



Pekka Nygren

Advancing the Everyone's Right to Forest Science: opening research data and computational codes in Silva Fennica

Nygren P. (2024). Advancing the Everyone's Right to Forest Science: opening research data and computational codes in Silva Fennica. Silva Fennica vol. 58 no. 1 article id 24016. 3 p. <https://doi.org/10.14214/sf.24016>

Research data and computer codes used for analysing them are important products of scientific research. Open data make the full evaluation of research results possible. Only open data make a study really reproducible. Data collected for solving a particular research problem may gain new value over time e.g., as a part of a time series, as a comparison point for analysing changes or as a part of the cultural heritage. New analytical methods may reveal phenomena that the original data analyses missed. Accumulation of data on a particular field like genomics makes global reanalyses possible. Open data offer also new cooperation possibilities for scientists, research groups and organisations. All this calls for data openness and reliable description, or metadata, of it.

Finnish Society of Forest Science, publisher of Silva Fennica, has been an early adopter of the open science. Whole text articles of Silva Fennica and its sister journal published in Finnish, *Metsätieteen aikakauskirja*, have been openly available over the Internet since 1998. Silva Fennica has been always strict on clear and reproducible description of methods. Since 2021, we have applied the Openness and transparency guidelines of the Center for Open Science at level 1, which requires clear articulation of methods, citation standards and availability of research materials in the author instructions (Nygren 2020). We also started to request depositing all original genomic data to one of the *International Nucleotide Sequence Database Collaboration* repositories as a condition for publication in Silva Fennica. Given the benefits of open data and codes, Silva Fennica and *Metsätieteen aikakauskirja* are requesting opening of research data and codes for all manuscripts submitted on or after 1 February this year.

Research data refer to the data that are generated and/or utilised in the research you report in your manuscript. They may consist of data from several primary sources. Opening of them requires a few clear steps:

1. Check that your research data file(s) is clear and save it in an archivable, machine-readable format. We recommend CSV format for alphanumeric data, txt for text data and tiff for graphics.
2. Prepare rich metadata on your data. Instructions on the contents of metadata can be found at <https://silvafennica.fi/page/metadata>.
3. Deposit your research data and the metadata in a trusted repository. The chapter 3.6 of

Silva Fennica's author instructions gives basic guidance on selecting repositories. The repository must give a persistent identifier (e.g., DOI or URN) and a Creative Commons licence for your data.

Now your data are open following the **FAIR principles** i.e., they are Findable, Accessible, Interoperable and Reusable.

We recognise that the raw data, or the data in their original form, as they were obtained directly from the source and before they were processed in any way, may be used for producing several research data sets. Thus, we encourage but do not request opening of the raw data. However, the metadata of the raw data must be deposited to the same repository as your research data. The same instructions apply as for the research data.

Several computational codes may be used for producing a manuscript i.e., for processing raw data to research data, statistical analyses of the research data and making the manuscript graphs from the research data. In a modelling study, the code may be the principal output of the research. In principle, all codes should be open. Silva Fennica's **author instructions** provide detailed guidance for opening codes.

There may be acceptable reasons for not opening your data. These include national security issues, confidential business information, personally sensitive data or exact location of threatened organisms or minor nature types. In these cases, anonymization or curation of the data for removing sensitive information is the first option. Only if the curation is not possible or very laborious in comparison to the benefit, you may ask an exception of the data openness policy of Silva Fennica. Please, note that some repositories, like the **Finnish Social Science Data Archive**, offer curation services for scientists depositing data.

If you use commercially available data (e.g., remote sensing data) or an official data base (e.g., a national forest inventory data or the forest property data of the Finnish Forest Centre) you may not have the right to open your research data. However, you must open the metadata that fully identify the data you used and information how to request access. Potential future use or patenting are not acceptable reasons for not opening your data. However, you may apply for one-year embargo in these cases. Your metadata must be always immediately open.

The main motivation for requesting data and code openness is the enhancement of research quality. The so-called reproducibility crisis draw attention in the mid-2010s after a low rate of reproducibility of original research was reported in several fields of study like medicine (Kaiser 2015) and psychology (Open Science Collaboration 2015). Main reasons for the low reproducibility are probably lack of transparency and openness. Adopting robust research practices, like preregistration of methods, open data and codes and transparent writing were proven to increase substantially the reproducibility of quantitative psychology studies (Protzko et al. 2023). No information on reproducibility or writing practices in forest sciences are available. However, sloppy writing has been shown to be very common in ecology (Frazer 2018). The experience in psychology shows that open data and codes together with rigorous writing help to improve the situation.

Some of the special characteristics of forest research call for improved openness. Forest research often takes years to complete and field conditions vary from a year to another. Thus, the research process is difficult to repeat – it is often inherently non-reproducible. This makes data openness even more valuable than in a research conducted under controlled conditions. Several years of field data are themselves a very valuable product of science. Opening them makes reanalyses possible for detecting the data quality. As the data are hard to measure or gather, they are valuable points in a time series. They may even be reused in a completely new study, either as they are or as a part of a larger data set. This opens new cooperation opportunities for the researchers who produced the data. May someone use the data without cooperating with the original producers of the data? Yes, that may happen. However, the experience of the forest scientists who open data is

that the data reuse happens mostly via cooperation. A scientist may also ask themselves: If I do not use anymore my data, is it better to hide them or let others use them?

Today, geneticists consider obsolete a research if genomic data are not open. We in Silva Fennica sincerely hope that open data will be as obvious in forest sciences in a near future.

References

- Fraser H, Parker T, Nakagawa S, Barnett A, Fidler F (2018) Questionable research practices in ecology and evolution. PLoS One 13, article id e0200303. <https://doi.org/10.1371/journal.pone.0200303>.
- Kaiser J (2015) The cancer test – a nonprofit's effort to replicate 50 top cancer papers is shaking up labs. Science 348: 1411–1413. <https://doi.org/10.1126/science.348.6242.1411>.
- Nygren P (2020) Improvement of reproducibility and research transparency in Silva Fennica. Silva Fenn 54, article id 10411. <https://doi.org/10.14214/sf.10411>.
- Open Science Collaboration (2015) Estimating the reproducibility of psychological science. Science 349, article id aac4716. <https://doi.org/10.1126/science.aac4716>.
- Protzko J, Krosnick J, Nelson L, Nosek BA, Axt J, Berent M, Buttrick N, DeBell M, Ebersole CR, Lundmark S, MacInnis B, O'Donnell M, Perfecto H, Pustejovsky JE, Roeder SS, Walleczek J, Schooler JW (2023) High replicability of newly discovered social-behavioural findings is achievable. Nat Hum Behav 8: 311–319. <https://doi.org/10.1038/s41562-023-01749-9>.

Pekka Nygren

Managing Editor

Journals of the Finnish Society of Forest Science