The early bird get the return: how to decide when to publish data

I prefer to publish when the fieldwork or synthesis research is done for many reasons.

Abstract   Publish your data when data collection is complete. Not novel, but not commonly done.

Benefits

a. Better than Excel files or Googlesheets because it forces me to tidy up data to make 'presentable' to some extent sooner versus later.

b. Open data are easier read into R for QA/QC and coupled with GitHub work can be kept all online and distributed with easier remote collaboration.

c. Much easier to share a link with all the meta-data bundled to data even for visual reviews - see above (saves so many emails with the team).

d. Fieldwork is done and paid for so data should be open. Peer review for papers can take time.

e. Data repositories support versioning so we can easily track data integrity.

f. It is a valid research object and can be assigned a citable DOI.

g. You can cite your own data in the paper then submit that paper for review.

h. Often the 'primary' purpose of the data collection is done and published in short paper, pilot, or field experiment. Then, the data are ready for reuse. So, if you feel you have satisfied your primary purpose with even a short experiment or pilot, sharing it just makes all our lives easier. Main reason.

i. It is gratifying to get something out in the world rapidly.
j. Sometimes other people catch errors too and email the authors before it gets to print. So awesome. That happened to me recently, before the paper was published. The referee checked the data and found an error in it. TOTALLY saved the vegan bacon.

Potential limitations

i. Now, could someone else find it and use it? Sure. And I say, enjoy.

ii. They are unlikely to write the same paper(s) we may or may never write about the work. I do not think they could because it would be so hard without being there in many instances.

iii. They could find a new use we are even less likely to do. My experience is not extensive here, but most data reuse is for another new purpose not really linked to the field methods so to speak. So, no risk to us of being scooped but you do risk losing attribution and provenace can be tricky or even outright ignored.

iv. People get the gem or concept of the idea from the data and meta-data and more rapidly do the new or additional work and publish the paper before you.

v. You are often obligated by funders to publish data, and this takes some to a lot of time depending on the meta-data language. This is time not avaible to spend on other activities like writing the paper.

Net outcome   For simple, clean data - publish before paper is written or published. It is more efficient, rapid, open, and ensures better final data with meta-data. Not novel, but unfortunately not novel either. So, we can make a change and accelerate some components of the research process. Similar to pre-registration,

pre-prints, and clinical trial registrations. A little bit more speed and openness will ensure that the early bird gets the return.