# Assessment of awareness of FAIR principles and data management practices for early career scientists (PhDs) in geography

Data Management Practices

FAIR Principles

About you

1st year

2nd year

3rd year

>3rd year

Consent for sharing data

Questionnaire structure

**Start of questionnaire**

Thank-you so much for coming here to fill out this questionnaire!

It will take no more than 10 Minutes and is mostly box-ticking, though I would appreciate extensive text-answers, too.

It is aimed at PhD students from all years in the Department of Geography at the University of Cambridge. It will ask questions about the awareness of data management practices for your PhD and data sharing. The questionnaire is done in support with JISC, the UK higher, further education and skills sectors’ not-for-profit organisation for digital services and solutions.

In this questionnaire 'data' does not refer to notes that were taken from reading a paper. 'Data' is defined as the newly generated or collected information in some sort of processable format. This questionnaire is focusing on digital data.

You will answer questions related to 4 sections:

- Research data awareness in general (this section)

- FAIR principles (second section)

- Research data awareness specific to your year of study (third section)

- Consent of sharing results (fourth section)

Top of Form

Assessing research data management practices of PhD students in Geography

Thank-you so much for coming here to fill out this questionnaire!   
It will take no more than 10 Minutes and is mostly box-ticking, though I would appreciate extensive text-answers, too.

It is aimed at PhD students from all years in the Department of Geography at the University of Cambridge. It will ask questions about the awareness of data management practices for your PhD and data sharing. The questionnaire is done in support with JISC, the UK higher, further education and skills sectors’ not-for-profit organisation for digital services and solutions.

In this questionnaire 'data' does not refer to notes that were taken from reading a paper. 'Data' is defined as the newly generated or collected information in some sort of processable format. This questionnaire is focusing on digital data.

You will answer questions related to 4 sections:  
- Research data awareness in general (this section)  
- FAIR principles (second section)  
- Research data awareness specific to your year of study (third section)  
- Consent of sharing results (fourth section)

\*Required

How many locations do you store your data at (syncing)?

* Bottom of Form
* one
* two
* three
* > three
* I don't sync my data

If you do sync your data, what do you sync it with? Please select all that are appropriate.

* Google drive
* dropbox
* onedrive
* personal hard drive
* departmental cloud storage ([files.geog.cam.ac.uk](https://www.google.com/url?q=http://files.geog.cam.ac.uk&sa=D&ust=1504953885033000&usg=AFQjCNGl93E03TBtWA7PmEoEeYWWlg9zoQ))
* Departmental hard drive
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

How many backups of your data do you have? \*

* none
* one
* two
* three
* more than three

If you do back up your data, where do you back it up to? Please select all that are appropriate. \*

* google drive
* dropbox
* usb
* personal hard drive
* departmental cloud storage ([files.geog.cam.ac.uk](https://www.google.com/url?q=http://files.geog.cam.ac.uk&sa=D&ust=1504953885038000&usg=AFQjCNEN-eEjutdFtRc2qsVbMydY9jbYig))
* Departmental hard drive
* hpc cluster space
* I have no back up
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

How often do you back up your data? \*

* more frequently than weekly
* weekly - monthly
* monthly or more
* I don't back up my data
* Other:
* How big is the average data file of yours? \*
* KBs
* MBs
* GBs
* don't know
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Have you ever experienced a data disaster ([nearly] lost important data)? If yes, tell me below in a few words. If no, please skip this question.

What format do you store your data in (multiple answers possible) ? \*

* Geographic Information System?GIS(.gpx, .kml)
* Computer aided design/CAD(.dwg, .dxf, .pln)
* Matlab (.m, .mat)
* Audio(.aiff, .mp3, .wav, etc)
* Web(.html, .xhtml)
* scripts or code
* Video(.avi, .mov, .mp4, etc)
* Database (.db, .mdb, .pdb, .sql)
* Statistical Data/SAS,SPSS (.sav, .sdq, .spv)
* Image(.bmp, .gif, .jpg, .png, .ps, .psd, .svg, .tif)
* Scanned documents(.pdf)
* Data/text(.csv, .dat, .xml, json.)
* Spreadsheet (.wks, .xls)
* Word processor(.doc, .docx, .log, .rtf, .txt)
* Don't know
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Do you think other people could navigate and understand the data you are generating for your PhD thesis if you gave it to them today? \*

* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Do you think that matters? And why?

