitment to upskilling ECRs from LMICs.

SRDS schools are designed to be 2-week, residential events where students are introduced to a broad range of tools requisite for efficient and responsible data-centric research. In contrast to many other data science training curricula (Song and Zhu, 2016; Demchenko, Comminiello and Reali, 2019), the SRDS curriculum is intentionally *broad and shallow* and covers both the technical aspects of Data Science and Responsible Research Practices. The intention is that students completing an SRDS school will have a good understanding of the evolving data science research landscape, sufficient expertise with selected tools to engage in their own data-centric research, and an awareness of the areas in which they require future training. A range of future training opportunities, researcher networks and online resources are also provided to facilitate this future upskilling.

The SRDS curriculum is differentiated from the more traditional machine learning/data science “bootcamp” (Feldon *et al.*, 2017) by offering more than purely technical content. While the students are introduced to a range of computing tools, they are also instructed in a range of non-computing areas including Open Science, Responsible Conduct of Research (RCR) and Research Data Management (RDM). The total content covered by the curriculum is:

- Open and Responsible Research (Bezuidenhout, Quick and Shanahan, 2020)
- the Carpentries introductory material (Wilson, 2006; Teal *et al.*, 2015) on the Unix command line, Git and R¹
- RDM
- Author Carpentry ((Caltech Library, 2017) – authorship in the 21st century)
- Data visualisation
- Machine Learning
- Information Security
- Computational Infrastructures (an introduction computing beyond a laptop or desktop computer)

As is demonstrated by [Figure 1](#), “open and responsible research” forms the central theme of the schools. It is the intention of the SRDS that alumni of the school are not only competent data scientists, but also responsible researchers who understand, respect and potentially contribute to the Open Science movement. The Open and Responsible Research curriculum is fully described by Bezuidenhout, Quick and Shanahan in their 2020 paper (Bezuidenhout, Quick and

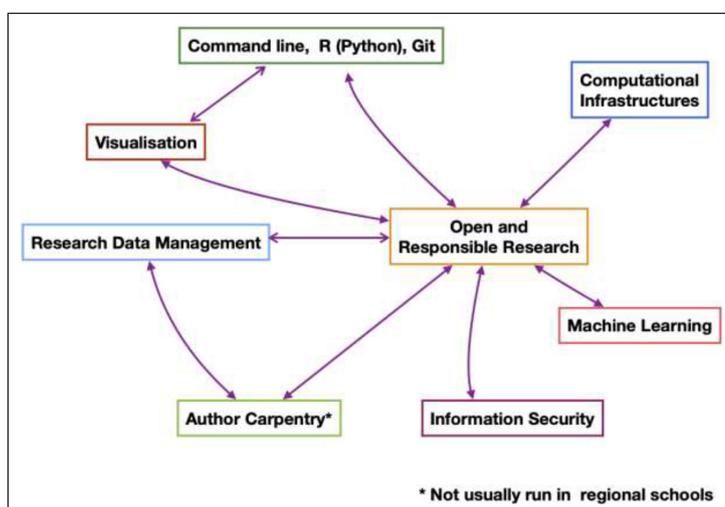


Figure 1 Diagrammatic representation of modules run in the ECR strand of the CODATA-RDA schools.

¹ In December 2019 the school in San José, Costa Rica used Python as the main language rather than R.

Shanahan, 2020). In brief, instruction includes formal lectures in Open Science, ethics and RCR as well as daily ethics reflection exercises that link ethical responsibilities to the daily research practices involving the computing tools being taught.

Since 2016 the SRDS has delivered 10 schools on 4 continents. These include 7 hosted by the International Centre for Theoretical Physics (Trieste, Italy: 2016, 2017, 2018, 2019; São Paulo, Brazil: 2017, 2018 and Kigali, Rwanda: 2019) and 3 hosted by national institutions (University of Addis Ababa, Ethiopia: 2019; University of Costa Rica, Costa Rica: 2019; University of Pretoria, South Africa: 2020). An abridged version of the school was also delivered in Brisbane, Australia in 2018. The organization and delivery of the schools is entirely done through volunteer activity and coordinated by a central committee of co-chairs. To date, the SRDS have trained around 400 students from over 40 countries. As the SRDS curriculum is disciplinarily agnostic, students have come from a wide variety of domains including Bioinformatics and other Life Sciences, Earth and Atmospheric Sciences, High Energy Physics and Social Sciences and Humanities.

Because the SRDS schools are residential schools, every effort is made to facilitate social networking between students (and instructors). There is a strong emphasis on teaching practical skills with team learning and ample opportunities for reflection and discussion. In addition, regular breaks, organized dinner and other social events facilitate these connections.

