ECOLOGY BITS

Open data, authorship, and the early career scientist

by Margaret Kosmala

June 15, 2016

About a year ago, my coauthors and I published a huge dataset of more than a million annotated images of animals from a camera trap network in the Serengeti. The lead author, Dr. Swanson, and I are both early career scientists, and we both put a ton of time and effort into this dataset. We made the decision to publish the dataset as its own product after more than a half-dozen researchers in other fields (computer vision, citizen science, education) contacted us to ask if they could use our data. Our graduate advisor (and PI-on-paper) wondered whether this was a good idea. If we published the data, he worried, other people could take it and do the sorts of community ecology research that we were hoping to do with it.

I've heard this worry a lot about open data. I've had this worry myself as a grad student. But as far as I can tell, having made this dataset (and others) available, is that the worry about being scooped is *way* overblown for most ecology datasets. That doesn't mean it can't or doesn't happen. But I think it's a rare case when it does. (Can anyone point to a time it's happened?) Instead, opening up the data has meant two great things. First, when people contact us about our data and camera trap network (which happens monthly), we can just point them to the dataset and it saves us a ton of time. Second, there are ecologists using our data in ways we never imagined, including looking at community ecology in groups of animals we don't (small mammals, lizards versus large mammals) and investigating wildlife disease.

Open data is great!

But. (You knew there was going to be a but.) Here's something I haven't heard proponents of Open Science talking about much. If you publish a dataset, you pretty much lose control over authorship.

Traditionally, the way data in ecology worked (and still mostly works) is that you go through a lot of effort to create a dataset. Then you keep it. Hopefully you're smart and you back it up and have other safeguards to make sure it doesn't get compromised. But usually it just sits on your desktop computer somewhere. Then people find out about your data. Probably you published something. Maybe sometimes through word of mouth. And if people want to use your data, they contact you and say, "hey, I have a great idea for an analysis and paper that needs your data. Can we collaborate?" Often this is code for, "if you give me your data, I'll give you co-authorship on the resulting publications."

And there's a reason for this customary tit-for-tat. Producing ecological datasets is far from trivial. It's also nice to know who is using your data and for what. As a data-creator, you want to make sure your data is not misused. Not only do you care about the science coming out right, but because your reputation is attached to the data, a misuse reflects poorly on you, even if it's done by someone else.

The LTER network has an explicit data policy that reads, "The Data Set has been released in the spirit of open scientific collaboration. Data Users are thus strongly encouraged to consider consultation, collaboration and/or co-authorship with the Data Set Creator." Not too long ago this policy was on a site-by-site basis and — at least for the sites that I used data from — contacting the data creator was a requirement for publishing using existing data.

For early career researchers, there's a super important reason for this custom of co-authorship when re-using data. Number of publications matters. It just does. If I have spent some sizeable fraction of my nascent career on developing a particular dataset, I *need* to get credit when that dataset is used for advancing science. And the truth is that number of publications counts way more than number of citations.

So here's the problem: anyone can use data from our big published dataset (please do!), and they will be right and proper to simply cite it. If we hadn't published the dataset, then people would have to contact us about collaborating and my coauthors and I could rack up more publications. Perhaps the data would be used less overall, because it's a bit more effort to exchange a few emails than to simply download a dataset. The crucial point is that Open Data may be good for *science*, but it may be bad for *scientists* — especially early career ones. Not because the authors of open data will be scooped, but because the authors lose credit for their data relative to authors who do don't make their data open.

Share this:

Permanent link to this article:

http://ecologybits.com/index.php/2016/06/15/open-data-authorship-and-the-early-career-scientist/

22 comments

2 pings

Skip to comment form ↓

Bobo

June 15, 2016 at 7:32 am (UTC -4)

"…because the authors lose credit for their data relative to authors who do don't make their data open."

So if journals mandate that all authors absolutely ~must~ make their data open, then we're all back on a level playing field.

Margaret Kosmala

June 15, 2016 at 9:03 am (UTC -4)

Well sort of. There's a difference between publishing a dataset and publishing an article with data. A dataset tends to be more comprehensive and (currently) has a higher bar for complete metadata description. If you publish an article, you usually only make data available that was specifically required for the analyses in that article, which is often only part of a larger dataset.

Even if all funders, institutions, and publishers required that all data be *open* (rather than just available upon request), we have an interesting situation in

which there's an unfortunate incentive to produce and work with small and easy-to-produce datasets. Ecology is moving towards larger datasets generally, but researchers who stick with small scale short term studies may be at an advantage.

