Moving toward a sustainable ecological science: don't let data go to waste!

Timothée Poisot, Ross Mounce, Dominique Gravel

Nov. 2012

Introduction

Claude Bernard (Bernard 1864) wrote that "art is *me*; science is *us*". This sentence has two meanings. First, the altruism of scientists is worth more to Bernard than the self-indulgence of mid-nineteenth century Parisian art scene. Second, and we will keep this one in mind, creativity and insights come from individuals, but validation and rigor are reached through collective efforts, cross-validation, and peerage. Given enough time, the conclusions reached and validated by the efforts of many will take prominence over individualities, and this (as far as Bernard is concerned), is what science is about. With the technology available to a modern scientist, one should expect that the dissolution of *me* would be accelerated, and that several scientists should be able to cast a critical eye on data, and use this collective effort to draw robust conclusions.

In molecular evolution, there exists a large number of databases (*GenBank*, *EMBL*, *SwissProt*, and many more) in which information can be retrieved. Such initiatives value (and promote) a new type of scientific research: building-on and extending the raw material of others, it is now possible to identify new phenomenon or evaluate the generality of previously studied ones. The job of scientists relying on these databases is not to *make* data, nor to *steal* them, it is rather to gather them and, most of all, look at them in a different way. This would not be possible, if not for the existence of public, free, online repositories. Depositing data in public repositories is so deeply ingrained in the culture of these disciplines that the "debate" on data sharing is non-existent. It is sadly impossible to be as enthusiastic when looking at current practices in ecology. Although there are many repositories available, their usage is entirely compulsory (*i.e.* let to the good will of authors), and there is often no way to have programmatic interactions with the data. This, in our opinion, goes a long way in explaining why there is no widespread data sharing culture among ecologists. Yet in the recent years, there has been a strong signal that some organizations are ready to invest time and money in data sharing. For example, *DataONE*

(Reichman, Jones, and Schildhauer 2011) is a large scale initiative, seeking to curate and make available observational data. We foresee that improving data sharing practices will be an important endeavor in the coming years, and the increasing awareness of the scientific community to these practices is a timely topic.

In this paper, using examples primarily taken from ecology and evolutionary biology, we will argue that improving our data sharing practices will improve both the quality of the science, and the reputation of the scientists. Although the exchange of data between groups is a widespread practice, we must be aware that it creates an intrinsic inequalities: those with good contacts have access to datasets, while other are left out. It would make sense that we collectively decide to abandon this practice, in favor of depositing data in open, free to access repositories. The recent emergence of several data-sharing platform (*DataDryad*, *figshare*), and the increase of mainstream attention they now receive, are the beginning of a disruption in the way we exchange and re-use data, that ecologists would benefit from. We illustrate how simple steps can be taken to greatly improve the current state of data sharing, and how we can encourage its practice of at different levels (Whitlock et al. 2010), and data citation, to encourage and reward sharing. Our most important point is that through sharing more data, we will increase both the quality and visibility of the science we produce. The contribution of synthesis centers, like NCEAS or NimBIOS, or NESCENT, speaks volume in support of this point, so one can only wonder how this impact would be increased if all the data collected had been made publicly available. We conclude this paper by showing that most of the technical aspects of data sharing can easily be mastered, meaning that data are ready to be liberated!

Why we ethically must

We strongly believe that data sharing is an ethical obligation for researchers. In this part, we point out the ethical aspects of data sharing, both with regards to other scientists, funding agencies, collaborators, and the civil society.

Data acquisition is (mostly) publicly funded

In contrast with other fields such as energy, or pharmaceutical research, most ecological and evolutionary research is funded through public grants or charitably-funded programs. Or in other words, most research is dependent on taxpayers. A recent HSBC report estimated that 80% of research publications across the world are funded by the public sector (Graham 2013). In some fields, most notably conservation biology, it is not uncommon for volunteers to participate in data gathering. For example, the French temporal survey of common birds (Jiguet and Julliard 2006), which resulted in 29 publications in peer-reviewed journals, is fed entirely through the work of amateur ornithologists. Given the direct (participatory) or indirect (financial, through public taxes) involvement of the public in ecological data collection, it is not surprising that some funding agencies have implemented data availability policies.