Do you know what happens with your PhD project's data after you finish your PhD and leave Cambridge? \*

* They will be all handed over to my supervisor
* They will be destroyed
* Valuable data will be shared publicly in a repository
* I don't know what will happen with my data
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

 a

* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## Section: Fair Principles

FAIR Data Principles act as guideline to enhance data-reusability. They aim to give guidelines on how make data Findable by everyone, Accessible to everyone, Interoperable (easily usable in your different analysis tools without spending hours on rearranging the input) and Reusable in general, by describing it enough so that there is enough context given with the data so other people can work with it.   
I want to understand how much you instinctively agree with these principles. You don't need to know any more about them.

Do you keep a notebook or README where you document how you changed your data during processing? \*

* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

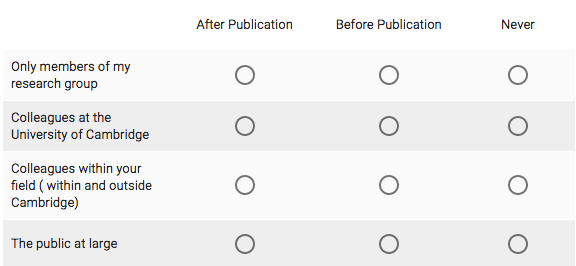
How long do you feel research data should be stored for? \*

* less than 1 year
* 1-5 years
* 5-10 years
* more than 10 years
* Indefinitely
* Don’t know

How important is it to share research data? \*



Please indicate your willingness to share your research data in general with different audiences. \*



What are the reasons for you not wanting to share your data?

Can you imagine anyone wanting to reuse your PhD data in particular? \*

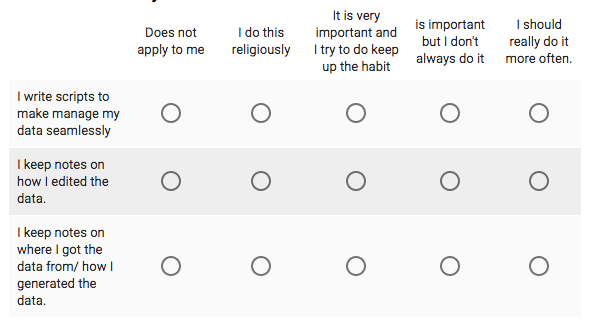
* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Why?

Is your data in a good shape to be easily fed into the appropriate kind of analysis software? \*

* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

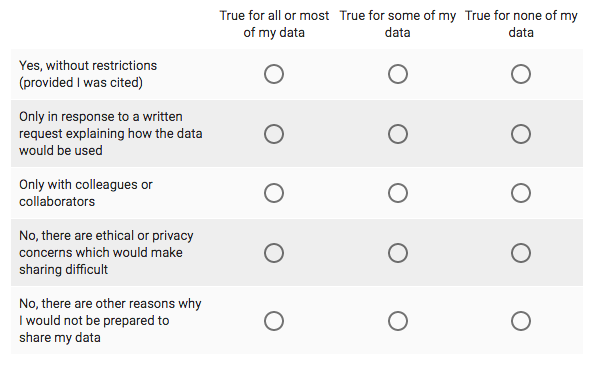
What describes you most: \*



Select where the data for your PhD comes from \*

* I generate/collect it myself
* From data repositories
* From the Web, but not held in a data repository
* From a discussion or analysis of the data in a published article/book
* From collaborators/supervisor
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Would you be happy in the future (e.g. after an embargo period of 1 or more years from the end of the research) to share the data that you are responsible for in your current research? \*



## Section: About You

Don't worry, this is very superficial and all completely confidential.