I think we as a culture have to somehow recognize (good quality) datasets as separate from articles and having intrinsic value in their own right.
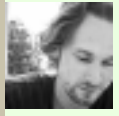
Koen

In a similar fashion there is little love for scientists who make messy data more accessible – even if it's not their own (e.g. fluxnet data). Considerable effort is put into reanalysis of existing datasets, but unless they are published as you did no one will acknowledge this effort. Valorization in this sense has value as well.

Similar problems exist with open source software / code. I only make code available for which I will get credit (or for which I don't care about the terms, still there is a license agreement included though). These days I will refuse to share any code unless properly acknowledged or involved in the project. Coding and maintaining code takes time (especially if you generate a user base, who at that point start to request features).

Remi Daigle

I totally agree that sharing data openly can lead to fewer publications related to said data with you as an author, which can be detrimental to your career. However, I don't believe (and as a early career scientist who shares all his data, I hope that) sharing your data will not have a net negative effect on ones career overall.

1) Putting your data out there is another form of advertising for YOU! Some people will know of your work from your publication, others will know your work from your data. Overall, getting your work out there will allow others to judge the merit of your whole body of work and if you put out quality work (pubs or data) they will want to collaborate with you.

2) Getting authorship(s) by holding your data hostage is unethical. I know that your statement of "Often this is code for, "if you give me your data, I'll give you co-authorship on the resulting publications."" is an accurate portrayal of the current situation, but wow is it ever unethical. It's even unethical on 2 fronts! The first is that in no journal I have ever seen does supplying data ever constitute grounds for being included as an author. I'm not saying it doesn't happen, I'm just saying it's against 'the rules". The other front, which I think is more important, is the inclusivity aspect and it's the same arguments as open access publishing for manuscripts. Just as big publishers (e.g. Elsevier, etc) don't have the moral license to act as gatekeeper for your pubs, you shouldn't hold on to your research data for your own personal gain (perhaps baring an embargo period to allow you to publish on it first). We as scientists may have worked hard to get that data, but we don't own it! It was our job to collect it and the public (generally) paid us to do so! I think it makes sense that the public gets the maximum benefit from what they paid for as opposed to it being selfishly hoarded away.

I know that I am personally more likely to seek out open data and collaborate with people who make their data open. I may not be the majority yet, but I think that your right about recognizing good quality datasets and valuing that as an important contribution to science. In the meantime I hope that by keeping my data open, I can

foster more real collaborations (as opposed to rubber stamping authorship-data trading). Anyway, I don't disagree that sharing data has some negative aspects, but I hope that the positives outweigh the negatives! I hope that you, I, and others keep putting their data out there!

Margaret Kosmala
June 15, 2016 at 12:18 pm (UTC -4)

Thanks for your comment, Remi.

For (1) I agree that there are positive aspects of having open data. But I'm not sure the positive outweighs the negative for early career folks. There's an opportunity cost to making data transparent, well-documented, easy to use, and actually publishing it. I think for my situation in which I have severe time constraints (having small children), my time is better spent on writing publications rather than grooming my data for dissemination. People were finding out about our data before we ever published it, so I'm not sure how much additional publicity it got from publishing, nor how to weigh the value of that. Being known in another field (say, computer vision) does very little for my chances of securing a permanent ecology position. I'm sure the pros and cons may balance out differently for different researchers. I just don't see many discussions about the potential negatives (other than, "agh! I might be scooped!").

For (2a) I'm not so sure it's unethical. Authorship is a peculiar beast and many others have written about it more than I have. Many researchers in ecology — I might even say most — would consider that production of a significant amount of the data going into a paper would merit the invitation of coauthorship. You say that journals would consider this "against the rules." But I've never seen it forbidden. I usually see something along the lines of "authorship should be limited to those who contributed substantially." I'd argue that the provision of data for a paper is a huge contribution. Do you disagree?

For (2b), I have two thoughts. You say "you shouldn't hold on to your research data for your own personal gain." I agree. I am pro Open Science. But it's naive to think that science happens in a world free of people. We have a scientific culture which is populated by scientists, who are human and therefore care about their own lives and careers. Scientific culture is competitive, as well as collaborative. If Open Science proponents want to open up science successfully, they're going to need to appeal to more than a starry-eyed sense of what everyone ought to do for the good of science. They need to figure out how to create cultural, social, and infrastructure systems that reward open practices while mitigating negative effects on individuals. You point out an embargo period for data. Embargoes are great example of a social construct that helps reduce the negative effects of sharing. There's no scientific reason to have them. Are there other practices we can come up with that also help reduce negative effects?