For example, *BBSRC* (UK) state that "[p]ublicly-funded research data are a public good, produced in the public interest", which "should be openly available to the maximum extent possible". They further add that "[t]he value of data often depends on timeliness[;] it is expected that timely release would generally be no later than the release through publication of the main findings". Similarly, *NERC* (UK) state that "[a]ll the environmental data held by the NERC Environmental Data Centres will normally be made openly available to any person or any organization who requests them.". Sanctions for not sharing data are also put in place, as "[t]hose funded by NERC who do not meet these requirements risk having award payments withheld or becoming ineligible for future funding from NERC". This perfectly mirrors one of the earliest drivers of the open access movement: science publications which are made possible through public investment must be made public. Publicly funded scientists, in most countries, are thus civil servants. Generating data is part of their job description, and there is no rational argument for which they should claim *property* of it (in addition to the fact that under most jurisdictions, data are not properties and cannot be copyrighted, a point we expand upon in the section on licensing issues). Claiming *paternity* of the data, as we discuss below, is a more legitimate claim than property is, but nonetheless does not prevent sharing them.

It improves reproducibility

Using journals to publish scientific information should not only serve the purpose of disseminating data analysis; it should maximize the ability of other researchers to replicate, and thus both validate and expand, results. It is arguably a perversion of the *publish-or-perish* mentality, that we think only in terms of papers. Interestingly, although editors and referees are very careful about the way the *Materials & Methods* sections of a paper are worded, it is extremely rare to receive any comment by referees about the data availability. However, some journals, including those from the Nature Publishing Group (Nature Publishing Group 2013) are now implementing policies to evaluate the quality of the data availability plan. Barring the availability of data, there is no certainty that the results can be reproduced.

This can cause problems at all steps of the life of a paper. How can a paper describing a new method be adequately reviewed if data are not available? How can you be sure that you are correctly applying a method if you can't reproduce the results? Can increased pressure to release the full dataset will help identify the (admittedly rare) cases of data falsification? The movement of *reproducible research* (see *e.g.* Mesirov 2010 for a recent perspective) advocates that a paper should be self-contained, *i.e.* be not only the text, but also the data, and the computer code to reproduce the figures. Even without going to such lengths, releasing data and computer code alongside a paper should be viewed as an ethical decision. Barnes (Barnes 2010) made the point that even though researchers are not professional programmers, computer code is good enough to be shared.

It will clarify authorship

It is well accepted that the final version of a scientific article reflects the diversity of backgrounds and scientific sensibilities of its authors (McGee 2011). Yet authorship, in the sense of deciding who gets to be listed as an author, and in which order, is still a key issue in several collaborations. Additionally, authorship deserves to be properly quantified (Tscharntke et al. 2007), to reflect the amount of work done by each contributor. Too strict rules of authorship will not award proper recognition, and rules too open will grant undue credit. To some extent, journals attempted to qualify the work of each contributor by having special sections, indicating who wrote the paper, conceived the study, or contributed data or reagents. This is far from being anecdotal, as it allows for increased accountability (Weltzin et al. 2006). By making dataset public and citeable, the contribution of data will become less and less of a criteria for authorship. Because the datasets can be cited independently from their original paper, they will also contribute to the overall scientific impact of the researcher who generated them, thus allowing to name as authors only those who analyzed the data.

Data cost money

Gathering data, either in the lab or in the field, costs money, as it requires the acquisition and maintenance of equipments and reagents, in addition to salaries. In this perspective, generating new data when existing ones are available and could bring answers to a question is a wasteful practice. So as to avoid this, we need to have an easy way to find suitable data, which require thorough indexing. The large amount of hard to access data was dubbed 'dark data' (Heidorn 2008). The fraction of data falling within this category is likely to increase. (Wicherts et al. 2006) surveyed the field of psychology, and showed that asking for the raw data often does not result in a successful data sharing outcome, even after 6 months of repeated inquiries. Authors can claim to have 'lost' the data, can be extremely slow to reply, can ignore emails, the given contact email address may be invalid and it can be difficult to find the 'current' contact address. Authors also die or retire, and sadly this can result in the loss of valuable scientific data unless it has been accessibly and discoverably archived elsewhere. Ultimately, authors can also flat out refuse to give the data. The practice of releasing data into the public domain with a CC0 waiver (best) or with minimally-restrictive licenses (some of which are explained in a later section), and associated with standards-compliant metadata, will help fight this effect. Overall, by making data easier to access, understand, and re-use, we will decrease the flow of funding going into data gathering, and thus decrease the financial pressure on labs.