In 2020 the SRDS turned 5 years old. This milestone, together with the rapidly expanding number of alumni, suggested that an impact survey needs to be conducted amongst the alumni. In particular, the survey was intended to assess whether alumni were still using the skills learnt on the SRDS schools, and whether they were actively engaged in research. Additional questions about networking and sustained contact and collaboration were also included to assess the impact beyond data science research. This paper provides an overview of this survey, highlighting not only the impact of the SRDS curriculum, but also its emerging role as an international network of interdisciplinary ECRs.

METHODS

An online survey was disseminated using SurveyMonkey. A copy of the survey is available at [10.6084/m9.figshare.12033888](https://doi.org/10.6084/m9.figshare.12033888). The survey was distributed to alumni of the SRDS via personal emails (one email with two reminders). It was also disseminated via SRDS accounts on Twitter and Facebook. Responses were collected between mid-January and mid-February 2020.

In total, 180 responses were collected (104 from email invitations and 74 from the weblink). This represented approximately 45% of the total alumni population. All data was collected anonymously, although alumni were invited to share their email addresses if they wished to participate in follow-up interviews. Consent for the re-use of the aggregated data was given upon completion of the survey. This was clearly elucidated on the survey landing page.

The disaggregated survey data was only available to the authors of this paper and the SRDS co-chair committee. The aggregated data (aside from the submitted email addresses) is available at [10.6084/m9.figshare.12040104](https://doi.org/10.6084/m9.figshare.12040104).

RESULTS

DEMOGRAPHICS

43% of respondents were female. Respondents were from a wide range of ages, with the 78% falling between 22 to 35 at the time of the survey (17% were 22 – 25, 35% were 26 – 30, 26% were 31 – 35). Nonetheless, 2% of respondents were between 18 and 21, and 2% were over 50. Similarly, the majority of respondents were completing postgraduate degrees at the time of their school attendance, with 42% being Master's students and 31% PhD candidates. 84% of respondents were registered at an institution in their home country at the time of the summer school.²

² In the selection process, students with LMIC nationalities who were either registered in an institution in their home country or in another LMIC were given priority placements over those registered in HIC institutions. The rationale for this was that students registered in HIC institutions were more likely to be able to access similar training through their institutions or other in-country networks.

All major disciplines were represented in the cohort, the highest being computer science (20%) and mathematics (16%) and the lowest humanities (3%) and business sciences (1%). This is largely in keeping with the SRDS alumni demographic. Alumni of all 9 SRDS schools held between 2016 and 2019 were represented in the responses. The distribution of these responses is detailed in [Table 1](#) below.

SPECIFIC SRDS	PERCENTAGE OF RESPONSES
Trieste 2016	7%
Trieste 2017	16%
São Paulo 2017	8%
Trieste 2018	13%
São Paulo 2018	10%
Kigali 2018	5%
Trieste 2019	21%
Addis Ababa 2019	13%
San Jose 2019	8%

Table 1 Distribution of survey responses according to specific SRDS.

In total, 46 nationalities were represented amongst the respondents. Unsurprisingly, high numbers of responses were received from countries in which there had been recent schools, such as Ethiopia, Brazil, and Costa Rica. The distribution of countries is detailed in [Table 2](#) below.

NUMBER OF RESPONSES	COUNTRY OF CITIZENSHIP
26	Ethiopia
22	Brazil
18	Costa Rica
10	India, Nigeria
8	Morocco
6	Ghana, Iran
5	Kenya
4	Sudan
3	Colombia, Indonesia, Italy, Philippines, Rwanda, South Africa, Uruguay
2	Algeria, Canada, Cuba, Dominican Republic, Peru, Tanzania, Tunisia, Uganda, Venezuela
1	Argentina, Belize, Bolivia, Botswana, Egypt, France, Germany, Ireland, Mexico, Mozambique, Namibia, Nepal, Netherlands, Nicaragua, Pakistan, Slovenia, Spain, Zambia, Zimbabwe
6	Undisclosed

Table 2 Countries of citizenship represented amongst respondents.

As is demonstrated in [Figure 2](#), the majority of respondents assessed the level of their expertise in data science as “beginner” before they attended the SRDS. A further 29% felt that they were competent. Only 11% felt that they were proficient or expert.

66% of respondents said that they were actively collecting data at the time of their attendance at the school. 90% of respondents said that they had continued to conduct research since finishing the school.

CONTENT OF COURSES

In two interlinking questions, respondents were asked about the courses offered in the SRDS curriculum. [Figure 3a](#) illustrates the selections for the three courses respondents felt that they had learned the most from. [Figure 3b](#) illustrates the course felt by the respondents to be more useful than expected.