My second thought on this is that researchers don't usually hoard their data away, while they also don't make it open. They are typically required to make it available, which means that they release it upon request. I don't think that's unethical, but it's also not ideal for searching and discovery of data sets.

Remi Daigle
June 15, 2016 at 2:04 pm (UTC -4)

Yes, for (1) it's probably case by case, but in general I would say that the more publicity the better. The issue you mentioned about being

known in the 'wrong field' is definitely an issue, but would the papers that are using your data and not including you as an author in your field anyway? But yes, I totally see your point about the loss of control for the authorship as a legitimate concern. I'm still just hoping the positives are greater than the negatives.

For (2a) I don't disagree that the authors of the data should be acknowledged for their contributions particularly if they contribute to the writing and or analysis. However, I don't believe that supplying the data alone would ever deserve co-authorship. For example, the authorship guidelines for Limnology and Oceanography:"A person claiming authorship or co-authorship of a scholarly publication must have met the following criteria: • Substantial participation in conception and design of
the study, in analysis and interpretation of data or in meta-analysis and synthesis of
research findings; • Substantial participation in the drafting or substantive editing
of the manuscript or scholarly submission; • Final approval of the version of the
manuscript to be published; • Ability to explain and defend the study in public or
scholarly settings. "

Just providing data does not meet the above criteria. Other journals may be more permissive or less specific, but I still think that in general including a 'data-only author' is unethical. That being said, I have included 'data-only authors' in my own papers before because that's just how things are currently done and some papers just wouldn't happen without being submissive to 'those with data". However, I made it clear to those authors that they also had to participate in the analysis or the writing (or something else) if they wanted to be included as an author.

I think it's understandable why it happens, but I don't think that makes it ethically OK. It exploits a weird mafia-esque power structure and essentially keeps those with power in power (insert the plethora of discussion about old-white-male-bearded-faculty here…).

Another concern is about meta-analyses. Is it really reasonable for authors of meta-analyses to invite all of their data sources as authors? Moreover, as someone who works on multi-disciplinary meta-analyses, I don't know those who hold potentially important data outside my area of expertise. If the data isn't open, I probably won't ever even find it!

For (2b) YES! Exactly, we definitely need to reward open data (and software, and other contributions) more effectively! I agree, we need more tools like embargos to limit the negative impacts, I don't know what those are, but I hope someone is working on it! I know there are always some who will not reward a highly cited dataset as much as a publication, but to a certain extent we as early career researchers we are part of the system and if you are proud of something and you think a hiring committee should take it into account, it's up to you (us) to convince them in your CV or coverletter. Put those alternative metrics right in your cv (data or software downloads, citations, etc). It won't work every time, but the system won't change until it reaches a critical mass.

Margaret Kosmala
June 15, 2016 at 3:06 pm (UTC -4)

Great follow-up comment. I think we're generally in agreement. I have a lot of discomfort with the culture of authorship as it stands currently, and like you, I do as the crowd does, because that's the reality of how things work currently. I would really like to see a better way of acknowledging contributions of all sorts — something more specific than "authorship" (which with the L&O definition you provided might mean some papers have ZERO authors!) and more meaningful than being in the acknowledgments section. I think the CRediT initiative is perhaps a good place to start http://casrai.org/credit (although I wish they had a better webpage), and I'm going to give it a try in my next publication. That's still mostly just putting people in the acknowledgments section, though, until (and if) it catches on.

I think meta-analyses are a great (and difficult) example. If you are using 25 different studies, then I would argue no one of them is a substantial part of the research. But here's a related question: what about studies like NutNet where you got (still get?) authorship by contributing data? This was a grass-roots distributed network of researchers who contributed their own time and money to creating data for a worthwhile project in exchange for authorship. Right now authorship is a currency. But how else do you get potential collaborators to participate in big data networks (e.g. NutNet, AmeriFlux) — or when creating a big database (e.g. TRY traits database)? Things that seem to work: offer money (or equipment in kind) or offer time (in the form of data management and processing) or offer prestige (i.e. authorship). Are there other currencies to replace authorship if a grassroots network can't offer money and data creators aren't interested in using the data they create (so their own processed data isn't valuable to them)? I don't know the answer. I'm just throwing the question out there.