Assuming that the increase of data sharing will result in enhanced recognition of the work involved in data collection and curation (which we detail later), data sharing can also be a way of valorizing "negative" results. Because the likelihood of a paper being published depends on the significance of the results it reports, the publication bias in favor of positive results is well documented across all scientific fields, and results in the accumulation of statistical bias over time (Scargle 2000). By dissociating the data from the paper, and recognizing data as a form of scientific production, it is possible to encourage the publication of "negative" results. This will allow (i) to produce research output even though the analysis is not conclusive (thus providing at least some return on investment), and (ii) to improve the planning of future experiments, because pre-existing data reporting both positive and negative outcomes will be available, thus allowing to make more informed decisions.

Which benefits it will bring us

In this section, we outline the ways in which sharing research data will benefit those who produced them, either because it will increase awareness about their research, or because it will allow others to measure their scientific production.

A proxy to your science

Datasets are an alternative means by which people can discover the research that you do. There is evidence showing that data availability improves reproducibility and adequate communication of results (Ince, Hatton, and Graham-Cumming 2012). Similarly, in some fields, releasing computer code under open source licenses (Vandewalle 2012) or sharing research data (H. A. Piwowar, Day, and Fridsma 2007) is associated with increased citation rates for your papers. Yet one of the argument often opposed by people reluctant to share their data is that they might risk loosing paternity of them. The previously cited analyses show that by *not* sharing data, we are exposed to a higher risk of our research being ignored, simply because other people cannot re-use or re-examine the data. By developing a culture of data sharing, and adequate citation of the datasets re-used, the origin of the data (and thus their paternity) will be made clear. It seems that by reserving intellectual *property* rights over data (although data cannot be considered as proprty), there are real risks of data not getting the usage it deserves, reducing scientists potential impact.

It stimulates collaboration and creativity

In our experience, releasing computer code (either scripts or full-featured packages) alongside a paper is a good way to get people to reproduce your work, and to use your results to build on (if only because it lowers the technical barrier to reproduce the approach). Some of these interactions result in collaborations, or in exchanges casting a new light on your previous work. In the same vein, releasing your data will allow people to explore new questions using them, which can potentially (i) lead them to interact with you so as to better exploit them, and (ii) show how your data can still provide valuable insights after you are done publishing them. The flow of data across research group is a promising way to increase the diversity of collaborations, which is viewed favorably by grant agencies (Lortie et al. 2012), and to a lesser extent, associated with higher citation rates (Leimu and Koricheva 2005).

It is a significant measure of your research impact

The NSF (US) Grant Proposal Guidelines for 2013 stopped referring to ‘Publications’ and instead refer to ‘Products’ (H. Piwowar 2013). This change was specifically performed to make it clear to scientists that research funders now see great value in research products, not *just* publications. Research products “include, but are not limited to, publications, data sets, software, and patents”. Thus published, shared datasets are now ‘first class research objects’ as they should be (`http://www.force11.org/white_paper`). We think this is a healthy move that will soon be copied by many research funders across the world. Modern science needs more than just publications, it needs shared data to function efficiently. By formally recognizing and encouraging applicants to put shared datasets on their CV’s and show the re-use of these datasets, the NSF is recognizing the immense and largely untapped value of data re-use. Just like publications, some datasets will be more re-used and cited than others. Thus research evaluation exercises will soon be looking to measure the impact of one’s data and software, not just publications.

How we technically can

In addition to the ethical and pragmatic arguments made above, we engage here in a more technical reflexion about how we should include data sharing early in the communication of scientific studies, so as to generate data in a format allowing their re-usability. We also briefly discuss the different licensing options.