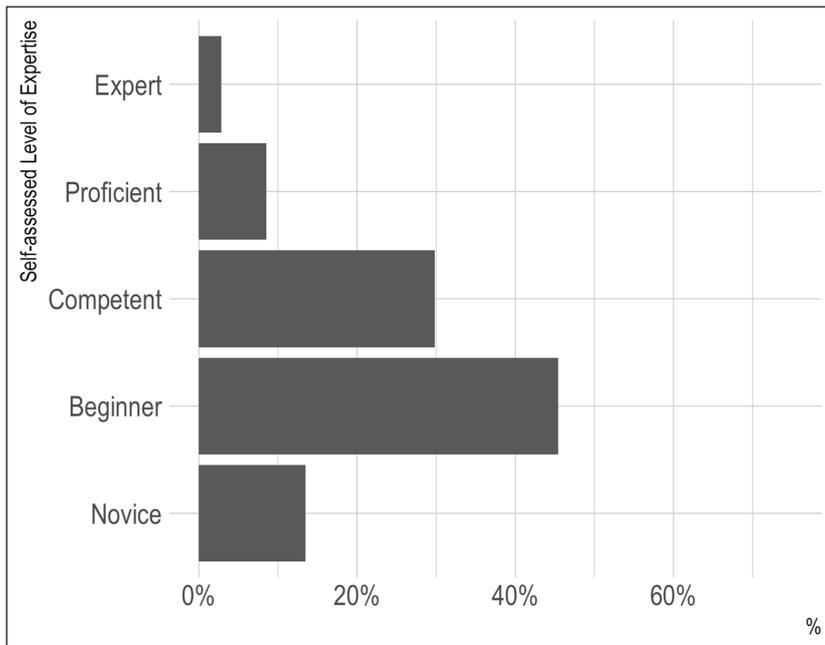


Figure 2 Respondents self-assessed level of expertise in data science before their attendance at SRDS schools.

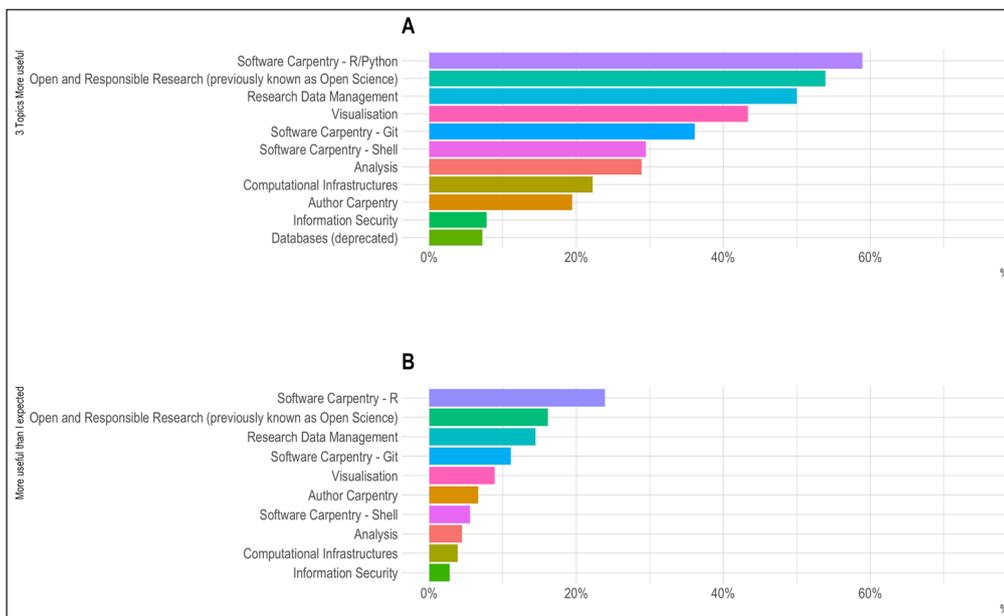


Figure 3 Demonstrating the responses to questions (a) Which three courses were the most useful? (b) Which course was more useful than expected?

The three courses felt to be most informative (*Figure 3a*) were R (60%) Open and Responsible Research (ORR, 55%) and Research Data Management (RDM, 51%). The same three courses were felt to be more useful than expected (*Figure 3b*).

Respondents were then asked about the frequency with which they continued to use the tools and skills learnt at the SRDS schools. As demonstrated in *Figure 4* above, usage varied considerably. Nonetheless, it was possible to say that respondents made use of all of the tools at least sometimes during their research. When asked why respondents had ceased to use certain tools taught at the SRDS, 74 respondents offered reasons. 39% of those answering the question said that the tool was not appropriate to their work, which is to be expected given the diversity of the SRDS curriculum. Nonetheless, 27% of those answering the question said that they needed more support to be able to use it, and 12% that they needed more instruction.³

CONTINUED ACCESS TO RESOURCES

The majority of alumni were based at academic institutions in LMICs. While the software used in the SRDS was free and open source software (FOSS), there was nonetheless the possibility that alumni might not be able to access the tools in their home institution due to complications

³ This is an important finding and one that the co-chairs committee is currently addressing through online training and webinars run through the new school alumni network.