And great point about highlighting contributions other than publications in CVs and coverletters. I wonder exactly how people word these things. I'm sure there's quite some variation and I'd be curious to know what other people do. Hmmm… I think I'm going to add a 'data products and code' section to my CV. I have it on my website, but my CV is a bit more conservatively organized…

Paul Caplat

I completely agree with you on this, Margaret.
Remi, something that is important to consider re. publications: sure, having a dataset to offer often won't (and shouldn't) get you instant publications, just based on that contribution (though I know surprisingly many people who do get those). But it will create opportunities for contributing to paper in a much easier way. So it is not unethical: if I have a big dataset (it also works with a ready-to-use model, btw) and someone invites me to be on a paper, I'll help thinking about the analysis & writing the manuscript. It'll still be one more paper that wouldn't have existed without that invitation. And not much time wasted. So I think the key concept here is what Margaret refers to as "tip-the-scale sort of way"…
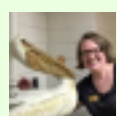
Megan N O'Donnell

So way back when, when I thought I wanted to be an artist*, I learned about a theory that says that once something becomes part of the public record the author(s) are no longer the sole "owners" of the work because the public have given it new meaning and context. You can see

this reflected in any number of ways but the easiest to understand is the meme.

Once a meme is made it takes on a life of it's own and it really doesn't matter who created it; what matters is how it's used.

I have come to view open data the same way. At some point the "who" becomes far less important then the "what." This is an idea academia isn't comfortable with since the currency of academia is based on citations (who/what/where) which have become a way to measure performance.

We often forget, that at their core, citations are audit trails for the origin points of information, and that "authorship" is merely a breadcrumb along the trail.

So, if we were able to accurately track and link how open data is reused/remixed/etc. does "authorship" still matter in the same way it matters today?

I am inclined to think not.

I think that once we can see all of the activity and influence derived from a work "authorship" losses some of it's power. Creating things will remain important but our need for control may lessen if we can easily track all of the outcomes and derivatives.

I also want to say that I completely agree with Margaret's statement of "I have a lot of discomfort with the culture of authorship as it stands currently".

[*Disclaimer: I'm a science and data librarian but my undergraduate degree is in Fine Art. It's a long story.]

Margaret Kosmala

June 17, 2016 at 9:07 pm (UTC -4)

Thanks for your fascinating perspective, Megan. I think maybe there's a tension between how things "ought to be" (according to some of us) and current scientific cultural practice. Scientists aren't used to really releasing their work to the public domain.
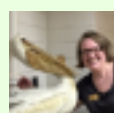
Megan N O'Donnell

June 20, 2016 at 12:57 pm (UTC -4)

Making things public can be scary – and sometimes for good reasons. There is certainly a lot of tension around these issues right now and while it can be tough it has also lead to some really positive changes. Thank you for sharing your thoughts publicly!

Jeremy Fox

June 15, 2016 at 1:32 pm (UTC -4)

Yes to this. Insofar as data sharing is expected and required of everyone, it's not something anyone should expect to receive any personal benefit from. Much as how it's expected that you won't falsify your data, and so shouldn't expect to receive any rewards from not falsifying your data. I don't know that we're all the way there yet. And I certainly wouldn't say that the traditional practice–inviting someone with data to collaborate with you–is unethical.* But that's the direction things are heading.

My own view is that this shift probably is a good thing for science as a whole, but not nearly as much of a good thing on balance as its strongest proponents claim. In large part because I don't think the overall rate of scientific progress is or ever was *that* limited by inability to find and obtain existing data. But I also think it's impossible to say with any objectivity.

For you as an individual early career scientist, I think this just means that you shouldn't rely on building your research program around people coming to you wanting to collaborate on analyses of your data and write papers with you. But honestly, I don't know that that was ever a common way to build a career in academic science. In general, if you want collaborations leading to co-authored papers, it's always been the case that you need to seek them out rather than waiting for potential collaborators to come to you. And in ecology, people who've collected datasets heavily used by others (collaborators or otherwise) mostly have built their careers on their own analyses of those datasets, at least as far as I know. And while I wouldn't presume to know your own situation better than you know it, I would tentatively hypothesize that the trend towards expecting/requiring open data would be a net benefit to the career prospects of someone with your programming skills and your broad experience and interests. Open data is a big opportunity for someone like you, I'd think, because you're better-equipped than most people to identify questions that can be addressed with existing data, find the data, clean it, and analyze it.

