

In the age of open science, repurposing and reproducing research pose their own challenges

May 12 2014

Growing numbers of researchers are making the data and software underlying their publications freely available online, largely in response to data sharing policies at journals and funding agencies. But in the age of open science, improving access is one thing, repurposing and reproducing research is another. In a study in the *Journal of Ecology*, a team of researchers experienced this firsthand when they tried to answer a seemingly simple question: what percentage of plants in the world are woody?

They thought the answer would be easy to find. After all, scientists have been distinguishing between woody and [herbaceous plants](#) for over 2000 years, ever since Plato's student Theophrastus—often considered the "father of botany"—made the distinction in 300 BC. Researchers already know when the first [woody plants](#) came to be, how wood develops and decomposes, and that woody plants like trees and shrubs evolve slower than herbs.

"We thought that if we just dug through the literature enough we would find the answer," said co-author Will Cornwell of the University of New South Wales.

But online searches weren't much help. Google didn't have the answer. Bibliographic tools like Web of Science didn't offer any clues, either.

Expert opinion didn't get them any closer. An informal survey of nearly 300 researchers from 29 countries revealed little consensus even among trained scientists, with guesstimates ranging from 1% to 90%.

"[Surprisingly] it didn't matter how much research experience they had, or how familiar they were with plants," said co-author Matt Pennell of the University of Idaho, who was a graduate fellow at NESCent at the time of the study.

Thankfully public data were available. Before they could turn to existing databases, however, they had to deal with an additional problem: Even the largest plant trait database to date—a global woodiness database containing nearly 50,000 species—contains less than 20% of the more than 300,000 plant species known to science. Simply calculating the fraction of species in the database that are woody gave misleading results, due to missing data and sampling bias towards economically important or temperate species.

By applying statistical tricks to account for sampling bias, the researchers were able to determine that between 45 - 48%, or just under half, of the world's plants are woody. "[The take home lesson is that] all big databases are biased, but by acknowledging that bias is universal and accounting for it we can make better use of them," said co-author Rich FitzJohn of Macquarie University

The researchers learned another lesson when they published their work. Their goal was to make enough information about their methods available such that other researchers could retrace their steps. Could someone—using the same data and code, but a different computer—get the same or similar results?

In an ideal world, reproducing the analyses should be as simple as installing the necessary software, downloading the data and hitting 'run.' But software changes from one version to the next. Analysis standards

evolve. Analyses that run on one machine don't always work on another.

Making a study easily reproducible, they found, requires a significant amount of time and technical skill. They made sure that everything needed to download and manipulate the [data](#) and even create the figures, was written into the code, and explained the thinking behind each snippet of code. They also provided links to tools that would enable researchers to compare changes between different versions of software and restore and run previous versions if need be.

"Nobody denies that researchers should try to make their work reproducible so that others can check their results, but actually making that feasible is easier said than done," FitzJohn said.

More information: FitzJohn, R., et al. (2014). "How much of the world is woody?" *Journal of Ecology*.
[dx.doi.org/10.1111/1365-2745.12260](https://doi.org/10.1111/1365-2745.12260)

Provided by National Evolutionary Synthesis Center (NESCent)

Citation: In the age of open science, repurposing and reproducing research pose their own challenges (2014, May 12) retrieved 23 June 2024 from <https://phys.org/news/2014-05-age-science-repurposing-pose.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.