PRACTICE PAPER

# Establishing, Developing, and Sustaining a Community of Data Champions

## James L. Savage[1] and Lauren Cadwallader[2]

[1] Department of Zoology, University of Cambridge, Cambridge, UK

[2] Office of Scholarly Communication, Cambridge University Library, University of Cambridge, Cambridge, UK

Corresponding author: James L. Savage (james.savage@cantab.net)

Supporting good practice in Research Data Management (RDM) is challenging for higher education institutions, in part because of the diversity of research practices and data types across disciplines. While centralised research data support units now exist in many universities, these typically possess neither the discipline-specific expertise nor the resources to offer appropriate targeted training and support within every academic unit. One solution to this problem is to identify suitable individuals with discipline-specific expertise that are already embedded within each unit, and empower these individuals to advocate for good RDM and to deliver support locally. This article focuses on an ongoing example of this approach: the Data Champion Programme at the University of Cambridge, UK. We describe how the Data Champion programme was established; the programme's reach, impact, strengths and weaknesses after two years of operation; and our anticipated challenges and planned strategies for maintaining the programme over the medium- and long-term.

## Background

Engaging researchers in effective Research Data Management (RDM) has proven a challenge, despite established incentives of better visibility, reproducibility, impact, and efficiency of research (Markowetz, 2015; Ingram, 2016), and despite growing pressure from institutions and funding bodies (Higman, Teperek and Kingsley, 2017; UKRI, 2018). Difficulties arise because researchers are either unaware of the benefits and evolving landscape of open science, do not possess the skills and resources to manage their own data effectively, or do not see its value because it is not appropriately rewarded by the current academic system. To bridge this knowledge gap, many institutions have set up specialised central units to advocate for and support RDM in their research communities. However, while these units possess a great deal of expertise, they often face substantial challenges targeting and engaging with researchers needing support because of the decentralised and diverse nature of the research community in their institution (Awre et al., 2015).

The University of Cambridge contains six academic Schools that collectively administer around one hundred departments and institutes, and employs around 11,000 staff of which 5,700 are academic or contract research staff that could benefit from RDM support. In addition, the University supports around 12,000 undergraduate and 7,000 postgraduate students that engage in varying levels of research activity and generally have little experience with RDM. Within the University, RDM support is primarily supplied by the Office of Scholarly Communication (OSC), established in January 2015. The OSC provides a central website (https://www.data.cam.ac.uk/) with information about good RDM practices, offers RDM training through short courses and consultancy, and provides support for researchers depositing data and scholarly output to the institutional repository.

From its inception the activities of the OSC have been unable to meet the demand, with frequently oversubscribed sessions. Despite this, some evidence shows that researchers were unaware of the support available for RDM (Higman, Teperek and Kingsley, 2017). Given the lack of additional central funding for support staff, an alternative community-based model for providing support was developed in the form of the Data

Champions Programme (Higman, Teperek and Kingsley, 2017; Teperek, Higman and Kingsley, 2018). Led by the Research Data Coordinator and overseen by the Research Data Facility Manager, this initiative was supported by the senior staff at both the University Library and the Research Office, who at the time were jointly responsible for the work of the OSC and their RDM team.

## The Data Champions Programme

The primary objective of the Data Champion Programme, as detailed in the initial paper on the Programme by Higman, Teperek and Kingsley (2017), was to create a community of RDM advocates and trainers with strong links to both the central support infrastructure and their local communities of research practice (**Figure 1**). Data Champions (DCs) are volunteers recruited from across the University, who advise members of their own research communities on proper handling of research data, promote good RDM, and support FAIR research principles: Findable, Accessible, Interoperable, and Re-usable data (Wilkinson et al., 2016).