What year of your PhD are you in? \*

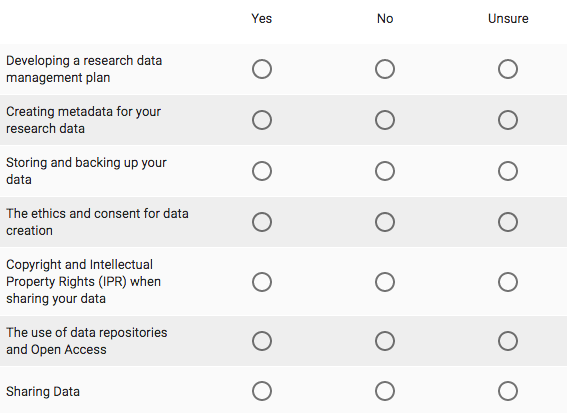
* 1st
* 2nd
* 3rd
* > 3rd

## About you (1st year) cont.

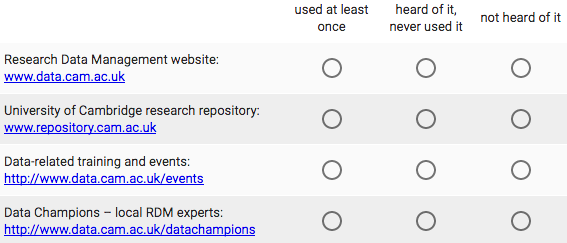
"A data management plan describes the data that is collected or generated in the course of your work and what happens to this data during its life-cycle (storage, publication, citation, long-term availability, anonymity, deletion, etc.). The goal of a data management plan is to meet the requirements of good scientific practice and to allow for reproducibility of research results." (ETH Zurich, Datenmanagementplan\_EN.pdf). Do you think such a plan would have been a helpful element in your first year Research Proposal, to get you thinking about your data during your PhD? \*

* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Would any training be useful to you in the following areas of research data management in general? \*



Have you heard of the following Cambridge resources that can help you with RDM? \*



Do you think that a general course in data management practices in the PhD induction would help improve your data management and efficiency for your PhD project and beyond? \*

* Yes
* No

## About you (2nd year) cont.

Don't worry, this is all completely confidential.

"A data management plan describes the data that is collected or generated in the course of your work and what happens to this data during its life-cycle (storage, publication, citation, long-term availability, anonymity, deletion, etc.). The goal of a data management plan is to meet the requirements of good scientific practice and to allow for reproducibility of research results." (ETH Zurich, Datenmanagementplan\_EN.pdf).

Do you think such a plan would have been a helpful element in your first year Research Proposal, to get you thinking about your data during your PhD? \*

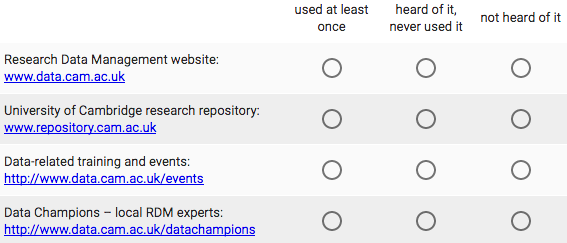
* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Looking back, do you think that a course in data management practices in the PhD induction would help improve your data management and efficiency for your PhD project and beyond? \*

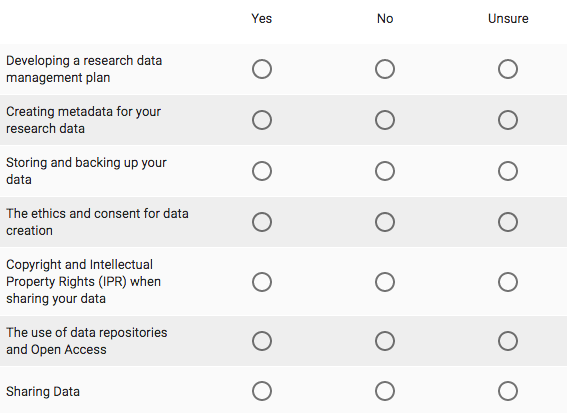
* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Do you think such a course would still be valuable for your PhD and future carreer? \*

* Yes
* No, I have found my own data management system.
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Have you heard of the following institutions that can help you with RDM in Cambridge? \*

Would any training be useful to you in the following areas of research data management in general?



Do you feel prepared to manage your research data after your PhD? \*

* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

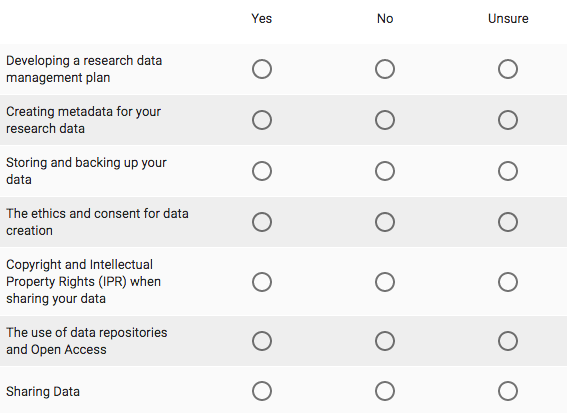
## About you (3rd year) cont.

Don't worry, this is all completely confidential.