Data representation

Except when they are deposited into large-scale databases, data usually live (in various states of dormancy) on the hard drives of researchers. These data are usually formatted in the way where they were used to produce the few figures or run statistical analyses used in the published account, which is to say mostly as a spreadsheet, or a raw text file (Akmon et al. 2011). Probably one of the most commonly used, the CSV (Comma separated values) format, is introducing significant risks for errors, notably because it lacks a formal specification (the chief problem being that the field delimiter will vary with the computer locale, and can interfere badly with the decimal separator or text characters). Yet, more robust and sharing-friendly formats exist, which should be taken advantage of as they offer an unprecedented way to organize information in a way maximizing accessibility. For example, the *JavaScript Object Notation* (JSON) (Crockford 2006) allows a context-rich representation of data, which can be based on templates (thus ensuring that several groups will present their data in the same way). Building upon this format, a working group can put together a syntax to represent a given type of ecological data, then provide JSON templates for other people to release these data. JSON templates (i) serve as a data-specification, and (ii) can validate the data, thus ensuring that no errors have been made. In addition, JSON is the *de facto* standard format in most APIs (Application Programming Interface, essentially a common, well-documented way to interact with, and re-use, a particular application or data-base). In the ecological sciences, there are now publications outlets focused only on methodological papers (e.g. *Methods in Ecology and Evolution*, and to some extent *BMC Bioinformatics*), and several other journals have sections for methodological papers. JSON parsers exist for almost all languages (notably C, Python, R, Java), which means that different applications will be able to access the shared information. Under this perspective, it is possible to build local databases. As long as they respect the specification, groups only need to share the access to these databases. A "global" access can still be achieved by wrapping all of the local data sources, through an API, as detailed in the following section.

Database linkage

An important obstacle is that maintaining a global database requires funding on a scale which is orders of magnitude higher (in terms of amount and duration) than what most grants will cover. The solution, building on an increased use of strict data specification, is to link several local databases (*e.g.* each research group can keep and take care of its own local database) through APIs (Fig. 1). In short, an API is an interface to application stored on a server, which will offer several *methods*, each returning a *reply*. For example, a *method* can be "retrieve all datasets containing species A", and the *reply* will be a list of datasets identifiers. If a particular data format is applied to more than one database, it becomes possible to query them at once. Under this perspective, the origin of the data does not matter, because the API will return them in a standardized fashion. When coupled with a data specification, this allows for seamless integration of different data sources. Each group implementing such a database can, in this situation, share the information related to data access. Instead of putting the raw data on a data sharing platform (some of which are reviewed below), the authors will give informations about the study, and informations about where the data are stored, and how to access them. Ideally, a good data exchange service will be agnostic to the location of the data. As soon as a specification is fixed, and used consistently, users can query both sharing platforms and home-grown database, as long as they know where the resource is located.

Legal issues - waivers, licenses and copyright law

Perhaps the point with which scientists will have less familiarity is the licensing or waivers under which data should be made available. Broadly speaking, a license is a text legally defining how content can be used, modified, and distributed. Fortunately, easy to understand, non-restrictive licenses exist, which are fit for scientific outputs. The most well known family of them is the *Creative Commons* (CC) set. This family of licenses arose from a need to relax the default restrictions of normal 'All Rights Reserved' copyright status, to expressly allow redistribution and re-use of content on the internet within the framework of existing copyright law (Lessig 2004). (Hrynaszkiewicz and Cockerill 2012) remind us that copyright does not apply to factual data, and so licenses should not be applied to this data. Where possible it is best to apply the Creative Commons Zero (CC0) Waiver to scientific data in most cases, to ensure that re-use is as frictionless and legally-unencumbered as possible. The CC0 waiver does not legally force citation of data when it is re-used. Nor should it. No one to our knowledge has ever sued another party for lack of academic acknowledgment of data re- use.