### *Establishing: Calls for participation*

The first call for volunteers to join the Programme was sent in September 2016 and appealed for 'anyone interested in research data management and sharing' to become a local expert by joining the community. This call included detailed expectations for the volunteers: to organise and run at least one workshop per year, act as a local expert, forward questions to the RDM Team, advocate for good RDM, and act as an RDM representative by attending the bimonthly forums. The potential benefits of becoming a DC were also listed (growing networks, learning new skills, increasing impact, and boosting your CV), along with the support provided by the central RDM Team (help with workshop preparation, 'train the trainer' sessions, feedback mechanisms to improve support). The call was advertised via mailing lists and social media as well as directly to specific individuals, such as those who had specific expertise that would benefit others if shared, and a total of 43 eligible applications were received, all of whom were accepted (Higman, Teperek and Kingsley, 2017).
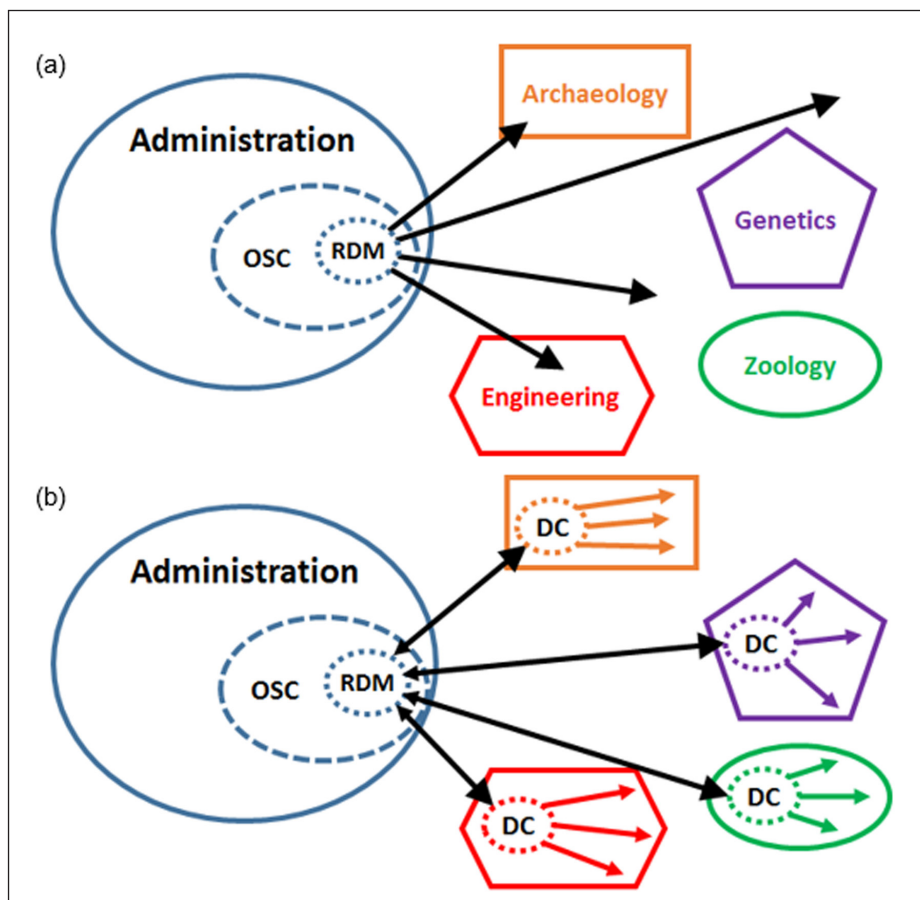


**Figure 1:** Conceptual diagram of Research Data Management (RDM) unit support to individual departments before **(a)** and after **(b)** implementation of the Data Champion (DC) Programme, illustrating the increased reach of RDM expertise within the University. The RDM unit is based within the Office of Scholarly Communication (OSC). The departments named are purely for illustration.

The second call was launched in February 2018 with one significant difference: the expectations were amended such that the volunteers were not required to deliver any particular type of support (OSC 2018). It had been recognised that the first cohort of DCs were delivering support and advocacy in a number of ways, but not necessarily delivering an RDM workshop as previously specified (see **Table 1** for examples). Given the volunteer nature of the Programme it was felt that DCs should have the option of delivering support and advocacy using whatever approaches they were most comfortable with and considered most valuable, rather than risk missing out on potential DCs that either felt unable to deliver a workshop or judged that activity a poor fit for their research community. The 2018 call was updated to reflect this, and included examples of the diverse ways DCs might support RDM in their communities. This second call resulted in 20 eligible applications, all of whom were accepted; the only applications deemed ineligible came from individuals not affiliated with the University of Cambridge.