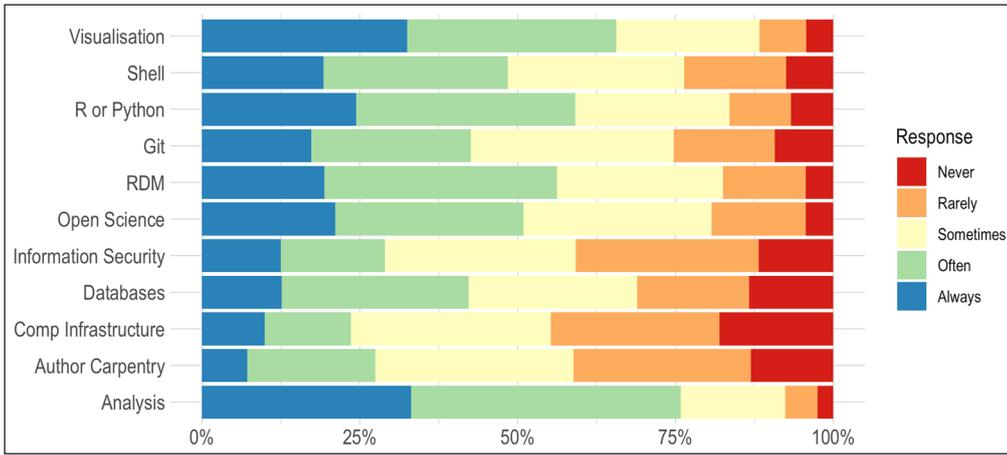


Figure 4 Continued use of tools learnt at SRDS schools.

including hardware availability. Nonetheless, 68% of respondents agreed that they were able to access all the tools used in the SRDS at their home institution.

Another key challenge for students returning home was the level of support that they received for their data science activities. As is evident from [Figure 5](#) below, the level of this support varied. While 68% of respondents felt supported by their peers, only 57% felt that they had the support of their supervisor. Moreover, less than half (45%) felt that they had the technical support that they needed for their data science activities.

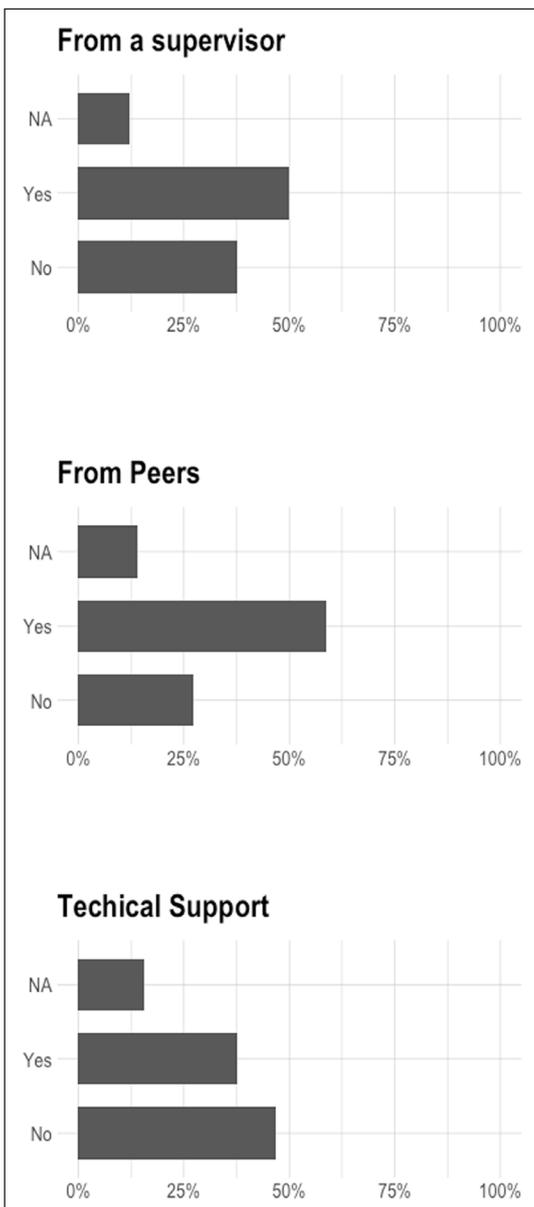


Figure 5 Respondents assessment of the support for data science activities in their home institution.

Nonetheless, it is well recognized in the literature that engaging in data science activities in LMIC research institutions can be hampered by a range of infrastructural and regulatory issues (Bezuidenhout et al., 2017). In order to gain a better understanding on how these issues impacted on the activities of alumni, respondents were asked to rate the influence of a number of different factors. As demonstrated in [Figure 6](#) below, the influence of these factors varied considerably. The most prevalent concern was difficulties relating to the purchase of hardware, and 40% felt that this was often/very often a concern. In addition, bandwidth (32%), power (24%), lack of technical support (31%) and lack of guidance about ethical/regulatory issues (30%) were all issues that respondents sometimes had to address.