These matters are not policed by legal courts, but rather the social and community norms of academics and thus have no need for legal protection by copyright law. Legally enforcing even just attribution via a licensing mechanism can and does cause *real problems* that are best avoided e.g. 'attribution stacking' (Mietchen 2012), thus CC0 is recommended for most data to avoid unnecessary complications. This particular waiver is used by *Dryad* (a data repository associated with, *e.g.*, *The American Naturalist*) and *fig**share*** (though only for datasets). Where the 'data' are more artistically-expressed (a prime example is color plates of organism pictures) they are covered under copyright law, and can if desired, be licensed. An acceptable license that minimally impedes science progress is the Creative Commons Attribution (*CC BY*) license, which allows use and reproduction of the data as long as the original data is cited in the manner specified by the author(s) and not in any way that suggests that they endorse the re-use (this license is used for all non- data submissions in *figshare*). Concerns over the use of CC BY in academia have been exhaustively answered by Creative Commons recently as so many academics in the UK were confused (`http://wiki.creativecommons.org/Data`). The *Creative Commons* website offers an intuitive free tool to choose a license (`http://creativecommons.org/choose/`). We urge readers to take heed of the above, and strongly encourage scientists to be aware of the pitfalls associated with the other more restrictive license modules available when selecting one (Hagedorn et al. 2011 ; Klimpel 2012).

How it should be encouraged

The role of journals

Journals are in the best position to make things move (Vision 2010) because a scientist's career depend on getting his work published. Although a bottom-up approach should always be preferred when possible, editors have in their hand a powerful lever to modify our collective behavior. Some journals are now asking the authors to deposit their ecological data in a public repository (Fairbairn 2011 ; Whitlock et al. 2010). This is mandatory for sequences in all journals (*GenBank*), and similar mandatory archiving of all data in TreeBase, DataDryad, or FigShare is becoming a common practice. The referees are, however, rarely asked to evaluate if the adequate data are released, and even more rarely given access to the data during the evaluation process. About this last point, an increased collaboration between journals and data sharing platforms, to allow referees to anonymously access the data, should be encouraged. In practice, authors are still free to release summary statistics instead of raw data, which allows to reproduce the paper, but not to confirm the validity of the approach. There are however signals that things are changing. The *Nature* journals family will implement a more robust data sharing policy, effective in May 2013, to reduce the irreproducibility of life science papers (Nature Publishing Group 2013).

Journal-led mandates cannot be the only solution used. When compliance with journal stipulations are retrospectively checked, even clinical trials data compliance (Prayle, Hurley, and Smyth 2012) and *GenBank* archiving of data are not universally adhered to, even in the 'best' journals of highest reputation (Noor, Zimmerman, and Teeter 2006). Journals must take care that data archiving mandates are enforced and not just 'rhetoric', be it through increased editorial control, or by asking the referees to evaluate the data sharing plans. In addition, journals should implement incentives for authors to cite the datasets, and not just the paper to which they are attached. Strong limitations on the number of references can currently impede this practice, as it will force authors to choose citations. In the context of meta-analyses, this can become especially problematic. The solution of having references part of the supplementary materials is not optimal either, as it comes with no assurance that they will be registered as a citation to the dataset, and will benefit from less exposure. To this effect, having an additional reference, as it will valorize the production of data as literature items.

The role of funding agencies

In our opinion, the first step that funding agencies can take to encourage good data sharing practices is to recognize the value of data contributions. We outlined some initiatives in this sense earlier in the text. In this perspective, the fact that datasets can be attributed a DOI, which allows one to cite them, is an important step forward. This allows to measure the long- term impact of a dataset, by following how citations accumulate. Especially for early-career scientists, it is frequent that the computer code of the datasets are available long before the paper is even in press. When applying to grants or positions, whether the funding agency recognizes "non- publication" research products can make all the difference.

On the other hand, there is a need for a collective discussion between scientists and funding agencies. In addition to recognizing the value of data, should agencies *request* their availability as a condition to obtain a grant? Round-tables between ecologists and representative of funding agencies during large ecological meetings (*ESA*, *INTECOL*, *EEF*, *BES* for example) can be a productive step forward, and can help drafting recommendations which will improve our data sharing practices. However, it is important that not much coerciveness goes in these measures, as it can render some needlessly hostile to the logic of data sharing, which in our opinion would only hinder scientific advancement. Although we clearly would appreciate enforcement of data sharing policies by funders, we think that this should be accompanied by a didactic effort to make the point that there are few downsides to data sharing.