A third, near-identical call was launched in January 2019. The most significant update was a line welcoming applications from 'those working in all subject areas under the broad auspices of the arts, humanities, social sciences and STEM disciplines.' This aimed to increase participation in the programme by researchers in the arts, humanities and social sciences, who make up a third of the University but only 20% of the Data Champions cohort, with no representation from the School of Arts and Humanities. In contrast DCs from the STEM disciplines make up 76% of the DCs but around 66% of the University. In the long term, the OSC plans for calls to occur annually and coincide with the start of the academic year (October). This will maximise participation time for those who are only in Cambridge for short periods, and allow DCs to plan their training and advocacy work in relation to other events occurring within their academic communities.

As of January 2019, prior to the third call, the DC Programme comprises 46 champions across five of the six Schools of the University, covering departments from Archaeology to Zoology (**Figure 2**). The remaining

**Table 1:** Documented Data Champion activities as of January 2019.

| Workshops & Presentations | Other Activities |
| --- | --- |
| File management | Open Data FAQ |
| Data sharing | Bite-size emails |
| Avoiding data disasters | Training needs analysis |
| Writing data management plans | 1:1 RDM and data analysis support |
| Repositories | Data Policy |
| Code management | Data Audit |
| GitHub introduction | Outreach materials |
| 'Bring your own data' workshop | Electronic Lab Notebook trial |
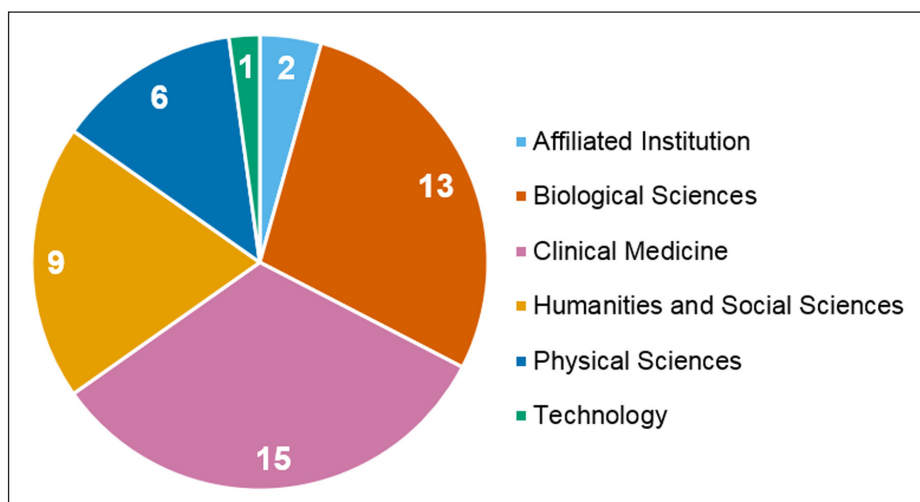


**Figure 2:** Data Champion primary School affiliations within the University of Cambridge (46 active champions in Jan 2019). No current DCs are affiliated with the School of Arts and Humanities. 'Affiliated Institution' incorporates areas of the University which do not come under the scope of the six academic Schools of the University.

17 individuals have withdrawn from the Programme, citing reasons such as leaving Cambridge, job changes, or no longer being able to commit sufficient time. The majority of DCs are postdoctoral researchers and PhD students, but academic and support staff are also well represented in the programme (**Figure 3**).

## Developing: Support for Data Champions

DCs are supported by the OSC in a variety of ways. Those that joined the Programme after the second call received a comprehensive induction, including an explanation of the programme, first-hand accounts from an existing DC, icebreaking activities to kick-start networking, and a detailed Welcome Pack containing information about the programme and available resources. When a new DC joins the programme the OSC also sends a letter to their Head of Department, explaining what the Programme is, what the DC role entails, and the associated benefits to the Department (Higman, Terepek and Kingsley, 2017). This aims to empower DCs to volunteer their time to the Programme, and ensure that departments are aware of the DC as a new source of local support for RDM.

Interaction between DCs is facilitated by regular two-hour forums, held bimonthly over lunch. These events give the DCs a chance to learn something new, and provide time for networking to help drive the Programme forward. The OSC organises forums with input from the DCs, and a typical forum might include an invited speaker, either internal, such as a representative of central IT services discussing high performance computing, or external, such as a publisher running a workshop around peer review of data. DCs also have opportunities to present during forums, both in the form of longer presentations on a topic of interest to the whole group, or shorter lightning talks to highlight ongoing or completed work. These talks raise awareness, kick-start networking and collaboration opportunities, and facilitate the development of presentation skills within a relatively safe environment.

