UK Data Service

Best practice in governance of data for research: Licensing and accessing
Wed, Apr 18, 2018 3:30 PM - 4:30 PM BST

Q&A

Q. What's the recommendation for data retention? (for personal research data)? Most national guidance documents (code of conducts etc.) advise that data is retained for 10 years. GDPR suggests that personal data are kept for the minimal time period necessary. However, there are exceptions for scientific/archival purposes. So is the rule of a thumb recommendation to store personal data for 10y for archival purposes?

SC: As an archive, we have a role for long-term preservation and curation. So we take materials into the collection, into the archive in perpetuity. Many archives will have this arrangement and will have successor arrangements in place, should they cease to exist. There's two strands to this: the archival role; and the data collected in order to undertake the research project initially which depends on the purpose. How long you keep it depends on the purpose. Whether you want to re-interview the participants or you're doing a longitudinal study is important as it means you have a duty to hold any personal information safely and for no longer than is necessary, according to data protection legislation. And whether or not the data that you're holding is personal will depend on whether or not data protection concerns apply.

LC: I think there's a lot of talk about 10 years, and I don't think it comes from anywhere in particular; it tends to be a rule of thumb that's used and often used by universities. There is nothing wrong with that, but you need to think about when you've got a longer term preservation role and whether it's personal data or not, because different rules will apply. There are no hard and fast rules apart from if you're collecting personal data to make sure you're doing it according to the law.

Data Access Committees

Q. Does the UKDS have a central DAC which could be used for controlled access datasets? Or is the requirement on depositors to set up DACs on their own?

LC: No we don't run a central DAC; that's not really our role. We could run a DAC on behalf of depositors and actually we are members of some DACS, but we don't run one ourselves. Most of the organisations with controlled access already have a decision-making body or committee that meets regularly to agree research applications. Depending on the size of the Committee, they tend to meet every month or so and review applications. There is no reason why you couldn't convene one yourself, but you may want to think about who actually is responsible for doing this decision making, who the ultimate owner of the data is.

JS: It is very important though to think about this, because what the Administrative Data Service does is provide secretariat and supporting functional work to enable the Committee to operate well. They handle applications, ensure that only useful and relevant business goes forward to the data provider; so it acts as a filter. It is very important to be working closely with a DAC to establish the right kind of kind of screening and triage mechanisms to make that work.

LC: And, just to say that can be an awful lot of work, because the detail needs to be there for the Committee to make the decisions, so it's really important that some human is screening applications beforehand. And quite often you'll have somebody presenting the case and they need to have a really good handle on it.

Q. What's the % of datasets in UKDS under controlled access, and what % of these datasets are managed by DACs?

LC: Overall it's over a very small proportion actually, it tends to be additional variables on top of data sets we have. I think it's something like 150 out of 7200 and they're all managed by some kind of data access committee. Researchers gain access through an approvals panel run by the data owner. For example, for the UK cohort studies, where we have additional Controlled Access variables available via a Secure Lab, applications are passed by their Data Access Committee that meet every month. The Committee is made up of the owners of the survey and various other experts and, we actually sit on that one, to help make decisions about disclosure risk. We would never just accept data solely for Controlled Access, but try to make sure we've got a Safeguarded data set as well. Because otherwise, you're really limiting the number of people who can go in and use the data, currently being limited to people who've got the analysis skills who may be using it for academic work. Undergraduates and Master's students do not gain access to use these data sets. So again, a word of warning, it's quite restrictive putting everything under Controlled Access.

Q. Can researchers from outside the UK deposit data with UKDS? What are the costs of data deposition for controlled access?

SC: Yes they can. We offer, we run an appraisal process where data can be offered: https://www.ukdataservice.ac.uk/deposit-data/how-to

LC: And there's no cost, but we would think very carefully about whether the data are have enough value for us to put it through our Controlled Access mechanism. And again it goes back to having various access levels of datasets in the collection because most people will want to use an End User version whereas specialist users will want the Controlled Access version and it does have inbuilt in-house costs associated with access to controlled datasets

LC: One of the criteria for deciding on appraisal, whether that would via a Controlled Access pathway, would be the amount of people wanting to use it already, and whether there's a big demand for it.

Q. Do you have experience with live, steaming data? If so, do you have any advice about topics that will require special attention?