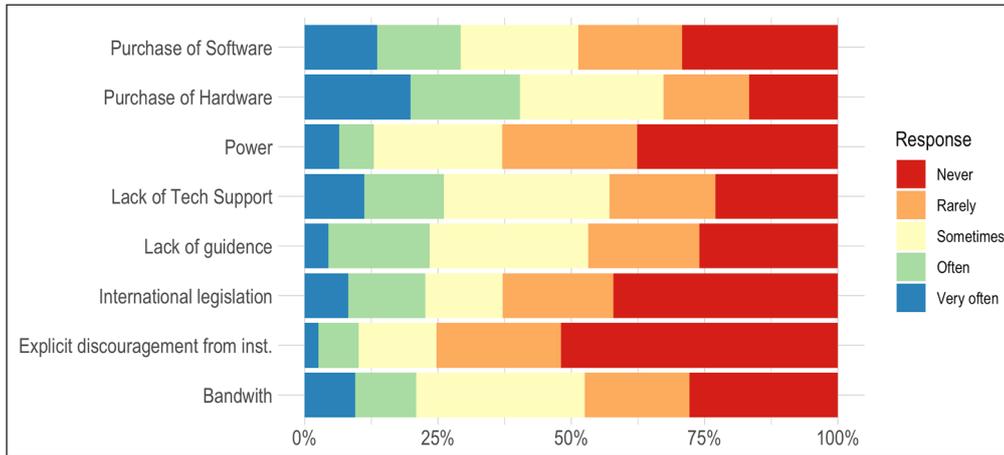


Figure 6 Challenges for using SRDS tools/skills in home institution.

SOCIAL NETWORKING

During the two weeks of each SRDS, the students had the opportunity to form social networks. A number of questions investigated whether these had led to persisting social contact. When asked whether they kept in contact with peers from the schools, 79% said that they kept in contact socially and 71% for academic-related activities. 28% kept in touch for other reasons such as sharing job opportunities and support for infrastructure ([Figure 7a](#)). This networking extended beyond solely social/information sharing, as 31% said that they had set up formal academic collaborations with school peers and/or staff ([Figure 7b](#)).

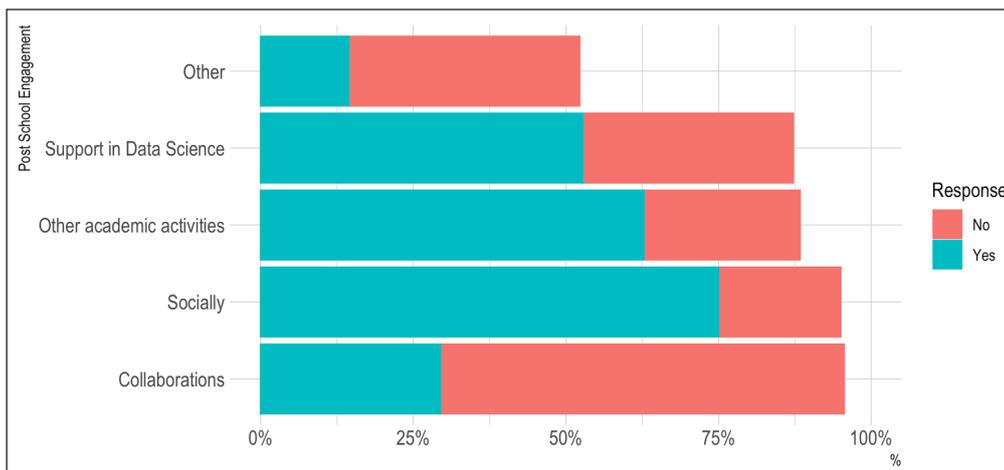


Figure 7 (a) Level of post-school social engagement **(b)** Level of post-school collaboration with peers and instructors.

OPEN SCIENCE AND DATA SCIENCE ADVOCATES

The introduction discussed the central ethos of the school, namely the commitment to Open Science practices. This content was introduced to students in the form of Open and Responsible Science Citizenship, an approach to teaching ethics that has been developed for the schools by one of the authors (LB). This course includes content on Open Science, RCR and bioethics. The immediate objective of this course was to educate students about Open Science and responsible research practices. A longer-term objective was that students would internalize this content and become Open Science advocates and RCR practitioners in their subsequent work.

In order to assess this long-term objective, respondents were asked whether they had continued to engage with Open Science practices after their return from the SRDS schools. As is displayed in [Figure 8](#) below, the respondents were engaged in a wide variety of Open Science practices. 63% advocated for Open Science in their institution and amongst their peers, while 30% had joined global communities (such as the RDA) that have a strong Open Science remit. Moreover, 74% of respondents said that they used FOSS, while 48% published in Open Access journals and 45% shared educational resources via open platforms.