Online resources for the DCs include a dedicated email list and Slack channel. DCs use these tools to share expertise and collaborate on activities, and the OSC asks for advice on matters that require specific expertise (e.g. dealing with imaging data). All DCs also have access to a shared University Google Drive that hosts a panoply of resources, including:

· Useful resources for 'being a Data Champion': welcome pack, advocacy materials, 'how-to' materials etc.
· Materials from forum meetings: presentations, minutes, collaborative documents etc.
· Resources for training, including presentations delivered by the OSC and other DCs.

DCs are encouraged to share materials they generate in the course of their local activities, and to reuse those of others (respecting licences). The ultimate aim is to build a comprehensive resource and knowledge base to underpin future DC activities and make it easier for new DCs to deliver support.
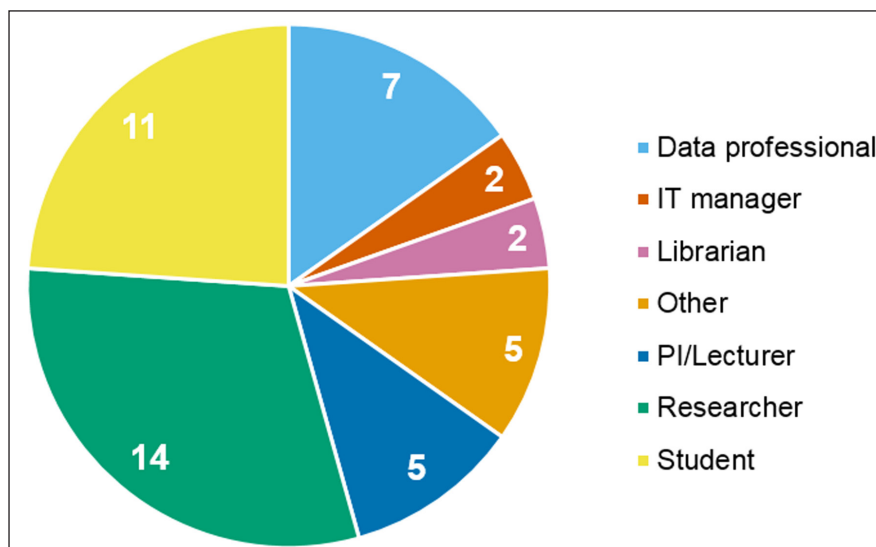


**Figure 3:** Data Champion roles (46 active champions in Jan 2019). Roles were assigned based on job titles. 'Student' includes all postgraduate students (Masters and PhD). 'Data professionals' are those who deal with research data directly in their day-to-day work but are not researchers. 'Other' includes roles such as administrators.

Feedback on institutional support has been positive. A short survey of DCs was carried out in September 2018, which aimed to get an overview of DC activities over the previous 12 months. Of the 20 responses received, 75% indicated that they had definitely or probably received enough support from the OSC. When asked what support they had received from their institution (i.e. Department/Faculty), 60% said they had at least a moderate amount of support, with only 15% reporting no support at all. Given the relative youthfulness of the Programme this is encouraging, however, it does indicate that more advocacy work can be done by the OSC to embed DCs fully into the University ecosystem. Going forward, the OSC must continue advocating for good RDM and demonstrate the value of the DC Programme, so that Departments feel able to fully commit to supporting their local DCs by facilitating activities, encouraging others to engage with the DC and join the Programme, and recognise the DC's contribution to enhancing the responsible research practices within that Department.

## Developing: Activities, costs, and impact

To date, DCs have organised and delivered a variety of support in their research communities, including formal RDM workshops and presentations on various topics, and diverse other contributions to support their local research community (**Table 1**). In addition to this support, DCs collect information about current practice and training needs to feed back to the central OSC, and provide specific expertise to OSC staff when required. Individual DCs provide local support in whatever form they feel comfortable with, reflecting the diversity of practice across disciplines.