Quibble: I don't see why it would reflect badly on you if someone downloads your open data and does something silly with it, without consulting you. And I'd be very surprised if anyone would think that it would. I mean, eugenicists justify themselves with bad evolutionary arguments, but does anyone think that reflects badly on Charles Darwin? (Ok, a few people do, actually, but everyone else thinks those people are nutters…) There's always a risk that, if you publish something–data, an idea, an approach, whatever–somebody else will do something silly with it. I know of cases where that's happened with ideas I and my colleagues have published. But I've never heard of anyone thinking badly of me or my colleagues for this reason, and I've never worried that anyone would.

*A position I find extremely odd and confess I don't understand at all…

Margaret Kosmala

"you shouldn't rely on building your research program around people coming to you"

For sure, for sure. I'm not saying that by publishing open data, I'm dooming myself. Not at all. I'm just pointing out that I may lose out on a few mid-authored publications in favor of publication citations. Does that matter to "building my career"? I think it does, not in a big way, but in a tip-the-scale sort of way. If I have relatively few publications (which I do), another two or three is a sizable addition. The question for us early career folks is less about whether we're doing good science and more about whether we're doing it fast enough to continue monetarily supporting ourselves. Lots of us are doing good science. Not all of us can stay. To win fellowships, grants, etc. you gotta have pubs, and the number of pubs you've got to have to be competitive is continually increasing. Yes, you've got to have good science. But to stay afloat, you've got to also hit the numbers.

" I don't see why it would reflect badly on you if someone downloads your open data and does something silly with it"

This may well be an overblown fear, where negatives are so rare they're not

worth really worrying about. But I have seen others' data cherry-picked and published to suggest something they don't really show. In particular, there was a long-term dataset where authors reusing it purposefully did not use the most recent data because it conflicted with the story they wanted to tell. I wouldn't want my name connected with that sort of thing.

Carl Boettiger

Hi Margaret,

Excellent post on an important issue. One thing I wonder about: do you really think it is that helpful to have a bunch of extra papers in which you are some middle author on topics that may be only loosely tied to your main research themes? In my limited experience, such papers are seen as evidence you gather nice useful data, but are no substitute for solid first author papers reflecting a clear, strong research program. To put it slightly more cynically, middle author papers where it is clear you are just a data-author may count just as much as the recognition you gain in publishing a widely used open dataset. Or perhaps it sounds more positive to say that by making your data open, annotated and user friendly & seeing it become well known, you already gained as much as had you gotten a handful of middle author papers instead. No, we shouldn't overestimate the value of that recognition, but we shouldn't overestimate the additional value of loads of middle author papers either.

And I might just add that in today's job market, recognition for an open data hit might do more to differentiate you from the competition than even the longest lasting list of middle author papers.

Of course my knowledge on this is more limited than my bias, but I'd be curious what you and others think of this claim.

Margaret Kosmala

Hi Carl,

Thanks for your comment. To be clear, I agree with you. I don't think a bunch of middle-author papers is any way a substitute for solid first author papers. But it may matter just a bit for early career researchers to have a few more papers — especially if they don't have many to begin with. Say I have two or three first-author papers — pretty typical for a recent PhD grad. And then maybe I have one or two versus four or five middle-author papers. I think those two situations look rather different (4 vs. 7 total papers) — especially for getting things like fellowships and passing automatic cut-offs bars for review for jobs. So I see it more of a tip-the-scales sort of thing, a marginal advantage. But I also think that review committees really do slice and dice things this fine at this career stage.

And Remi also pointed out that it could be a selling point, especially if you draw attention to it in your CV or letter. I think it's a great point that you could use this to differentiate yourself. Of course, most review/search committees will tend to be made up of more established researchers who might not be as aware of the value of open data, so I think there's definitely a burden on the researcher to highlight and explain the impact of their open dataset.

Paul Caplat

Carl,

good question, but in practice how does one distinguish between middle author on some analysis-thinking-intensive paper and a data providing author? I see lots of people considered by their peers as "successful" because they have lots papers (including a majority of middle-authored ones). But hopefully there is a trend towards recognizing different contributions and stop focusing on # of papers.