Conclusion

In the last two years, there were an important number of media outbursts, and public indignation, about the role of science and scientific conduct. They may all have been avoided if the practice of putting data publicly online was widespread. The so-called *climategate* (Jasanoff 2010) could have been largely averted if all data were made public in the earlier days of the affair, as it was later clearly demonstrated that the apparent lack of transparency eroded public trust in scientists (Leiserowitz et al. 2010 ; Ravetz 2011). Even more recently, the controversy over a study on the carcinogenicity of GM maize (Séralini et al. 2012) was thickened by the refusal of both sides (Monsanto and the French research group) to release the full data, in addition to many undisclosed conflicts of interests (Meldolesi 2012).

When journal editors publicly discussed the matter, they called this *data archiving* (Fairbairn 2011 ; Whitlock et al. 2010). We would exhort other scientists not to use this expression too much. Data *archiving* evokes cardboard boxes, in which data are put to collect some dust. Whether this happens in the hard-drive of a scientist or in a well-maintained repository only differs in the fact that the later solution comes with a DOI. We think that the process or making data available should be called in a way which reflects its objective: *data sharing*. We have the technology in place to give data a second life, in which the scientific community can appropriate them, recognize the paternity of those who generated them, and acknowledge this through citations. Data are all we care about. They make science, and especially in such data-hungry fields as ecology, possible. Sharing them ensure that people needing data to feed models, test routines, or perform meta-analyses can do that, and people contributing these data are recognized for their effort. Data bring answers to our questions, and much better, questions to our answers. After serving us so well, they deserve better than to be *archived*.

Acknowledgments: We thank Karthik Ram for offering us the opportunity to write this paper, and many people who gave feedback during the writing. This paper was developed in an open *GitHub* repository (`https://github.com/tpoisot/DataSharingPaper`), and is archived on ***figshare***. TP is a ***figshare*** advisor. TP was funded by a FQRNT-MELS post-doctoral scholarship.

Legends

References

Akmon, Dharma, Ann Zimmerman, Morgan Daniels, and Margaret Hedstrom. 2011. "The application of archival concepts to a data-intensive environment: working with scientists to understand data management and preservation needs." *Archival Science* 11 (3-4): 329–348. doi:10.1007/s10502-011-9151-4.

Barnes, Nick. 2010. "Publish your computer code: it is good enough." *Nature* 467 (7317): 753. doi:10.1038/467753a.

Bernard, C. 1864. *Introduction à l'étude de la médecine expérimentale*.

Crockford, Douglas. 2006. "The application/json Media Type for JavaScript Object Notation (JSON)." http://tools.ietf.org/html/rfc4627. http://tools.ietf.org/html/rfc4627.

Fairbairn, Daphne J. 2011. "The advent of mandatory data archiving." *Evolution* 65 (1): 1–2. doi:10.1111/j.1558-5646.2010.01182.x.

Graham, Dan. 2013. "Academic Publishing: Survey of funders supports the benign Open Access outcome priced into shares." *HSBC Global Research*: 1–36.

