UK Data Service

Online workshop: Why and how to document your workflow

Dissertation Week 2021

UKRI Economic and Social Research Council

Table of Contents

1	• **What is a workflow?**
2	• **Why do workflows need to be documented?**
3	• **Great British Science Off**
4	• **What (and how) to document?**
5	• **When to stop documenting**

What is a workflow?

Broadly, a workflow is an ordered series of actions that produced an outcome.

In a scientific or research context, it is all the details of all the (ordered) steps that would allow another researcher to produce your "same" outcome.

Why do workflows matter?

Allowing other researchers to produce your "same" outcome (a.k.a. reproducibility or replicability) is a key part of the scientific method!

Since early science (especially social sciences and medicine) have been struggling with a crisis of reproducibility.

Why should scientists document workflows?

- Allows others to reproduce, validate or extend the work
- Helps others to understand the work and its results
- Contributes to trust in the work and its results
- Promotes a culture of open science!

But why should **I** document **my** workflows?

- Well-documented workflows = easy methods sections

- Documentation contributes to memory/understanding

- Fast/easy way to:
  - Identify and correct errors,
  - Incorporate new/different data,
  - Re-run tests, re-create images, etc.
  - Apply work to new topics/areas

- Probably more

Realistically, a workflow comes from…

Keeping VERY good records of:
- All the work that you did,
- All the data you used (source, version, dated, etc.),
- All the processes you employed,
- All the decisions you made,
- All the analyses you ran,
- All the tools you used (software, packages, materials, etc.)
- All the visualisations you chose…
- Etc.

Second check in for questions or comments

Now is a good time to take a few minutes break.

Ask any questions or share any comments you might have on what has been covered so far in the Zoom Q&A.

Also, please stand up, stretch, get a drink, etc.

Great British Bake Off illustrates reproducibility

Specifically, the "Technical Challenge" round.

Great British Bake Off technical round

Contestants have to produce perfect (e.g. the "same" as an ideal) classic recipe using :

- Vague instructions
- Identical ingredients (may not all be needed)
- Equivalent equipment (may include choices)
- Fixed time limit

Comedy ensues.

Great British Reproducibility Off

The Bake Off would be less entertaining if the technical round took out all of the guesswork.

In contrast, well documented workflows actively seek to remove all the guesswork.

What should I document? And how?!?

Let's get interactive!

What are some things you do that need to be documented?

literature search

focus groups

literature references

research questions

qualitative research

interviews

surveys

shadowing

data analysis

data exclusions

essays

references

observations

different interactions

transcribing interviews

How would you document the things you do?

excel spreadsheet

spreadsheets

recorder

word

write methods as go along

cloud

make a table in word

statistical code

stata do files

keep a diary

pdf of surveys

written notes

written

transcriptions

bullet points

audio - video recordings

methods in word document

table in word

What are some of the decisions you make that need to be documented?