In the September 2018 survey, when asked about the activities they had carried out in the last 12 months, 35% spent at least 1 day per month on DC-related activities and 60% between a few hours and half a day per month. When taken across the whole year, this can add up to around a weeks' worth of time (or more). When asked if they thought their activities had delivered impact, 65% said probably or definitely yes, 20% were not sure and 10% thought not (1 person did not respond). The main explanations given for the response on impact were that awareness of good RDM practices had noticeably increased after activities. However, it is as yet unclear how far this perceived increase in awareness has translated into improvement of practices.

The DC Programme is relatively cheap. Its staff costs are approximately 0.6 FTE split across two roles: the Research Data Coordinator (~0.3 FTE) who is responsible for the Programme, and administrative support from another member of the OSC (~0.3FTE). Current staff believe that more time could be usefully spent on running the Programme, but that it functions well at the current level of support.

The largest non-staff cost is catering for the bimonthly forums. Venues are provided by individual DCs who can access them for free in their Departments, and otherwise the Programme only uses generic University resources already available to staff. Other minor costs have included branded advocacy materials (flyers etc.), and providing conference support for one DC to attend the 2018 SciDataCon conference. However, despite the relatively low ongoing cost, longer-term funding for the programme is being sought to ensure that it can continue in its current form and cope with future increases in DCs.

## Sustaining: Challenges

All programmes built on networks of volunteers with central support face a number of challenges that must be proactively mitigated for the programme to enjoy sustained success (Studer and von Schnurbein, 2013). For the DC Programme, the main risks are failing to attract new volunteers, lack of productive activity by existing DCs, and loss of central support from the University (Higman, Teperek and Kingsley, 2017). Both a lack of volunteers and a lack of activity by DCs could occur either because individuals lack the ability to provide support within their departments, or because there are insufficient incentives to devote time to the programme. The OSC, together with the DCs, are exploring several initiatives, as explained below, as options to improve both DC numbers and the efficiency of the network at inducting and training new DCs in supporting their research communities.

## Sustaining: Finding & integrating new Champions

Many DCs are PhD students or short-term research staff, and thus high turnover inevitably occurs as DCs complete their studies or contracts. Consequently, the aim should be for DCs to join the Programme as early as possible, and be rapidly inducted and trained to maximise their potential contribution. A particularly vulnerable period is immediately following induction, as initial enthusiasm can wane before DCs develop working relationships with others in the network. Beyond the induction sessions and welcome pack, the OSC provides events, tools and resources to develop and link DCs as efficiently as possible (see "Developing: Support for Data Champions" section).

In January 2019 the OSC instigated a call for mentors within the DC community to support the new cohort of champions. In addition to reinforcing and expanding on information from their induction on how the Programme works, mentors are intended to provide new DCs with a 'friendly face' at the bi-monthly forums and introductions to other DCs and University staff with relevant expertise. A related initiative is the incipient 'legacy' programme, which aims to mitigate the high turnover rate and ensure RDM support and links to the OSC within individual departments are maintained when DCs leave. DCs nearing the end of their contracts are asked to identify a possible replacement, and together with the OSC help to induct the new DC into the Programme and explain the work that has previously been done within that community.

An additional problem is the bias towards the STEM subjects in the demographic of the current DCs. The Programme is designed to work for all disciplines and it is vital for the OSC that RDM support reaches the many and varied disciplines in Cambridge. The language of the third call and related application form for DCs (January 2019) has been edited to make it more inclusive to those working in the Arts, Humanities and Social Sciences (AHSS), although the call remains advertised to all fields.

### Sustaining: Activity by existing Champions

Motivating DCs to provide support is a matter of minimising workload and rewarding engagement. DC activities occupy valuable time and most DCs are researchers, a time-poor group who are professionally judged primarily on their research output and not on voluntary activities even if these are closely related to research. Rewards must also operate in both directions: DCs need to gain skills, networks, or prestige sufficient to offset lost research time, and the University must likewise receive sufficient demonstrable benefit from DC activities. A key element here is visibility and creating public goods as outputs: the existence and contributions of champions must be publicised, and whenever possible the support and resources they provide must be documented, shared, assessed, and iterated.

