CAREERS

ADAPTED FROM GETTY

DATA MANAGEMENT

For the record

Making project data freely available is vital for open science.

BY QUIRIN SCHIERMEIER

When Marjorie Etique learnt that she had to create a data-management plan for her next research project, she was not sure exactly what to do.

The soil chemist, a postdoc at the Swiss Federal Institute of Technology (ETH) in Zurich, studies the interaction of trace elements in sediments and water. While preparing a grant proposal for the Swiss National Science Foundation last October, she learnt of the funder's new data rules. These require applicants to provide a written plan for the organization and long-term storage of their research data, to help minimize the risk of data loss and provide guidance for other scientists on how to use the data in the future.

Etique found the task daunting. "Data management is really not my primary skill," she says. "I had absolutely no idea how to go about it." She was able to get advice from her supervisor and from ETH's digital library service. Other researchers might not be so lucky, and might not even know what a data-management plan is — let alone why they would need one and how to produce it. Here, we answer these questions.

WHAT ARE DATA-MANAGEMENT PLANS?

A data-management plan explains how researchers will handle their data during and after a project, and encompasses creating, sharing and preserving research data of any type, including text, spreadsheets, images, recordings, models, algorithms and software. It does not matter whether the data are generated by large pieces of research equipment, such as imaging tools or particle accelerators, or from straightforward field observation.

Many funders are asking grant applicants to provide data plans. Requirements vary from one discipline to another. But in general, scientists will need to describe — before they begin any research — what data they will generate; how the data will be documented, described, secured and curated; and who will have access to those data after the research is completed. They must also explain any data sharing and reuse restrictions, such as legal and confidentiality issues. Researchers can consult their funder and their host institute's digital library services for assistance. Colleagues who have previously produced data plans may also be able to help (see 'Keeping stock').

WHO NEEDS THEM?

Data management is one example of the way in which public research sponsors and research institutions are implementing 'open science', the push to make scientific research and data freely accessible. Many funding agencies have made data-management plans mandatory for grant applicants in the past decade or so. All US federal agencies, including the National Science Foundation and the National Institutes of Health, have such policies. Data-management plans must also now be included in grant proposals to the European Research Council and other European Union-funded research programmes. And many national funding agencies in Europe — including the UK research councils and the London-based Wellcome Trust, world's largest biomedical research charity — also ask for data plans.

Many scientists already practise data management by default. Astronomers, for example, have been doing so for decades when calibrating their observations and archiving huge amounts of telescope-survey data in standardized, machine-readable catalogues for reuse.

Geneticists, too, use special data repositories to archive the vast amounts of DNA and genome-sequencing data (see go.nature.com/2omlrbe). But less data-intensive fields of science and social research also benefit from data management. For example, geochemists analysing soil bacteria and mineral products in different environments can use it to ▶

▶ collaborate more easily. "In the emerging era of open science, any researcher must be prepared to open up their research processes and results," says Eloy Rodrigues, library director at the University of Minho in Braga, Portugal, who coordinates FOSTER, an EU-funded open-science e-learning portal.

Still, many scientists are unsure about open-data provisions, and what grant applicants need to do. A 2017 survey of early-career researchers in Europe found that many were unaware of new open-data policies (see go.nature.com/2oqx9s5). Only one-quarter of the 1,277 respondents to the survey, carried out by the European Commission and the European Council of Doctoral Candidates and Junior Researchers (Eurodoc), had actually written a data-management plan; another quarter said they didn't even know what such a plan might be. Most said they'd not received any relevant training or support from their institutions.

"Data management is inevitably going to be an essential skill in the open-science era," says Eurodoc's president, Gareth O'Neill, a linguist at Leiden University in the Netherlands. "And yet, many scientists are scarcely familiar with what it is all about." The situation in the United States is hardly different, adds Stephanie Simms, a research-data specialist with the California Digital Library (CDL) in Oakland. "We are still at the beginning of a profound shift in research culture," she says.

WHERE CAN I GET HELP?

