Professionalization: Open Science

GIS 710, November 12, 2018

Class

Although there is an ongoing discussion about what are definitions of words such as reproducible and replicable, there is a basic agreement about what is open science and an more and more voices saying that open science, at least in the basic and perhaps vague understanding, is the way forward. Watson (2015) says that *[open] science is the future, and it will replace closed science. I encourage you to embrace it.* Open science is usually understood and defined through describing the availability (and sometimes licensing) of different parts of the research process. In many cases, however, open science or is part is described without even mentioning “open science” focusing instead on concepts such as reproducibility, replicability or reusability. Another related concept, recomputability, is especially relevant in computational and geospatial sciences, although its relation to reproducibility and replicability is not yet completely resolved. Whatever the formal definitions are, at the end, applying open science principles means working with open data repositories and open source software. Rocchini & Neteler (2012) give GBIF, DataONE, R and GRASS GIS as examples. Consequently, learning about specific tools which are enabling open science is a necessary part of learning about open science.

In the lecture portion of the class we will go over history of and motivation for open science and we will discuss some of the controversies. In the exercise portion of the class we will split into groups discussing different stands in relation to open science. For the blog assignment, you will get to describe your own position.

Open science components

Watson (2015) lists and states:

- Open methodology
    - *An open methodology is simply one which has been described in sufficient detail to allow other researchers to repeat the work and apply it elsewhere.*
- Open access
    - *Who should pay? The funder, of course; and when there is no funder, or there are no funds, then there are preprint servers ... and institutional repositories.*
- Open data
    - *Scientists who hoard data, far from pushing back the boundaries of human knowledge, instead act as barriers to discovery.*
- Open source

    - *[If] you use software as part of the scientific method, then the source code should be available to read (preferably via a website such as GitHub or SourceForge), the software should compile and run and there should be a description of the core algorithms.*
- Open peer review
    - *...is about making peer review a collaborative process between authors and reviewers.*
- Open education
    - *...refers to the open and free availability of educational resources.*

The “re” words

Ince et al. (2012) distinguishes direct and indirect reproducibility in computational science and says: *Direct reproducibility refers to the recompilation and rerunning of the code on, say, a different combination of hardware and systems software, to detect the sort of numerical computation and interpretation problems found in programming languages… Indirect reproducibility refers to independent efforts to validate something other than the entire code package, for example a subset of equations or a particular code module.* However, many different authors defined these or similar ways in other contexts or even in the same context. Here is one possible list of “re” words and set of definitions:

- replicability: independent validation of specific findings
- repeatability: same conditions, people, instruments, ... (test–retest reliability)
- recomputability: same results in computational research
- reproducibility: obtain again same results from the raw data
- reusability: use again the same data or methods

Open Source Software and Open Science

Software plays a major role in open science especial in the geospatial field where software is almost always involved as a dependency, something unchanged and simply used as is in the research, or it is involved as new code created to apply a new method, perform computational experiment, or to address data processing issue. Ince et al. (2012) say that *unambiguous descriptions are no guarantee of reproducibility* and further continue *[as] a result, to maximize the chances of reproducibility and consistency, not only would we urge code release, but also a description of the hardware and software environment in which the program was executed and developed.* Rocchini & Neteler (2012) argue for employing open source software: *If FOSS were more widely employed in ecology and the code used in data analysis provided in scientific papers, more researchers would be able to rely on and replicate peer-reviewed functions.*

The Open Source Definition

Open source doesn't just mean access to the source code. The distribution terms of open-source software must comply with the following criteria:

1. The license shall not restrict any party from selling or giving away the software…
2. The program must include source code…
3. The license must allow modifications and derived works…
4. …The license must explicitly permit distribution of software built from modified source…
5. The license must not discriminate against any person or group of persons.
6. No Discrimination Against Fields of Endeavor
7. The rights … must apply to all to whom the program is redistributed…
8. License Must Not Be Specific to a Product
9. License Must Not Restrict Other Software
10. License Must Be Technology-Neutral

(Source: https://opensource.org/osd, shortened)

Free Software Definition

A program is free software if the program's users have the four essential freedoms:

- The freedom to run the program as you wish, for any purpose.
- The freedom to study how the program works, and change it.
- The freedom to redistribute copies so you can help others.
- The freedom to distribute copies of your modified versions to others.

(Source: https://www.gnu.org/philosophy/free-sw.en.html, shortened)

The Open Definition

It can be summed up in the statement that: Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness). (Source: http://opendefinition.org/, full definition available ibid)

Class Exercise

Discuss open science and science in general in 3 groups, each group one of the following topics. Then get back together, summarize your argument and discuss with the whole group.

- In favor of open science: Why is open science necessary for science? Why is open science good? (...)
- Opposed to open science: What are the reasons not to do open science? Why is open science bad? (...)
- Middle way: Is there something in between? Is something like partial open science is sustainable? How to implement something in between to satisfy both groups or something transitional? (...)

If you are not comfortable defending the particular position of the group you are in, you can just formulate your thoughts as “scientists who hold this position argue that…”

Blog Assignment

Write a blog post in the “And, But, Therefore” format about where you stand in terms of open science or about where you want to be and what is stopping you. 250 words. Due ...

References

- Watson, M. (2015). When will ‘open science’ become simply ‘science’?. Genome biology, 16(1), 101. https://doi.org/10.1186/s13059-015-0669-2
- Ince, Darrel C., Leslie Hatton, and John Graham-Cumming. “The case for open computer programs”. In: Nature 482.7386 (2012), pp. 485–488. http://doi.org/10.1038/nature10836
- Rocchini, Duccio and Markus Neteler. “Let the four freedoms paradigm apply to ecology”. In: Trends in Ecology and Evolution (2012). http://doi.org/10.1016/j.tree.2012.03.009

Additional resources

- Morin, A et al. “Shining light into black boxes”. In: Science 336.6078 (2012), pp. 159–160. http://doi.org/10.1126/science.1218263
- Piwowar, Heather A., Roger S. Day, and Douglas B. Fridsma. “Sharing Detailed Research Data Is Associated with Increased Citation Rate”. In: PLOS ONE 2.3 (Mar. 2007). http://doi.org/10.1371/journal.pone.0000308
- Lees, Jonathan M. “Open and free: Software and scientific reproducibility”. In: Seismological Research Letters 83.5 (2012), pp. 751–752. http://doi.org/10.1785/0220120091
- Marwick, Ben. “Computational reproducibility in archaeological research: basic principles and a case study of their implementation”. In: Journal of Archaeological Method and Theory 24.2 (2017), pp. 424–450. http://doi.org/10.1007/s10816-015-9272-9
- Kieran Healy. Plain Person’s Guide to Plain Text Social Science. (by Associate Professor of Sociology at Duke University). Website: http://plain-text.co (PDF: http://kieranhealy.org/files/papers/plain-person-text.pdf)
- Eva Amsen. 2014. What is open science? F1000Research blog. https://blog.f1000.com/2014/11/11/what-is-open-science/

Summer distance education class

GIS 595-601 Special Topics in Geospatial Information Science
Tools for Open Geospatial Science
https://ncsu-geoforall-lab.github.io/open-science-course