Several initiatives are being explored to support this: the community Google Drive resources have been restructured to make it easier for DCs to find and reuse material, reducing the preparation time needed to deliver support. It is also hoped the mentoring initiative will increase collaborative activities and therefore reduce the workload of individual DCs without reducing engagement with those who need RDM support. Finally, the DC webpages (https://www.data.cam.ac.uk/intro-data-champions) are undergoing changes to increase the visibility of the DCs to those who might need their help but also to clearly give them a presence and credibility to their work.

An additional issue is keeping established DCs engaged with the bimonthly forums. After the second cohort of DCs joined the Programme, it became clear to the OSC that there was a need to balance the forum agendas to suit both new and established DCs who may have significant differences in their knowledge base about RDM. Acknowledging this issue with DCs has been an important first step, and careful planning of the 2019 forums aims to mitigate the problem. Thus far, interactive activities have been used to allow new DCs to learn from their colleagues while simultaneously giving established DCs space to discuss issues at the level they prefer. As the programme grows, we anticipate increasing use of formal sub-groups to allow DCs with shared interests to focus on specific issues both within and outside the forums. For example, we envisage groups of DCs who work with similar data types to collaboratively run training or support sessions outside DC forums, or to work together in forums to look at issues from their common perspective. This will allow a greater exchange of ideas that are relevant and timely to both the established and new DCs.

### Sustaining: Central support from the University

Fundamentally, continued support for the DC Programme from the institution relies on it delivering value for money. To do this, it must be both cost-efficient and able to demonstrate a positive impact on research communities. A key element here is documentation, with DCs encouraged to make available any materials they create in the shared Google Drive and to report back on successful activities, as it is very difficult to measure their impact over such a large institution. Given the size of an organisation like the University of Cambridge, it is also imperative that this Programme aligns with the wider strategies and aims of the institution.

Moving forward, we plan for the Data Champion Programme to build stronger links with other institutional organisations and initiatives. For example, in Cambridge a new Research Support Ambassador Programme aims to upskill library staff in RDM and related topics, creating further expertise in staff embedded within academic units (Sewell and Kingsley, 2017). It is envisaged that these librarians will work closely with DCs within their department, providing DCs with further expertise on available support within the institution, and librarians with a closer understanding of data issues encountered during research.

One unknown factor in the future sustainability of the programme is what the long-term size of the DC community will be. Each call brings in new members, and although some established DCs leave each year, it is unclear at what number of DCs the programme will stabilise, and hence whether the current level of staff resources provided by the OSC will be sufficient going forward. Consequently, future challenges may arise in ensuring sufficient central funding is committed to the programme.

## Conclusions

The DC Programme represents an experiment in community-driven RDM support, but one that is heavily facilitated by a central unit. This hybrid approach has proven effective in developing expertise in RDM across many departments. Despite the support from the OSC that has been required for the Programme to develop and thrive, costs have been minimised by taking advantage of existing institutional resources and – fittingly – encouraging open science principles within the Programme itself. The Programme has thus far been successful in boosting researcher engagement in RDM, but challenges remain in navigating the high turnover rate of Champions and in demonstrating impact and engagement across the institution.

Ultimately, the success or failure of any particular initiative supporting RDM may primarily reflect the existing context or community within the institution, rather than the strengths and weaknesses of the initiative itself. Consequently, we strongly encourage further discussion and reporting of structured efforts to provide RDM support across as wide a range of institutions as possible. A number of potentially successful frameworks exist, from a large, active central research data unit, to embedded full-time Data Stewards within each faculty. to programmes such as the Data Champions that utilise volunteer community networks (Bryant, Lavoie and Malpas, 2017; Teperek et al., 2018). It is currently unclear whether one of these approaches is particularly effective, or if different strategies make sense under different institutional contexts. This is a question that can only be answered by collating data and case studies from across the sector.

## Data Accessibility Statement

The data in support of this publication are freely available under a CC BY licence from the University of Cambridge repository, Apollo, at the following link: https://doi.org/10.17863/CAM.35725.

## Ethics and Consent

The data presented from the survey of the Data Champions has been fully anonymised. The Data Champions were sufficiently informed that the data would be used in the publication of this article.

## Acknowledgements

## Competing Interests

The authors have no competing interests to declare.

## Author Contribution

JS & LC conceived the article, wrote the manuscript, and approved final submission.

## Author Information

James Savage is a Data Champion and Research Associate. Lauren Cadwallader is Deputy Head of Scholarly Communication (Research Data Management).