The University of California Curation Center, part of the CDL, and the Digital Curation Centre in Edinburgh, UK, provide examples of data-management plans written by researchers from various fields. The centres also provide online tools for writing data-management plans that meet the demands of most funding organizations in both countries. Versions of the tools are also available for scientists in several other European countries, as well as for those in Australia, Canada and South Africa (see go.nature.com/2oquiyz).

Simms recommends that grant applicants who are unfamiliar with open-data provisions consult funding-agency programme officers about any field-specific requirements. For more technical guidance, on requirements for machine readability of data protocols, say, or on file formats used by institutional data repositories, scientists should consult their host institute's digital library services, she adds.

Etique did just that. Staff members at the ETH's digital-curation office briefed her about Switzerland's new open-data policies, and provided her with a generic template for drawing up her data-management plan in line with the requirements of the Swiss National Science Foundation.

"It was a bit tricky to address some of the questions, such as file-naming conventions and metadata standards," she says. But after speaking with information-technology services and ETH library staff, she spent two weeks producing a five-page plan that met all of the funder's requirements.

Complying with data-management rules is not just another box to tick, says Rachael Ainsworth, an astrophysicist at the University of Manchester, UK. "Your primary collaborator is yourself six months from now, and your past self doesn't answer e-mails," says the open-science advocate, who regularly hosts data-management workshops. "So handling and storing your data in an organized way might save you time and resources."

DO THE PLANS VARY ACROSS DISCIPLINES?

Data-management demands vary widely, and different research communities (and funders) have different customs and practices. The plans needed for collaborative particle physics, where powerful accelerator facilities generate huge volumes of experimental data, look very different from those used in smaller research projects, such as Etique's.

Sarah Jones, a researcher for the Digital Curation Centre who is based at the University of Glasgow, UK, says any data that serve as evidence for a researcher's claims and results should be archived (the centre was set up in 2005 to champion the management of research data at UK higher-education institutions). This does not mean that a researcher should preserve all of their records, including their lab journal, for posterity, she adds. Indeed, many scientists whose thesis might rely on a limited number of field observations might need to archive only a small amount of data. And if a project does not generate or reuse any data, as could be the case in purely theoretical science or conceptual work, a data-management plan might not be necessary.

Archived research data must be accompanied by appropriate metadata describing their origin and purpose, so that others will be able to find, read and understand them. Scientists who are unsure about metadata requirements, or about which protocols and digital archives to use for their data proper, should contact their host institute's library services, says Jones.

Scientists who generate data should specify who will curate the information after the research project is complete. This is essential because scientists spend only so long at a given institute or department. And to guarantee long-term data availability, they should assign that curation responsibility to an office — usually a library department at their current host institute — rather than to a person.

Library departments typically do not curate individual data sets; rather, they archive and maintain institutional repositories so that any data stored there can be accessed indefinitely.

WILL THEY IMPROVE MY SCIENCE?

Access to research data preserves the rights of researchers anywhere to reach independent conclusions about published science. So it's a good idea for scientists to keep track of their data in case other researchers fail to reproduce the same results, says Jones, or in case legal or ethical problems arise after a paper is published. But not all data types and records can be generously disclosed and freely shared. For example, patient data and health records normally must be anonymized. The same applies to some interview recordings used in empirical social research, such as political surveys or those on personal behaviour.

ADAPTED FROM GETTY
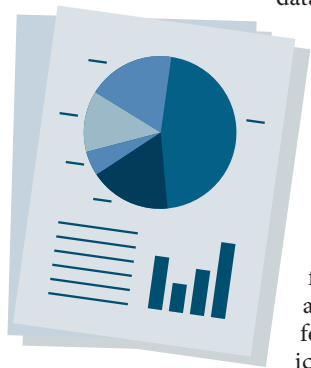
Data-management plans must also state any constraints regarding confidentiality or copyright, for example. These might relate to collaborations between academic scientists and industry researchers or military services. "Carefully consider data privacy and ethical aspects when writing your plan," says Ainsworth, adding that ethical, legal or other constraints should be noted.