Jeremy Fox

My experience is that Carl's right. Being one of many middle authors on a few more papers doesn't really make any difference to your prospects for a faculty position. Search committees at research universities want to hire people who project as future leaders in their fields. Which means you need to have a strong, coherent research program. That doesn't mean being a one-trick pony or narrowly focused–most people have research programs comprised of multiple lines of research that might only be loosely related. But each of them is a sustained line of research, not a bunch of one-off collaborations on unrelated topics to which they only made middle-author contributions.

@Margaret: re: automatic cutoffs, there are some universities that use scoring rubrics for searches (https://dynamicecology.wordpress.com/2015/09/24/guest-post-many-american-universities-use-score-sheets-to-rank-faculty-job-applicants/). But those rubrics are just a device to make sure that all competitive candidates are considered fairly from all angles, which is something all search committees try their best to do anyway. I've never experienced or heard of any search for which otherwise competitive people with <X papers just have their applications binned. And if anyone did use such a criterion, the threshold would be set very low (say, 0-2 papers), so that it was just a quick way to eliminate obviously non-competitive candidates.

Margaret Kosmala

The thing is, I'm not just talking about faculty positions. To get even that far, first I have to get (multiple) postdoc positions and/or fellowships. One also needs publications for awards and other recognition that eventually get you to the point where you're competitive.

And while I point out in my post that people are doing things with my data that are unrelated to my focus, that doesn't mean that there isn't also related research being done with the data. The same thing hold true about authorship even on related research. In fact, it's the very act of "hey, neat data, can I have it?" that often starts conversations that lead to genuine collaboration. If one can just download the data, there's less of an incentive to even start that conversation.

Again, as I've said on other comments here, I'm not saying this issue is a make-or-break one. It's certainly fairly moot for established researchers. But I do still think it can have a small effect for early career researchers at a time when small effects can matter.

Douglas in Norway
June 18, 2016 at 7:21 am (UTC -4)

An important topic! Open access to data can benefit everyone, but we need some standards and norms to help ensure fair credit to data producers. I have seen this from both sides in various guises and agree outcomes can seem unfair.

I think a suitable compromise is to require data users to sign-up and agree to offer co-authorship to data producers. The data producers then decide if they are willing and able to be a coauthor (or are rather thanked in acknowledgements). If they agree to be coauthors they then are expected to provide all the input and feedback as required from any author — this can substantially benefit the paper overall as data producers are often well/better placed to judge interpretations and implications. A sign-up system can also ensure that sensitive information is suitably controlled (e.g., locations of some rare and vulnerable species etc.).

A good example is the TEAM network … see

http://www.teamnetwork.org/data/query

Are there reasons why these sign-up and invite approaches are not more widely applied?

Margaret Kosmala
June 18, 2016 at 2:25 pm (UTC -4)

"Are there reasons why these sign-up and invite approaches are not more widely applied?"

Probably because of the overhead involved. Networks like TEAM and LTER have the resources to really think through data sharing for multiple datasets and set up websites for them. Our dataset, by contrast, was produced by a single lab without resources to do this properly. Instead we published our data in a journal specifically created to disseminate datasets. But the publication doesn't have the same sort of policy — probably because cultural norms vary among disciplines or because open data is still a fairly new thing. Maybe as various networks come up with varying policies that scientists seem comfortable with, we as a discipline will decide upon a standard best practices. Then it will be easier for single researchers and single labs to come up with policies.
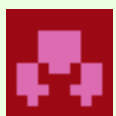
Douglas in Norway
June 18, 2016 at 2:45 pm (UTC -4)

Thanks (thought I was lost in verification there a moment so re-posted the comment at Dynamic Ecology)

I agree the norms will take time to evolve. As I say, I think the basic concepts used in TEAM have the balance about right. If you cannot manage the online systems I think a simple option is attaching a clear request for users to contact and invite data producers as authors. Hard to ignore and not unreasonable.

Bob Reed
June 28, 2016 at 3:21 am (UTC -4)

Hi Margaret, I wrote about a blog about your blog. I thought you might find it interesting. You can check it out here:

https://replicationnetwork.com/2016/06/28/bob-reed-the-problem-with-open-data-would-requiring-co-authorship-help/

[…] Kosmala on a potential downside of open data–people no longer have to contact you and invite you to collaborate in exchange for access to […]

[…] Data may be good for science, but it may be bad for scientists — especially early career ones," writes Margaret Kosmala. (Ecology […]