Research data should be available long-term…but who is going to pay?

There is now a broad consensus that sharing and preserving data makes research more efficient, reproducible and potentially innovative. As such, most funding bodies now require research data to be stored, preserved, and made available long-term. But who is going to pay for this to happen? **Marta Teperek** *and* **Alastair Dunning** *outline how the costs of long-term data preservation are not eligible for inclusion as part of any funding body's grants. Neither is it currently realistic for these costs to be absorbed by research institutions. With discussions between funding bodies and institutions yet to bear fruit, perhaps it is time for joined-up national (or international) efforts on data preservation.*

There are lots of drivers pushing for the long-term preservation of research data, and to make it Findable, Accessible, Interoperable, and Reusable (FAIR). There is a consensus that sharing and preserving data makes research more efficient (no need to generate the same data all over again), more innovative (data reuse across disciplines), and more reproducible (data supporting research findings made available for scrutiny and validation). Consequently, most funding bodies require that research data be stored, preserved, and made available for at least ten years.

For example, the European Commission requires that projects "develop a Data Management Plan (DMP), in which they will specify what data will be open: detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and reuse, and how it will be curated and preserved."

But who should pay for that long-term data storage and preservation?

Given that most funding bodies now require that research data is preserved and made available long-term, it is perhaps natural to think that funding bodies would financially support researchers in meeting these new requirements. Coming back to the previous example, the funding guide for the European Commission's Horizon 2020 funding programme says that "costs associated with open access to research data, including the creation of the data management plan, can be claimed as eligible costs of any Horizon 2020 grant".

So one would think that the problem is solved and that funding for making data available long-term can be obtained. But then… why would we be writing this blog post? As is usually the case, the devil is in the detail. The European Commission's financial rules require that grant money can only be spent during the timeline of the project (and only for the duration of the project).

Naturally, long-term preservation of research data occurs only after datasets have been created and curated, and most of the time only starts at the time the project finishes. In other words, the costs of long-term data preservation are not eligible costs on grants funded by the European Commission [1].

Importantly, the European Commission's funding is just an example. Most funding bodies do not consider the costs of long-term data curation as eligible costs on grants. In fact, we are not aware of any funding body which would consider these costs eligible [2].

So what's the solution?

Funding bodies suggest that long-term data preservation should be offered to researchers as one of the standard institutional support services. The costs of these should be recovered within overhead/indirect funding allocation on grant applications. Grants from the European Commission have a flat 25%-rate overhead allocation. Which is already generous compared with some other funding bodies which do not allow any overhead cost allocation at all. The problem is that at larger, research-intensive institutions overhead costs are at around 50% of the original grant value.

This means that for every €1m researchers receive to spend on their research projects, research institutions need to find an extra €0.5m from elsewhere to support these projects (facilities costs, administration support, IT support, etc.). Therefore, given that institutions are already not recovering their full economic costs from research grants, it is difficult to imagine how the new requirements for long-term data preservation can be absorbed within the existing overhead/indirect costs stream.

Reflection

The problems described above are not new. In fact, these were previously discussed with funding bodies on several occasions (see here and here for some examples). But not much has changed so far. There were no new streams of money made available; not through direct grant funding, nor through increased overhead caps for institutions providing long-term preservation services for research data.

In the meantime, researchers (those creating large datasets in particular) continue to struggle to find financial support for the long-term preservation and curation of their research data, as nicely illustrated in a blog post by our colleagues at Cambridge.

Since discussions with funding bodies held by individual institutions do not seem to have been fruitful, perhaps the time has come for some joined-up national (or international) efforts. Could this be an interesting new project to be tackled by the Dutch National Coordination Point Research Data Management (LCRDM)?

[1] Some suggest that the costs are eligible if the invoices for long-term data preservation are paid during the lifetime of the project. However, this is only true if the invoice itself does not specify that the costs are for long-term preservation (i.e. says that the invoice is simply for "storage charges", without indicating the long-term aspects of it). Which only confirms the fact that funders are not willing to pay for long-term preservation and forces some to use more creative tactics and measures to finance long-term preservation.

[2] Two funding bodies in the UK, NERC (Natural Environment Research Council) and ESRC (Economic and Social Research Council), pay for the costs of long-term data preservation by financing their own data archives (NERC Data Centres and the UK Data Service, respectively) where the grantees are required to deposit any data resulting from the awarded funding.

This blog post originally appeared on the Open Working blog and is published under a CC BY 4.0 license.

Featured image credit: janeb13, via Pixabay (licensed under a CC0 1.0 license).

Note: This article gives the views of the authors, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our comments policy if you have any concerns on posting a comment below.

About the authors

Marta Teperek is the Data Stewardship Coordinator at TU Delft, the Netherlands. She previously worked at the University of Cambridge, leading the creation and development of its Research Data Management Facility, supporting researchers in good management and sharing of research data. While at Cambridge, Marta initiated and oversaw the Data Champions programme and the Open Research Pilot. Marta is a scientist by training; she completed a PhD in epigenetics at the University of Cambridge. She is an advocate for open research and better transparency in science, and tweets @martateperek.

Alastair Dunning is Head of 4TU.Research Data, based at the Technical University of Delft, the Netherlands.