## References

**Awre, C, Baxter, J, Clifford, B, Colclough, J, Cox, A, Dods, N, Drummond, P, Fox, Y, Gill, M, Gregory, K, Gurney, A, Harland, J, Khokhar, M, Lowe, D, O'Beirne, R, Proudfoot, R, Schwamm, H, Smith, A, Verbaan, E, Waller, L, Williamson, L, Wolf, M** and **Zawadzki, M.** 2015. 'Research Data Management as a "wicked problem"'. *Library Review. Emerald Group Publishing Limited*, 64(4/5): 356–371. DOI: https://doi.org/10.1108/LR-04-2015-0043

**Bryant, R, Lavoie, B** and **Malpas, C.** 2017. *Scoping the University RDM Service Bundle. The Realities of Research Data Management, Part 2*. Dublin, OH: OCLC Research. DOI: https://doi.org/10.25333/C3Z039

**Higman, R, Teperek, M** and **Kingsley, D.** 2017. 'Creating a Community of Data Champions'. *International Journal of Digital Curation*, 12(2): 96–106. DOI: https://doi.org/10.2218/ijdc.v12i2.562

**Ingram, C.** 2016. *How and why you should manage your research data: a guide for researchers, JISC*. Available at: https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data (Accessed: 27 January 2019).

**Markowetz, F.** 2015. 'Five selfish reasons to work reproducibly'. *Genome Biology*, 16: 274. DOI: https://doi.org/10.1186/s13059-015-0850-7

**Office of Scholarly Communication University of Cambridge.** 2018. *Call for Data Champions, 3 April 2018*. Available at: https://www.data.cam.ac.uk/sites/www.data.cam.ac.uk/files/mem_archive-call2018_v1_20180403.pdf (Accessed: 27 January 2019).

**Sewell, C** and **Kingsley, D.** 2017. 'Developing the 21st Century Academic Librarian: The Research Support Ambassador Programme'. *New Review of Academic Librarianship. Routledge*, 23(2–3): 148–158. DOI: https://doi.org/10.1080/13614533.2017.1323766

**Studer, S** and **von Schnurbein, G.** 2013. *Organizational Factors Affecting Volunteers: A Literature Review on Volunteer Coordination, Voluntas*. DOI: https://doi.org/10.1007/s11266-012-9268-y

**Teperek, M, Cruz, MJ, Verbakel, E, Böhmer, J** and **Dunning, A.** 2018. 'Data Stewardship addressing disciplinary data management needs'. *International Journal of Digital Curation*, 13(1): 141–149. DOI: https://doi.org/10.2218/ijdc.v13i1.604

**Teperek, M, Higman, R** and **Kingsley, D.** 2018. 'Is Democracy the Right System? Collaborative Approaches to Building an Engaged RDM Community'. *International Journal of Digital Curation*, 12(2): 86–95. DOI: https://doi.org/10.1101/103895

**UKRI.** 2018. *Common principles on data policy*. Available at: https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/ (Accessed: 27 January 2019).

**Wilkinson, MD, Dumontier, M, Aalbersberg, IJ, Appleton, G, Axton, M, Baak, A, Blomberg, N, Boiten, J-W, da Silva Santos, LB, Bourne, PE, Bouwman, J, Brookes, AJ, Clark, T, Crosas, M, Dillo, I, Dumon, O, Edmunds, S, Evelo, CT, Finkers, R, Gonzalez-Beltran, A, Gray, AJG, Groth, P, Goble, C, Grethe, JS, Heringa, J, Hoen, PA't, Hooft, R, Kuhn, T, Kok, R, Kok, J, Lusher, SJ, Martone, ME, Mons, A, Packer, AL, Persson, B, Rocca-Serra, P, Roos, M, van Schaik, R, Sansone, S-A, Schultes, E, Sengstag, T, Slater, T, Strawn, G, Swertz, Ma, Thompson, M, van der Lei, J, van Mulligen, E, Velterop, J, Waagmeester, A, Wittenburg, P, Wolstencroft, K, Zhao, J** and **Mons, B.** 2016. 'The FAIR Guiding Principles for scientific data management and stewardship'. *Scientific Data*, 3: 160018. DOI: https://doi.org/10.1038/sdata.2016.18