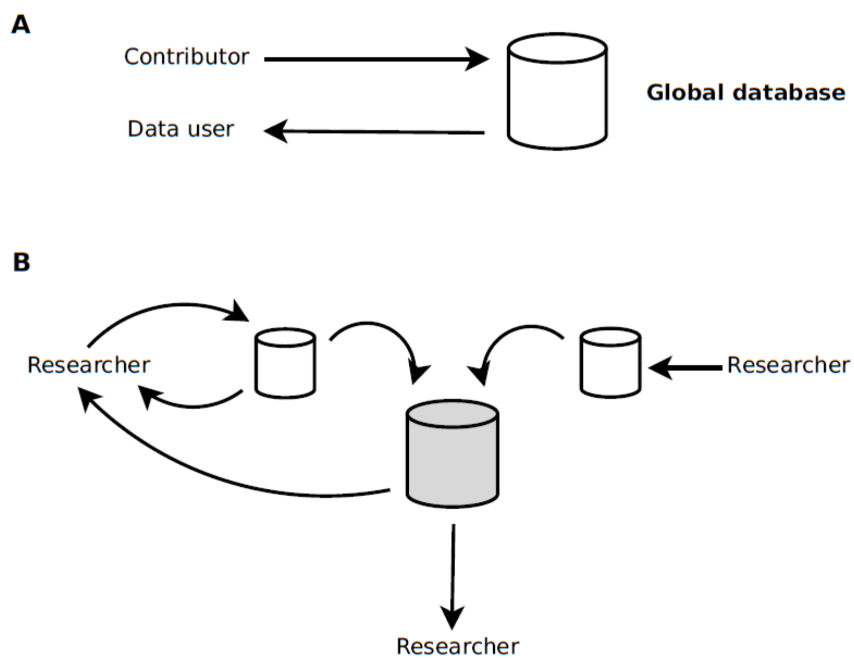

Figure 1: The differences between a large, global database (*e.g.* Genbank, **A**), and the interactions between different databases (**B**). In both diagrams, arrows represent the flow of information (*i.e.* data) between users, through databases. In the first situation, a global database centralizes all of the information. In the second situation, each group maintains its local database, with which it can interact. In addition, local databases are unified through an API (here stored on the grey server), allowing every one to access the data, including replicating them on other servers to ensure redundancy.

Hagedorn, Gregor, Daniel Mietchen, Robert Morris, Donat Agosti, Lyubomir Penev, Walter Berendsohn, and Donald Hobern. 2011. "Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information." *ZooKeys* 150 (0): 127–149.

Heidorn, P. Bryan. 2008. "Shedding Light on the Dark Data in the Long Tail of Science." *Library Trends* 57 (2): 280–299. doi:10.1353/lib.0.0036.

Hrynaszkiewicz, Iain, and Matthew Cockerill. 2012. "Open by default: a proposed copyright license and waiver agreement for open access research and data in peer-reviewed journals." *BMC Research Notes* 5 (1): 494+.

Ince, Darrel C., Leslie Hatton, and John Graham-Cumming. 2012. "The case for open computer programs." *Nature* 482 (7386): 485–488. doi:10.1038/nature10836.

Jasanoff, Sheila. 2010. "Testing time for climate science." *Science*.

Jiguet, F., and R. Julliard. 2006. "Suivi temporel des oiseaux communs. Bilan du programme STOC pour la France en 2005." *Ornithos* 13: 158–165.

Klimpel, Paul. 2012. *Free knowledge based creative commons licenses*. Berlin: Wikimedia Deutschland. http://www.vlaamse-erfgoedbibliotheek.be/sites/default/files/bron/2725/klimpel-consequences-risks-side-effects-cc-non-commercial-2012.pdf.

Leimu, Roosa, and Julia Koricheva. 2005. "Does scientific collaboration increase the impact of ecological articles?" *BioScience* 55 (5): 438–443. http://www.bioone.org/doi/abs/10.1641/0006-3568(2005)055%5B0438:DSCITI%5D2.0.CO%3B2.

Leiserowitz, Anthony, Edward W. Maibach, Connie Roser-Renouf, Nicholas Smith, and Erica Dawson. 2010. "Climategate, public opinion, and the loss of trust." *Social Science Research Network*.

Lessig, Lawrence. 2004. *Free culture: the nature and future of creativity*. New York: Penguin Press.

Lortie, Christopher J., Lonnie Aarssen, John N. Parker, and Stefano Allesina. 2012. "Good news for the people who love bad news: an analysis of the funding of the top 1% most highly cited ecologists." *Oikos* 121 (7): 1005–1008. doi:10.1111/j.1600-0706.2012.20109.x. http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0706.2012.20109.x/abstract.

McGee, Glenn. 2011. "The Ethics of Authorship: Does It Take a Village to Write a Paper?" *Science Careers*. http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2001_03_30/noDOI.2580479737297545632.

