

University of Cambridge

The University of Cambridge is a large research-intensive university with research interests spanning a broad spectrum of disciplines.

Expectations for research data

The University of Cambridge has no formal requirements for research data management, but there are guidelines and recommendations for best practice laid out in the University's [Research Data Management Policy Framework](#) [1].

Researchers who are in receipt of funding from external funding agencies are expected to comply with the RDM requirements of their funders.

Research data management support

The first resources for research data management (RDM) at Cambridge were developed as part of the two-year 'Incremental' project [2] in 2010-12. This project resulted in the development of a set of RDM webpages and the provision of some RDM training for staff. Once the Incremental project funding ended, the resources remained available, but were not developed further.

In 2015, with the deadline for the EPSRC expectations for research data looming, RDM support was revisited. Feedback from researchers was used to determine the range of researchers needs. This resulted in a [Research Data Facility](#) [3] providing the following RDM support:

- provision of comprehensive information via a website
- support with data management plans
- provision of an [institutional repository](#) [4]

Anyone wishing to see how researchers initially responded to the new Research Data Facility at the University of Cambridge are advised to look at this [infographic](#) [5]. Recent information about the uptake of data deposit opportunities are available in the '1000 datasets' [blogpost](#) [6]. More information about the development of the Research Data Management Facility are available as an [Unlocking Research blogpost](#) [7].

Service sustainability

When the current RDM facility was started up in 2015, its single staff member was paid for from funding designated to develop the University's research support services in the area of scholarly communication.

From 2016 – 2018, the facility was funded from central University funds, but provided as two 'non-recurrent' grants. These awards are

STAFFING LEVELS

Total research-active staff = 5000

RDM staff = 3.0 FTE

Ratio of staff to RDM support = 1330 : 1

MAJOR FUNDING SOURCES

- Wellcome Trust
- UKRI (RCUK)
- European Research Council

RDM SUPPORT

- Research Data Manager (1 FTE)
- Research Data Coordinator (1 FTE)
- Repository Manager (0.25 FTE)
- Repository Integration Manager (0.25 FTE)
- Outreach and Engagement Coordinator (0.25 FTE)
- Scholarly Communication Assistant (0.25 FTE)

intended to be one-off bridging funds designed to support a service until it achieves sustainability, but two rounds of this funding was needed due to the complexities of developing a business model for the Research Data Facility.

Initial sustainability proposal

Guidance from RCUK (April 2013) – *it is permissible to recover costs from grants through direct charges or overheads, but institutions must not charge twice... it is permissible for institutions to recover costs of RDM facilities as other Small Research Facilities...*

Cambridge's proposal

- Establish the Research Data Facility as a Small Research Facility according to the Transparent Approach to Costing (TRAC) methodology.
- Recover facilities costs from grants as directly allocated (DA) costs. This option had two significant advantages – transparency to the funder of the Facility's activities and awareness raising with researchers for the need to consider RDM seriously (both time and money).
- The total direct annual cost of the Facility would be less than £200,000, and would cover 3.2 FTEs, and other service costs, but not repository costs, which would be charged directly.

How much to charge?

The Cambridge team investigated many different models for charging RDM services to grants:

- Ideally, the Facility cost would be accurately measured based on information provided in a Data Management Plan, but not all funders require DMPs, and DMPs do not necessarily estimate the amount of RDM assistance required during the lifetime of the grant.
- Charging based on the number of people in a research team. Since the training component of the Facility was measurable by attendees to workshops, this was investigated as an option. But, this information was not easy to extract from grants and this could cause problems in charging for collaborative grants.
- Charging every grant using the Facility services an amount proportional to the size of the award. But, there was no evidence that large grants required more Facility assistance.
- Charging every grant using the Facility a flat rate. This model had the fewest 'buts...' and was also the most cost-effective in terms of administration. This model was included in the eventual business case.

Funder feedback

The business case for funding the Research Data Facility from grant income was put out for consultation with members of the research community at Cambridge. The business case was then presented and cleared through the necessary committees of the University. Once mature, the business case was shared with research funders to ensure that they would be willing to support the proposal.