LC: Yes, we have had our first experiences with live streaming data last year. We're actually running a project with UCL to set up a Smart Meter Research Portal, which will stream live energy data from selected UK households, for the longer term. There are quite a few issues raised, not only from the ownership, consent status of the data, because with things like smart meter data, it's actually the householder who owns the consent to share the data. So you need very explicit consent about what can be done with that. So the key issues are in clarifying consent around these kinds of data, how you're going to deal with ingesting or streaming it fast moving data.

There are the technologies in place to be able to do that. We've actually set up a Hadoop system to be able to handle such large amounts of columnar data. Then also thinking about the curation role, how do you curate the data, and then how do you describe and publish it? How do you add citations to that? Is it around the chunks when people take bits away or are you time-stamping a citation on data, - every day, week, or month? So there's quite a lot of issues around that and yes we are working on all of these things at the moment. And of course, on the licensing, governance and ethics side, because smart meter data and streaming data is only often useful when you interlink it to other personal household attributes. The data be available in a Controlled Access environment.

Q. Can I just check that current version of standard license reflects GDPR needs, or is UKDS still doing work on aligning the latter with former

LC: Yes, we are actually working on updating our licence agreement to reflect GDPR needs. We will be updating our data management deposit information with GDPR compliant License forms and End User Agreements and adding information to our website. We've got a team on that at the moment, but we are not ready to do it now, because we're still having meetings with funders, but we certainly want to advertise when we do have it, so please watch this space.

Q. Are there any special specifications for researchers based in non-UK, non-EU universities for data request and access?

LB: There are different specifications for different data sets. So, for example, Controlled Access data which can only be accessed in a secure lab, can only be accessed in the UK, so non-UK users would need to be a visiting researcher at a UK institution. Otherwise, with Special License datasets, a lot are available anywhere in the world, but there are some that are limited just to the UK. And then End User Agreement data mostly is available to all once you're registered, and that could be used outside the UK.

LC: Anyone can register free with the Data Service. Mostly the data sets are for use for non-commercial purposes, so as long as you can justify your project and you're not going to be selling data, then you can come and use it; and all mostly downloadable data. We do issue user credentials for people who do not have a UK Federation account. That is federated access authentication so that people who are part of UK universities can join easily using their local credentials. Yes, we do have users from all over the world, using data.

Q. Data are often collected by project teams that dissolve after the project. Who therefore runs the DAC after the project? Do you expect, eg, universities to maintain their interest or does the DAC devolve to the UKDA when the data are deposited?

LC: Because we've been archiving and sharing data for 50 years, a person depositing data 30 years ago might have said that 'I want to know who's using my data and I'm going to need to authorise use by seeking my permission'. Of course, within 10 years they've moved or retired or are no longer interested. It's not really sensible to have an individual as a person giving permission, so we tried to guard against that and ask them to either delegate to someone else who's going to here longer than they are, or their your institution could give permission. But, actually, there does need to be some perpetuity around that, and that needs to be built in. It could be up to the department, I guess or the University. In a couple of cases of depositors, they have deposited with us long (long) ago, and was nominated as the decision maker myself, until we negotiated an Open Data license some years later. More often than not either we try and renegotiate away the condition so it no longer needs permission, or if it's a lot of years later then people might be dead anyway so the data present less of a risk. We have had a large program of renegotiation, to make sure these bespoke conditions are removed where possible. I would suggest probably a central department or somebody with a curation role in the institution could do the vetting and approval, but it can be devolved if you want. It's really up to the data owner to decide what they want to do, and if it's delegated ensure that risk is passed back to them, i.e. add a disclaimer.

JS: I think it's also worth thinking about whether the DAC model is really necessary for a given situation. Is there the depth of potential for reuse to really justify going to that really rigorous governance for it? Would it be more effective especially if these sorts of problems are anticipated about being able to go back to associated people? Is it better just to think really from a service point of view, let's have something where you don't have to put that level of rigor around it? There will still be plenty of opportunity for reuse if we can put the data into a situation where it's Safeguarded instead of Controlled.

LC: Okay, I think we've covered all the questions and come to the end of our session, so we want to thank you very much for listening, to us and hope we've answered your questions. Feel free to mail us offline if you want to ask anything else. We will try and summarise some of the questions and put them as an FAQ and we will make we will make them available as well. So, thank you very much and we hope to see you at another webinar or an event okay.

Please see the webinar recording and slides hosted on our past event website:
https://www.ukdataservice.ac.uk/news-and-events/eventsitem/?id=5234