Meldolesi, Anna. 2012. "Media leaps on French study claiming GM maize carcinogenicity." *Nature Biotechnology* 30 (11): 1018.

Mesirov, J. P. 2010. "Accessible Reproducible Research." *Science* 327 (5964) (jan): 415–416. doi:10.1126/science.1179653. http://www.sciencemag.org/cgi/doi/10.1126/science.1179653.

Mietchen, Daniel. 2012. "Attribution stacking as a barrier to reuse." http://wir.okfn.org/2012/01/27/attribution-stacking-as-a-barrier-to-reuse/.

Nature Publishing Group. 2013. "Raising standards." *Nature Immunology* 14 (5): 415–415. doi:10.1038/ni.2603. http://www.nature.com/ni/journal/v14/n5/full/ni.2603.html.

Noor, Mohamed A. F., Katherine J. Zimmerman, and Katherine C. Teeter. 2006. "Data Sharing: How Much Doesn't Get Submitted to GenBank?" *PLoS Biol* 4 (7): e228+.

Piwowar, Heather. 2013. "Altmetrics: Value all research products." *Nature* 493 (7431) (jan): 159–159. doi:10.1038/493159a. http://www.nature.com/nature/journal/v493/n7431/full/493159a.html.

Piwowar, Heather A., Roger S. Day, and Douglas B. Fridsma. 2007. "Sharing detailed research data is associated with increased citation rate." *PloS one* 2 (3): e308. doi:10.1371/journal.pone.0000308.

Prayle, Andrew P., Matthew N. Hurley, and Alan R. Smyth. 2012. "Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study." *BMJ* 344.

Ravetz, J. R. 2011. "'Climategate' and the maturing of post-normal science." *Futures*.

Reichman, O. J., Matthew B. Jones, and Mark P. Schildhauer. 2011. "Challenges and opportunities of open data in ecology." *Science* 331 (6018): 703–5. doi:10.1126/science.1197962.

Scargle, Jeffrey D. 2000. "Publication bias: the 'file-drawer' problem in scientific inference." *Journal of Scientific Exploration* 14 (1): 91–106.

Séralini, Gilles-Eric, Emilie Clair, Robin Mesnage, Steeve Gress, Nicolas Defarge, Manuela Malatesta, Didier Hennequin, and Joël Spiroux de Vendômois. 2012. "Long term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize." *Food and chemical toxicology* 50 (11): 4221–31. doi:10.1016/j.fct.2012.08.005.

Tscharntke, Teja, Michael E. Hochberg, Tatyana A. Rand, Vincent H. Resh, and Jochen Krauss. 2007. "Author Sequence and Credit for Contributions in Multiauthored Publications." *PLoS Biology* 5 (1): e18. doi:10.1371/journal.pbio.0050018. http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371%2Fjournal.pbio.0050018.

Vandewalle, Patrick. 2012. "Code Sharing is Associated with Research Impact in Image Processing." *Computing in Science and Engineering*: 1–5.

Vision, Todd J. 2010. “Open Data and the Social Contract of Scientific Publishing.” *BioScience* 60 (5): 330–331. doi:10.1525/bio.2010.60.5.2.

Weltzin, Jake F., R. Travis Belote, Leigh T. Williams, Jason K. Keller, and E. Cayenne Engel. 2006. “Authorship in ecology: attribution, accountability, and responsibility.” *Frontiers in Ecology and the Environment* 4 (8): 435–441. doi:10.1890/1540-9295(2006)4[435:AIEAAA]2.0.CO;2. http://www.esajournals.org/doi/abs/10.1890/1540-9295%282006%294%5B435%3AAIEAAA%5D2.0.CO%3B2.

Whitlock, Michael C., Mark A. McPeek, Mark D. Rausher, Loren Rieseberg, and Allen J. Moore. 2010. “Data archiving.” *The American naturalist* 175 (2): 145–6. doi:10.1086/650340.

Wicherts, Jelte M., Denny Borsboom, Judith Kats, and Dylan Molenaar. 2006. “The poor availability of psychological research data for reanalysis.” *American Psychologist* 61 (7): 726–728.