geographical regions

choice of analysis method

ethnicity

data cleaning

inclusion criteria

boundaries of research

data to include or exclud

power dynamics

research questions

personal inspirations

data inclusion criteria

software to use

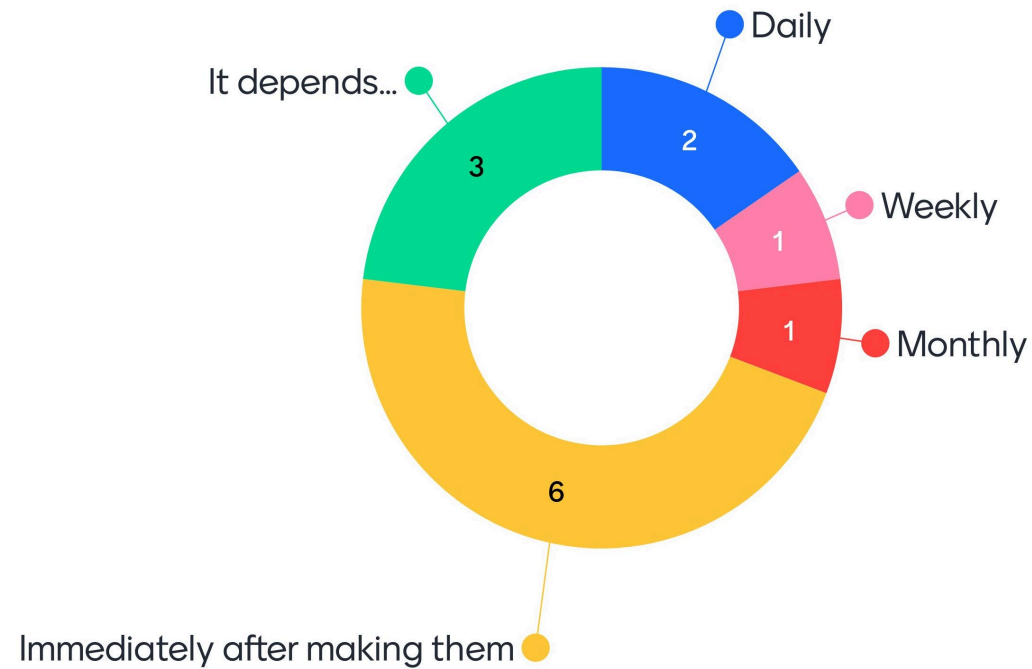comply to gdpr

survey questions

which databases accessed

data excluded or included

changes to excl criteria

ethical considerations

How frequently do you think you should document the decisions you make?

- Daily — 2
- Weekly — 1
- Monthly — 1
- Immediately after making them — 6
- It depends… — 3

How do you document the ideas that influence your work?

Checklist

bookmark tab; save article;

Don't have a good system for this but want one!

Word document

Keep notes on mobile device, take screenshots, record audios

Notepad of ideas and then discuss with tutor/research team to get focused on what we'll actually take on. Then may pull the relevant papers to cite

Life stories

How do you get data?

Interviews, Surveys, Questionnaires

Interviews mostly

Life stories

Primary or secondary data collection (books, journals, websites, interviews, surveys, observation)

Previous research papers, focus groups, transcripts, questionnaires, apps/websites, government publications

How do you document your data acquisition?

Haven't been doing it. That's why I'm attending this session :-)

ethics application and ongoing documentation in methods section

Well-documented workflows can include:

- Details of data (source, volume, descriptive statistics, representativeness, etc.)

- Data processes (how stored, recoded, linked, anonymised, analysed, etc.)

- Step-by-step of any experiments run, observations made, models changed, etc.

- Description of any materials, software, etc. used

- Digital resources (raw data, final data, synthetic data, code written, R files, etc.)

- Justifications for any decisions (reference lists, theories, frameworks, written explanations, etc.)

- Anything else that is relevant for your work

A major tip

- Make conscious decisions as early as possible.
- Be realistic.
- Automate where possible!

AVOID AT ALL COSTS!!!

- Manual citations and references
  - This = nightmare of formatting styles, double checking citations to refs and refs to citations, etc.
  - Better = reference management software (EndNote, Bibtex, Zotero, Mendeley, etc.)
- Save files under different names for different versions
  - This = nightmare of indecipherable version names.
  - Better = version control software (a native feature of MS O365)
- Emailing files around
  - This = nightmare of everyone having different versions.
  - Better = load them to a central repository that everyone can access.

Third check in for questions or comments

Now is a good time to take a few minutes break.

Ask any questions or share any comments you might have on what has been covered so far in the Zoom Q&A.

Also, please stand up, stretch, get a drink, etc.

When to stop documenting

Do not document or share anything that you do not have a right to document or share

Also, stop documenting when…

You are spending more time on the documentation than on the actual work.

If you find the documentation to workflow balance is off try:

- Automating the boring stuff

- Using better tools/processes

- Scaling back/rescoping the project

- Recruiting help for sanity checks

- Other suggestions?

Thank you.

Dr. J. Kasmire

Email to j.kasmire@manchester.ac.uk

@JKasmireComplex on Twitter

@JKasmireComplex on Twitch