Wellcome Trust – would not allow direct charging to grants for this facility. They considered these costs to be overheads, which they do not pay.

Cancer Research UK – was positive about the transparent approach to costing. On discussion with senior management at CRUK, they agreed to consider RDM costs as direct costs on grant applications, but only on a case-by-case basis, if appropriately justified in the application.

RCUK – considered the proposal but decided that these costs had to be charged to grants as overheads instead of direct costs.

Eventual Sustainability Solution

In 2017, the Office of Scholarly Communication (of which the Research Data Facility is part) was subject to a University review [8] of service provision and staffing levels. The highly positive outcome of this review supported an application to the annual University Planning Round for ongoing central funding. This was a lengthy process (almost a year), but the application was granted, meaning that the service – in the form outlined in the sidebar above – is now financially stable.

The data repository storage is hosted in the main University Data Centre and is paid for by the University Library. The cost for long-term storage of research data is charged to researchers' grants at a rate of £4/GB, for datasets over 20GB. Storage for datasets under this size is not charged to grants.

The eventual outcome from the work invested in making the Research Data Facility sustainable was good, but the process was long. The business case development started in March 2015 and a long-term funding model was finally secured in September 2018. Throughout this entire process, the Facility staff have had to be kept on temporary contracts. There has been a huge opportunity cost in training people for these roles.

Costing RDM activities

During the development of the business case for RDM support, the staff in the Research Data Facility put a lot of effort into making sure that the management and financial systems of the University were prepared and able to manage the costing of RDM services into grants. Work was also done to ensure that all of the administrative staff involved in managing research grants (over 100 people at the University of Cambridge) were aware of the potential need to cost RDM services into research grants.

Due to rejection of the business case by research funders and the (final) funding of the Research Data Facility from central University funds, the systematic addition of RDM costs to grants has not happened.

There are anecdotal examples of instances where RDM costs have been paid for from grants, but these are few and far between:

Some researchers have told the Research Data Facility that they have money costed into their grants for depositing large datasets in Apollo (the institutional data repository).

One project which shares a large number of video files via Apollo made the decision not to employ two additional people on the grant, so that they could pay the data deposit costs instead.

The Research Data Facility has no involvement in the grant submission process beyond offering to review data management plans (and they calculate that they probably review fewer than 1% of the DMPs which are submitted with applications each year), so they have no mechanism beyond continuing advocacy to improve the rate at which RDM is costed into grant applications.

The only situation in which money costed into grant applications can be accessed by the RDM Facility is when it's intended to pay for long-term storage of a large dataset in Apollo. The Facility charges for the deposit of large datasets to pay for maintenance and upgrades of infrastructure to ensure the long-term availability of the data.

Acknowledgements

Many thanks to Lauren Cadwallader and Danny Kingsley who provided the information for this case study, both directly, in response to questions, and via the ever-informative [Unlocking Research](#) [9] blog.

Links

- [1] University of Cambridge Research Data Management Policy <https://www.data.cam.ac.uk/university-policy>
- [2] The Incremental Project <http://www.dcc.ac.uk/projects/incremental>
- [3] University of Cambridge Research Data Facility <https://www.data.cam.ac.uk/>
- [4] University of Cambridge Data Repository <https://www.repository.cam.ac.uk/>
- [5] Research data sharing at Cambridge <http://www.lib.cam.ac.uk/librarians/oa/data-poster.pdf>
- [6] 'Milestone – 1000 datasets in Cambridge's repository' <https://unlockingresearch-blog.lib.cam.ac.uk/?p=1654>
- [7] Data sharing – build it and they will come <https://unlockingresearch-blog.lib.cam.ac.uk/?p=221>
- [8] Summary of 2017 Review of OSC <https://osc.cam.ac.uk/open-research/joining-scholarly-communication-discussion/open-research-working-group/summary-2017>
- [9] Unlocking Research Blog <https://unlockingresearch-blog.lib.cam.ac.uk/>

This case study was written by Mary Donaldson as part of the Funder Requirements for Research Data project. This project was funded by JISC as part of the wider Research Data Management Business Case and Costings (RDM-BCC) project.

©MMD August 2018

This resource is made available with a CC-BY 4.0 licence.