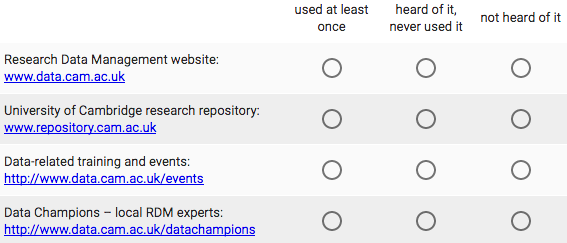
"A data management plan describes the data that is collected or generated in the course of your work and what happens to this data during its life-cycle (storage, publication, citation, long-term availability, anonymity, deletion, etc.). The goal of a data management plan is to meet the requirements of good scientific practice and to allow for reproducibility of research results." (ETH Zurich, Datenmanagementplan\_EN.pdf). Do you think such a plan would have been a helpful element in your first year Research Proposal, to get you thinking about your data during your PhD? \*

* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Would any training be useful to you in the following areas of research data management in general? \*



Have you heard of the following institutions that can help you with RDM in Cambridge? \*



Looking back, would you have profited from guidance on data management at the beginning ( for example during the induction events) of your PhD? \*

* Yes
* No, I have come up with my own data management system.
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Do you feel prepared to manage your research data after your PhD? \*

* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

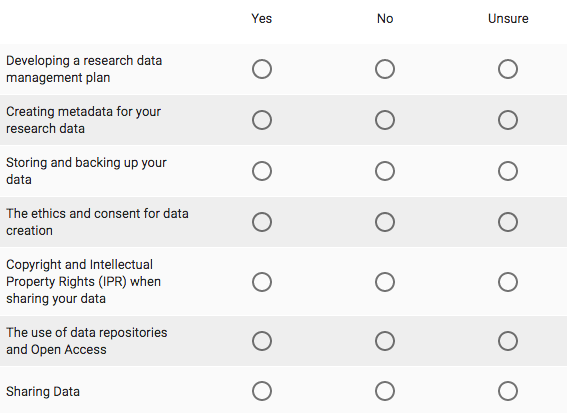
## About you (>3rd year) cont.

Don't worry, this is all completely confidential.

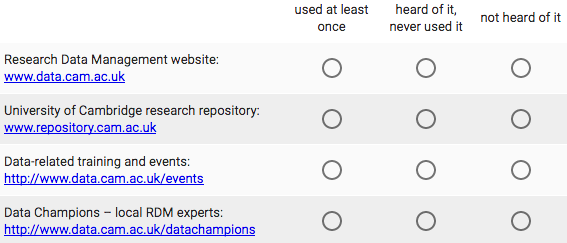
"A data management plan describes the data that is collected or generated in the course of your work and what happens to this data during its life-cycle (storage, publication, citation, long-term availability, anonymity, deletion, etc.). The goal of a data management plan is to meet the requirements of good scientific practice and to allow for reproducibility of research results." (ETH Zurich, Datenmanagementplan\_EN.pdf). Do you think such a plan would have been a helpful element in your first year Research Proposal, to get you thinking about your data during your PhD? \*

* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Would any training be useful to you in the following areas of research data management in general? \*



Have you heard of the following institutions that can help you with RDM in Cambridge? \*



Looking back, would you have profited from guidance on data management at the beginning ( for example during the induction events) of your PhD? \*

* Yes
* No, I have come up with my own data management system.
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Do you feel prepared to manage your research data after your PhD? \*

* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## Consent to share results

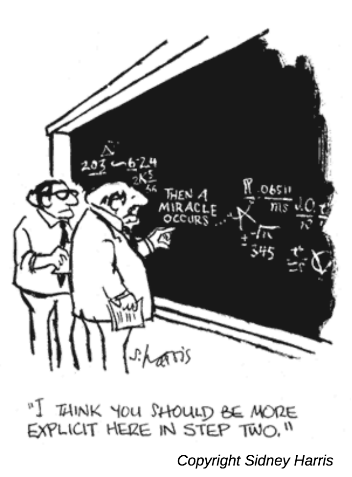
Before submitting this form, there is one last question. Although you have not actually provided me with any sensitive data, I still need to ask you whether you are happy for me to make these anonymous survey results of yours publicly available. Final anonymous answers will be shared publicly on an open access repository as a good example for data sharing.

Would you be willing for me to make these anonymous survey results publicly available? \*

* Yes
* No
* Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

If you would be willing to volunteer for a follow-up interview, please leave your email address below.

Supervisor feedback:



Note that some questions in this survey were reused with permission