European research funders will address confusion over open-data policies by setting out minimum standards for discipline-specific data-management plans. The exercise should be completed in a year. "It just doesn't make sense that different bodies have different rules and requirements when the overarching aims are all the same," says Peter Doorn, director of data archiving at the Royal Netherlands Academy of Arts and Sciences in Amsterdam, who chairs a joint working group on the topic. "Researchers would rather have clear, not-too-detailed instructions all in one place."

Scientists needing guidance can check the EU-funded FOSTER portal for webinars and training material on data-management plans (see go.nature.com/2oq4byo). A toolkit, tailored for applicants to the EU's Horizon 2020 research programme — a 7-year, €77-billion (US$95-billion) research-funding programme — becomes available in May, says Rodrigues.

Etique, meanwhile, hopes that the data plan that she has submitted with her grant proposal will be reviewed favourably. She expects a funding decision about her project later this year. "It was an opportunity to consider my handling of my research data — it makes sense to think early on about the types and amount of data you will collect with each method and instrument, and how to organize those data for effective use," she says of her first foray into data management. Such a plan, she notes, can also help scientists to avoid potential problems with data loss and reproducibility. "It may save you a lot of unforeseen trouble," Etique says.

Unlike the volatile mercury compounds she wants to study, her data are designed to endure. ■ **SEE EDITORIAL P.286**

Quirin Schiermeier *is Nature's senior correspondent in Germany.*

TURNING POINT
Benefit balancer

Jens Magnusson, a stem-cell biologist at the Karolinska Institute in Stockholm, works extensively with mice, despite becoming a vegetarian in 2012. He explains how he reconciles his personal values with his programme of research.

Why did you become a vegetarian?
It was mainly down to environmental concerns. Meat consumption has a big impact on the environment, partly because animal-derived foods require a lot of energy and resources to produce. I think that as a society, we must shift to a more sustainable diet and that vegetarianism will be a crucial part of the shift. My vegetarianism is motivated by my utilitarian perspective. For me, the positive taste experience that accompanies eating meat is smaller than the negative environmental effect of meat consumption.

Your published research has involved working with animals. Does that pose a problem for you now?
I don't like the fact that biological research often requires experimentation on animals. However, it is not always possible to replace research animals with other systems such as cell cultures or computer models. For example, when developing pharmaceutical drugs for use in humans, we must find out how the drugs behave in whole organisms before we administer them to people. Similarly, understanding fundamental processes in biology requires the study of such processes in the context of living organisms. I think that the benefits of animal experimentation, which include increasing scientific knowledge and creating new medicines, outweigh the negative ethical consequences such as animal suffering. The advantages of animal research are sufficiently important to motivate experimentation on animals.

Do you have an ethical objection to eating meat or fish?
No. But the more that we learn about how the brains of humans and other organisms work, the more we realize that the subjective experiences of pain, stress and discomfort are not unique to people. I think that this will make it increasingly hard to argue, from an ethical point of view, that animals should be killed for food.

How do you reconcile your concerns with the need to conduct research?
I do not re-evaluate my fundamental moral position every time I do an experiment. Yet, at the same time, I aim to be guided by my concern for animal welfare. When I plan an experiment, I try to evaluate the level of discomfort that the animals would experience and I let that evaluation strongly guide the experimental design. I also attempt to reduce the total number of animals that will participate in experiments, for example, by using mice that are unsuitable for other researchers' work.

Do you try to persuade colleagues to become vegetarians?
When I initially became a vegetarian, I was eager to share my thoughts with other people. But I found that this did not really make any of my friends or family change their behaviour.

Has being a vegetarian had an effect on your career progression?
I don't think it has changed the kind of research projects that I want to be involved in. I feel comfortable with my decision that the benefits of animal research outweigh the moral costs. But I try to re-evaluate that decision occasionally, and if I were to change my mind, I would steer my career towards projects that do not require animal research, such as those involving cell culture.

How do your vegetarian or vegan colleagues balance their own concerns?
Different people have different ways of reconciling their beliefs, some of which seem incoherent to me. For example, one colleague feels guilty about doing research on animals, so they ask another colleague to perform the actual experiments. ■

INTERVIEW BY MARTA PATERLINI
This interview has been edited for length and clarity.