



I CAN'T PUBLISH MY DATA

LAME EXCUSES AND SOME ANSWERS

Markus Demleitner
Zentrum für Astronomie der Universität Heidelberg, Germany

SPONSORED BY THE



Federal Ministry
of Education
and Research

You already know why publishing data is the right thing to do, don't you? It's just that exactly your data is an exception, right? It cannot be published because...

Here's a collection of reasons we've heard. And some replies we've always wanted to give (but mostly swallowed).

People will contact me.....

... and ask about stuff.

Well, science is about exchange. Think how much *you* learned by asking other people. Plus, you'll notice that quite a few of those questions are actually quite clever, so answering them is a good use of your time. As to the stupid questions – well, they *are* annoying, but at least for us even those were eye-openers now and then.

People will misinterpret my data

Good documentation and standards mitigate this.

As for what remains – well, if you're publishing prose (i.e., in a journal), how many of your readers do you think actually get what you're writing?

My data is boring...

... or at least not very interesting.

Leave that decision to others. You'd be surprised how much "boring data" people point-and-click out of printed graphs or tables in the sweat of their mice. Every day.

I might want to...

... use it in a research paper.

If you've not done so so far, will you at all? When? Too much data obtained for uncounted kilodollars (or megadollars, for that matter) is gathering dust, waiting for the "real soon now". Be fair to the world and to the people funding you and your research and publish the data. If you're really worried, put a one-year embargo on the material. Procrastinate's Law: If over a year you don't get to do it, chances are overwhelming you'll never do it.

I'm not sure I own the data

That's amazingly common. So: Are you sure you cannot find out who does? If you made a reasonable effort to figure that out but failed, the likelihood is high you've orphaned data on your hands, obtained by people who've long left science for greener pastures. To avoid similar uncertainties with your data later on, please consider assigning explicit licenses to it – ideally CC-0. Do not worry that people will not give credit just because of a Free license. We're in science, and so this is a matter of scientific conduct rather than the law.

My data is embarrassingly bad

Everyone's is. Good data is just bad data that more eyes have seen and more hands have improved.

My data is too complicated

If it's too complicated to explain: are you sure you've understood it yourself? Be that as it may: Try explaining anyway, the improved understanding you'll get will reward you plenty.

I'm busy...

... and it's not a priority.

A-ha! Here we're talking. True, the current system of rewards in science doesn't actually encourage data publishing. But publishing is the right thing to do, anyway, even before the system gets back on a path of recovery. And: more and more funding agencies at least sound requirements for data publishing and preservation.

It's too much work

No, it's not. For example, the GAVO data center is there to help you. Unless your data is particularly funky, you'll not have to spend more than half a day from the half-documented, messy goo that's on your disk to a shiny, blinking, proper, VO-registered, be-proud-of data service. And we'll take care of it henceforth.

Ask around at the booth for more information

Acknowledgment: The good parts of this were blatantly snarfed from Charly Strasser's brilliant web page <http://datapub.cdlib.org/closed-data-excuses-excuses/>

GAVO is funded by the German Ministry of Education and Research (BMBF).