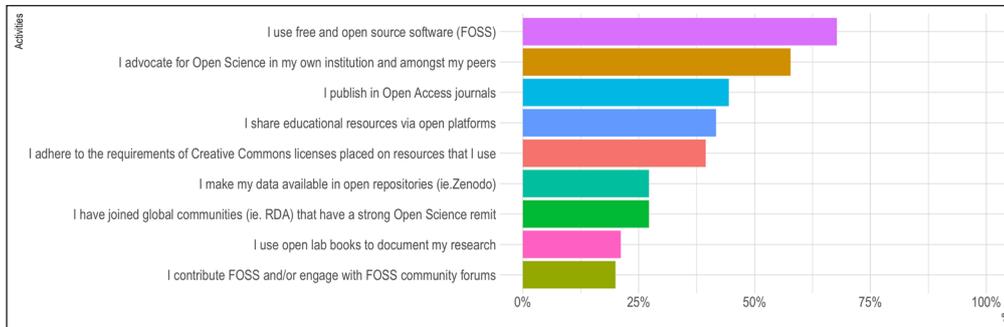


Figure 8 The range of Open Science activities in which respondents engaged after their attendance at the SRDS schools.

In addition to advocating for Open Science, alumni were actively engaged in disseminating their data science training. [Figure 9](#) below illustrates these activities. As can be seen, 76% of respondents said that they had informally shared skills from the schools with their peer community, while 46% had presented something at a group meeting. Moreover, 30% had made a formal presentation in their institution and 38% had organized a workshop or course.

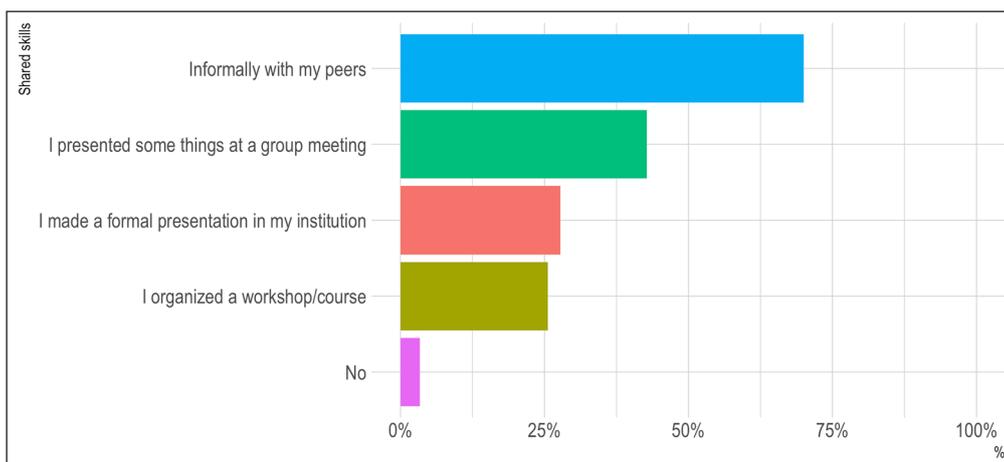


Figure 9 Level at which respondents shared skills learnt at the schools in their home institution.

DISCUSSION

THE IMPORTANCE OF MIXED CURRICULA

The SRDS curriculum was designed to be a “broad and shallow” introduction to data science. This differentiated it from a number of other approaches to data science training (such as The Carpentries), that provided focused instruction on specific tools. The “broad and shallow” approach was designed to specifically provide students with an overview of the tools that are needed to engage in efficient data-centric research.

As demonstrated in [Figure 4](#), the alumni continued to make use of the skills learnt at the SRDS in their research. This finding contrasts to recent studies, such as the one conducted by Feldon et al that suggest that short-term training has no impact for life science postgraduate students (Feldon et al., 2017). Indeed, the findings of the survey demonstrate that providing students with a broad overview of the field of data science - together with introductory skills and detailed advice on improving expertise - is a valuable approach to data science pedagogy.

The commitment to a *broad and shallow* curriculum provided the opportunity for two additional elements to be integrated, namely the integration of non-computing content and the use of

FOSS. As the curriculum provided a broad overview of data science practice, it was possible to include non-computing content relating to research practice and responsible conduct. The inclusion of these courses meant that students were not only instructed in the practice of data science, but became aware of the range of activities needed to become a *responsible* data scientist. These included introductions to Open Science, RCR, RDM and open authorship.

The Open Science and RCR content was grouped together under the concept Open and Responsible Science Citizenship. This concept (as described in Bezuidenhout 2020) broadly outlined the areas of responsibility required of an individual researcher, including responsible research conduct and civic responsibility. This concept foregrounded the reciprocity that characterises the Open Science movement, namely the importance of contributing resources/skills/expertise as well as benefiting from resources.

The exposure to ethics was not limited to formal lectures that introduced topics such as Open Science and RCR. The SRDS curriculum includes daily ethics reflection exercises that are directly linked to the computing content being taught (for a full description see Bezuidenhout 2020). This is a novel approach to ethics instruction, and is intended to assist students in making connections between high-level ethical values and daily research practice.

