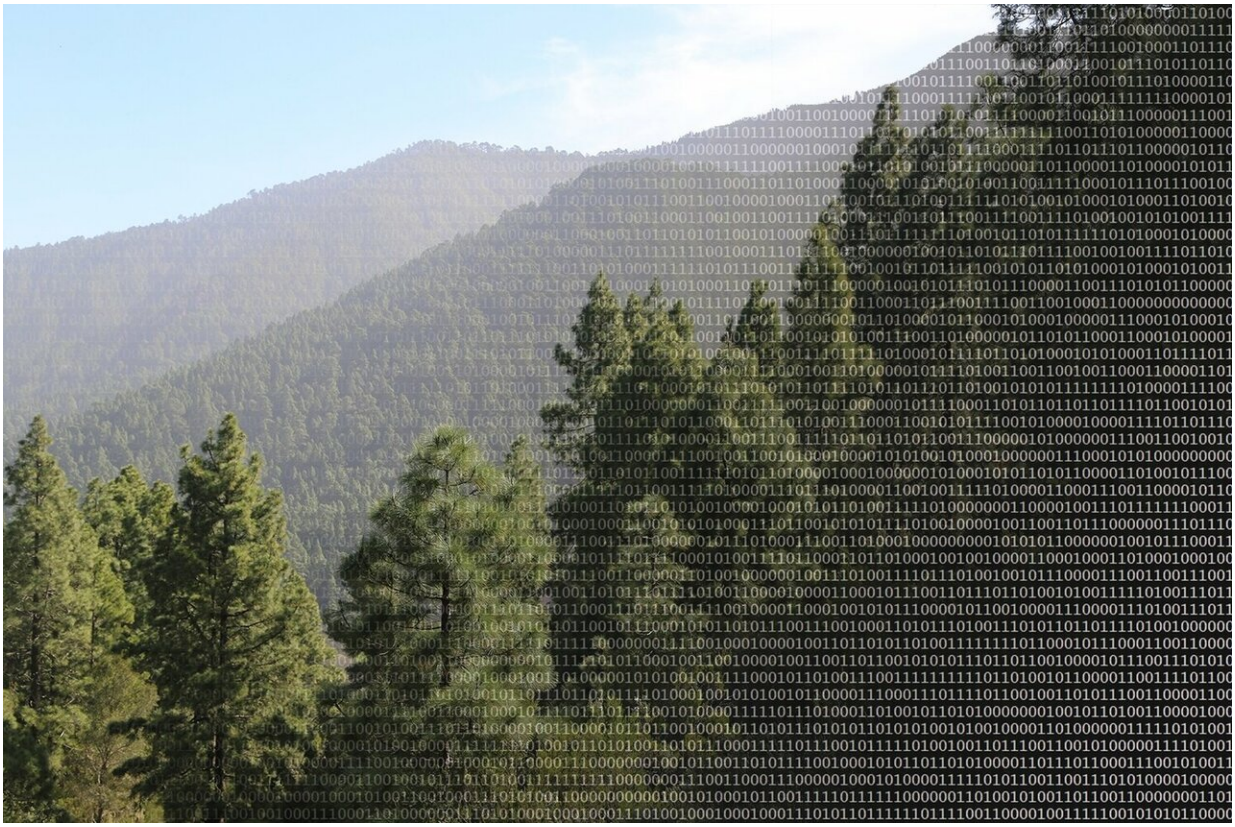
What 'Big Data' reveals about the diversity of species

March 29 2019



Big data is playing an increasingly critical role in biodiversity research. Göttingen researchers are now showing how this potential can be best utilised. Credit: Holger Kreft

Big data and large-scale analyses are critical for biodiversity research to

find out how animal and plant species are distributed worldwide and how ecosystems function. The necessary data may come from many sources: museum collections, biological literature, and local databases. Researchers at the University of Göttingen have investigated how this wealth of knowledge can best be integrated so that it can be transported into the digital age and used for research. The results were published in the journal *PLOS Biology*.

Biodiversity data can describe different aspects of life, for example the geographical distribution of organisms, their ancestry, ecological characteristics or interactions with their environment. In addition, each of these aspects can be described by different data types. For example, some researchers characterise the range of distribution of a species using individual occurrences, others use systematic counts or checklists of regional species. The individual data types not only differ in their availability, but can also have a significant influence on the scientific results and conclusions. The authors advocate for biodiversity research to use as many concurrent data sources as possible to better understand the complex interactions in nature, especially considering the increasing threat to species and ecosystems worldwide.

"The resolution of the data is crucial for the accuracy and reliability of studies on biodiversity," says the first author Dr. Christian König from the Department of Biodiversity, Macroecology and Biogeography. The fundamental compromise here is that the more detailed the data, the lower its availability and representativeness on a global scale—and often the data gaps are particularly large where species diversity is particularly high. The researchers prove this relationship by means of two case studies in which they model global patterns in the growth habit and seed size of plant species using different data types. "Our biodiversity models will improve if we link all the available biodiversity data together and make the best possible use of them, but there has been a lack of conceptual understanding of their basic characteristics and common

synergies," adds Professor Holger Kreft, Head of Department.

More information: Christian König et al, Biodiversity data integration—The significance of data resolution and domain, *PLOS Biology* (2019). DOI: 10.1371/journal.pbio.3000183