While students may have been initially surprised by the inclusion of non-computing content in the SRDS curriculum, the results of the survey illustrate that they saw benefit in this design. Indeed, as demonstrated in [Figure 3](#), Open Science and RDM ranked in the top three courses that students felt to be most useful. Moreover, as illustrated in [Figure 8](#), a high percentage of respondents went on to engage in Open Science practices within their own research institutions. Moreover, 63% of respondents said that they advocated for Open Science within their own research environments ([Figure 8](#)). These results strongly suggest that the inclusion of non-computing content within data science training has long-term positive implications for open and responsible research practices.

The commitment to Open Science was further reiterated by the use of FOSS throughout the course. As was demonstrated in [Figure 8](#), 74% of respondents continued to use FOSS in their research post-SRDS. They also shared their knowledge of FOSS with peers and more formally in their institutions ([Figure 9](#)). The combination of the sustained use of FOSS, together with the commitment to Open Science is a very positive reflection on the SRDS. 46 nationalities were reflected amongst the respondents, the majority of which are LMICs. These countries continue to be poorly represented within Open Science and FOSS communities. The possibility that SRDS alumni act as advocates for Open Science within these countries is a very positive contribution towards capacity building and representation on the global level. Similarly, Open Science discourse will benefit from the expertise and lived experiences of researchers in LMICs.

NETWORKS AND LONG-TERM ENGAGEMENT

The results of the survey demonstrate that attendance at an SRDS school provided students with the opportunity to form lasting social and academic networks. Nonetheless, organizing residential schools is both expensive and time-consuming when compared to other forms of delivery, such as online teaching. In a time when many research networks and training organizations are taking training content online, the SRDS faces regular questions about the efficacy of its model. Issues relating to the roll-out (global distribution, limitations of numbers), cost and global reproducibility (availability of instructors) are regularly raised as challenges of the 2 week residential school.

Nonetheless, the results of the survey clearly illustrate the benefits of residential schools. In all the schools the students have developed a strong community identity, and rapidly organise the social aspects of this community. This includes whatsapp and facebook groups, as well as peer-to-peer connections. The survey clearly demonstrates that this social networking persists long after the completion of the school, with 79% of respondents maintaining contact with their school peers ([Figure 7](#)).

In addition to the social support, 31% of the respondents confirmed that they formed active collaborations with peers or instructors ([Figure 7](#)). This illustrated two important benefits of the residential format. First, students and instructors had enough time to form connections and establish the trust relationships that lead to future collaborations. Second, the multi-disciplinary

nature of the student and instructor bodies meant that students were able to discuss their research with individuals that they might not normally interact with. The success of this model has already been described in a number of student-authored blogs.⁴

Due to the strong evidence supporting the residential model, it is unlikely that the SRDS will alter its format and migrate online. Nonetheless, there is room for further mixed-modality instruction. A free text question about further activities highlighted respondents desire for community-building activities online. These included refresher courses, train-the-trainer instruction, community resource sharing and a curated alumni network. Such activities are necessary - particularly as the SRDS network expands - to ensure that the community identity established during the schools is not dissipated, and that individuals engage beyond their particular school community with the broader SRDS network.

AMPLIFYING SUPPORT FOR ALUMNI AND OTHER LMIC DATA SCIENTISTS

The establishment of strong social networks was also important for SRDS alumni for reasons relating to their home research institution. Many LMIC institutions struggle with issues relating to research resources and infrastructures. Moreover, in many LMIC institutions research capacity is low (Fosci *et al.*, 2019). The combination of these issues can mean that SRDS alumni find it challenging to implement their newly-acquired skills within their own research.

As demonstrated by [Figures 5](#) and [6](#), key challenges experienced by respondents were a lack of access to hardware (40%) and technical support (55%). As the SRDS curriculum is not intended to make alumni into data science experts, this can mean that students returning home can encounter implementation problems that are individually unsolvable. As a result, the social networks formed during the schools, as well as their engagement with FOSS (and other Open Science) communities can mean the difference between persistent use or not. As the majority of respondents continued to make use of the tools taught in the SRDS ([Figure 4](#)), and 90% of respondents continued to engage in research after they returned home, the importance of this support cannot be under-estimated.

Nonetheless, reliance on social networks for technical support is not unproblematic. Indeed, previous studies on FOSS uptake (Vermeir *et al.*, 2018) have illustrated that the lack of dedicated technical support is perceived as a significant barrier to the use of FOSS. While the SRDS alumni may be able to effectively overcome these issues through their social networks and peer-to-peer contact, such options may not be available for their colleagues. This may hamper their efforts as FOSS/Open Science advocates and impair the spread of Open Science practices to LMIC regions.

Understanding the key role that physical resources and hands-on technical support play in the uptake of both data science practices and FOSS usage requires further consideration. In particular, Open Science communities need to engage with researchers in LMICs to understand how they are able to offset these challenges. Indeed, the provision of more online content or webinars cannot address issues to do with older hardware, lack of data storage options, absence of institutional technical support and ICT infrastructure.

CONCLUDING COMMENTS

The findings of this survey clearly demonstrate the importance of face-to-face interactions for early career researchers engaging in data science⁵. The findings also demonstrate how face-to-face instruction can be used as a means of fostering buy-in to Open and Responsible Science Citizenship. These observations provide an important counter to the prevalent trend of moving capacity building and training activities online. The value of forming long-term social connections and academic collaborations cannot be overlooked, and foregrounds the need for critical attention to be paid to the use of online courses and remote participation as the default means of engaging LMIC researchers (both in training and conferences).

⁴ For example, this blog describes a collaboration between a statistician student and a sociologist instructor <http://www.afox.ox.ac.uk/2019/12/17/breaking-barriers-in-academia-and-research/> (accessed 27/03/2020).

⁵ All students attending the SRDS have their travel and accommodation covered by a variety of grants and donations managed by the co-chair committee. This enables the schools to offer face-to-face training without a financial burden to the students.

While the SRDS alumni network is already flourishing, there is much that needs to be done if it is to be maximally effective. Although self-organizing networks are good, their impact continues to rely on the availability of volunteer effort. While this model has been productive for many networks - including the Research Data Alliance (RDA) - it is salient to recognize that maximal productivity is achieved with dedicated (and compensated) administrative support. Such support provides the overview, oversight and assistance to fully harness the enthusiasm and expertise within the volunteer population.

In conclusion, the survey presented in this paper offers an overview of the impact of the SRDS - not only as a means of building data science capacity, but also a burgeoning network of interdisciplinary early career researchers in LMICs. The paper demonstrates the power of RDA activities that start as interest groups, illustrating how community-initiated and -led activities have the power to become an expanding network of over 400 alumni.

ADDITIONAL FILE

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

- The data from the survey can be found here: https://gitlab.com/malfaro2/DSJ_repo

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Louise Bezuidenhout  orcid.org/0000-0003-4328-3963

University of Oxford, GB

Sara Drummond-Curtis

CODATA-RDA Schools for Research Data Science, GB

Bridget Walker

Research Data Alliance, GB

Hugh Shanahan  orcid.org/0000-0003-1374-6015

Royal Holloway University of London, GB

Marcela Alfaro-Córdoba  orcid.org/0000-0002-7703-3578

University of Costa Rica, CR

REFERENCES

- Bezuidenhout, L**, et al. 2017. Beyond the Digital Divide: Towards a Situated Approach to Open Data. *Science and Public Policy*, 44(4): 464–475. DOI: <https://doi.org/10.1093/scipol/scw036>
- Bezuidenhout, L, Quick, R** and **Shanahan, H**. 2020 “Ethics When You Least Expect It”: A Modular Approach to Short Course Data Ethics Instruction. *Science and Engineering Ethics*. Springer, 1–25. DOI: <https://doi.org/10.1007/s11948-020-00197-2>
- Caltech Library**. 2017. *AuthorCarpentry homepage*. DOI: <https://doi.org/10.7907/z96h4ffz>
- Demchenko, Y, Commiello, L** and **Reali, G**. 2019. Designing Customisable Data Science Curriculum Using Ontology for Data Science Competences and Body of Knowledge. DOI: <https://doi.org/10.1145/3322134.3322143>
- Feldon, DF**, et al. 2017. Null effects of boot camps and short-format training for PhD students in life sciences. *Proceedings of the National Academy of Sciences of the United States of America*, 114(37): 9854–9858. DOI: <https://doi.org/10.1073/pnas.1705783114>
- Fosci, M**, et al. 2019. *Research Capacity Strengthening in LMICs: Rapid Evidence Assessment Prepared for DFID*. London. Available at: www.research-consulting.com (Accessed: 30 March 2020).
- Song, I-Y** and **Zhu, Y**. 2016. Big data and data science: what should we teach? *Expert Systems*. Blackwell Publishing Ltd, 33(4): 364–373. DOI: <https://doi.org/10.1111/exsy.12130>
- Teal, TK**, et al. 2015. Data Carpentry: Workshops to Increase Data Literacy for Researchers. *International Journal of Digital Curation*, 10(1): 135–143. DOI: <https://doi.org/10.2218/ijdc.v10i1.351>
- Vermeir, K**, et al. 2018. *Global Access to Research Software: The Forgotten Pillar of Open Science Implementation*. Available at: www.globalyoungacademy.net (Accessed: 28 September 2018).

TO CITE THIS ARTICLE:

Bezuidenhout, L, Drummond-Curtis, S, Walker, B, Shanahan, H and Alfaro-Córdoba, M. 2021. A School and a Network: CODATA-RDA Data Science Summer Schools Alumni Survey. *Data Science Journal*, 20: 10, pp. 1–12. DOI: <https://doi.org/10.5334/dsj-2021-010>

Submitted: 30 March 2020

Accepted: 15 January 2021

Published: 18 February 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.