Annals of Forest Science (2014) 71:523–525
DOI 10.1007/s13595-014-0375-3

EDITORIAL

The open data debate: a need for accessible and shared data in forest science

Bruno Fady · Alain Benard · Christian Pichot · Marianne Peiffer · Jean Michel Leban · Erwin Dreyer

Received: 22 March 2014 / Accepted: 27 March 2014 / Published online: 6 May 2014

During the last decade, large data sets have been increasingly used to address key questions in the field of forest science, including: (1) the impact of climate change on productivity and species distribution; (2) the long-term course of carbon, water, and nutrient cycles; (3) the spread and virulence of pathogens; (4) the genetic basis of local adaptation; and (5) sustainable socio-economic strategies (Rehfeldt et al. 2001, 2002; Diaz-Balteiro and Romero 2008; Cappa et al. 2012; Benito-Garzón et al. 2013; Porth et al. 2013; Stephenson et al. 2014). Sound data are difficult to produce in forest science because trees are long lived, are elements of complex ecosystems, and are not easily amenable to simple experiments. Yet, foresters have observed, monitored, and measured trees and forest ecosystems for a very long time, producing impressive data sets. International provenance tests are carefully monitored since the early twentieth centrury (Rehfeldt et al. 2001, 2002); they compare in common gardens trees from seeds collected in different localities (provenances) in order to record the genetic diversity of traits of importance for forestry and adaptation. Similarly, long-term records are now available for the carbon budget, water use, and nutrient cycling of a large number of forest ecosystems in temperate, boreal, as well as tropical forests (Luyssaert et al. 2007).

However, like in other fields of research, the fate of these data remains, in many cases, uncertain, which has certainly detrimental effects for the advancement of forest science and for the improvement of forest ecosystem management. In the best of cases, they were published (usually not in the form of raw data) along with companion articles discussing the results. In most cases, they are stored under heterogeneous formats in the personal files of researchers and risk disappearing when these researchers change interests or retire. Recent European and international projects have taken this concern very seriously and have initiated the construction of large metadata and databases (e.g., TreeBreedex and Evoltree for genetic data, European Fluxes Database Cluster like ICOS, Carbo-Extreme, GHG-Europe, InGOS, for long-term ecological monitoring). International networks that monitor functional and morphological changes in forests are pushing in this direction. Institutes and research departments simultaneously consolidate available data into standardized and interconnected databases. Such a wealth of data requires very specific database management and data sharing procedures (Michener and Jones 2012).

In addition to producing increasingly large data sets from monitoring and automated machine collections, forest science also relies on a large number of short-term experiments whose main results are published in scientific journals, usually without providing the corresponding data that are sometimes lost after publication. There is a general feeling that such data are often under-analyzed by their authors and that they should be made available for re-use through synthesis and analysis to generate novel ideas and test theories at unprecedented scales.

Handling Editor: Erwin Dreyer

B. Fady · C. Pichot
INRA, UR 629, Ecologie des Forêts Méditerranéennes,
Domaine St Paul, Site Agroparc, 84914 Avignon, France

B. Fady
FRB, Centre de Synthèse et d'Analyse de la Biodiversité (CESAB),
Technopôle de l'Environnement Arbois-Méditerranée,
Aix en Provence, France

A. Benard · E. Dreyer (✉)
INRA, UMR 1137 Ecologie et Ecophysiologie Forestières,
54280 Champenoux, France
e-mail: dreyer@nancy.inra.fr

A. Benard · E. Dreyer
Université de Lorraine, UMR 1137, Ecologie et Ecophysiologie
Forestières, Faculté des Sciences et Techniques, 54500 Vandoeuvre,
France

M. Peiffer · J. M. Leban · E. Dreyer
INRA, Annals of Forest Science, Editorial Bureau,
54280 Champenoux, France

Given these challenges, Annals of Forest Science reached two important decisions: (1) incite all authors of accepted papers to provide an access to the primary data that enabled them to reach the conclusions described in the paper, and (2) launch a new category of papers, called "data-papers," devoted to the publication of valuable data sets in the field of forest science, from evolutionary to functional ecology, from local to landscape and region-wide monitoring and experiments, and from abiotic to biotic processes.

By valuable data sets, we mean the following: (1) data sets where metadata are clearly described using internationally recognized standards for metadata, where reasons why the data were collected are explained and where the potentially far-reaching interest of the data (both for empirical and theoretical work) are presented, and (ii) data sets with a scientific or practical relevance as assessed by the per-review process. We are providing (https://metadata-afs.nancy.inra.fr/ressources) a framework for the presentation of metadata under the form of a spreadsheet to be completed by the authors of the data set. The metadata will be maintained by the editorial board of AFS and made publicly available in a specific website linked to the papers and to the actual database. The latter will remain under the control of the authors, who should either maintain it on a server with precise and publicly known access rules or deposit it into a public database repository such as Dryad (http://datadryad.org/) or the KNB repository (Knowledge Network for Biocomplexity) (https://knb.ecoinformatics.org/).

Why launch this new section in a forest science journal? Valuable data sets for science should not remain unknown, hidden, or under very restricted access! This is also true for forest science, and we hope to convince researchers and experiment managers to make their data sets available to the research community for further analysis and valuation. We believe this will be an appropriate process for this essential part of science to be recognized, used, and cited by others, more focused on theoretical problems and on models and in need of sound data. Many fields of research in forest science lack data to back theory. For instance, the fate of marginal populations in terms of adaptation to changing environments has been theorized (Chevin and Lande 2011) but remains to be tested in natural environments. What governs local adaptation and genetic diversity in terms of gene flow, selection, and phenotypic plasticity has been addressed from a theoretical perspective (Le Corre and Kremer 2003) but remains to be challenged by data. In the field of forest ecology, the conditions under which niche vs. neutral processes govern community patterns (Rosindell et al. 2010) also still need to be addressed.

In the field of genetics and genomics, data resulting from sequencing (but not genotyping!) are now stored in databases. In fact, storing DNA sequences with an appropriate accession number from such repositories as GenBank or NCBI is a prerequisite for publication in most journals in the field of genetics and molecular ecology. The tide is turning for phenotypic and ecological data as well, and several leading publications require that phenotypic and ecological data be stored in public repositories, the leading one being Dryad (Rausher et al. 2010). Phenotypic data as well as ecological (biotic and abiotic) data are now crucial for linking evolutionary and functional processes and need to be available. Collecting phenotypic and ecological data in a sound way is not trivial, and in Annals of Forest Science, we wish to recognize that even if a valuable data set may not have given rise to a number of hypothetico-deductive publications, it nonetheless deserves recognition as a useful contribution to the scientific debate. In the field of genetics, again, there is currently a widespread recognition that data-driven science can lead to unforeseen discoveries and new scientific avenues. Annals of Forest Science wishes to embark on this course with forest science data: from question-driven science to data-driven science, and back!

Annals of Forest Science is joining a growing number of journals in ecology that recognize the need to make data available to the community. We hope that many authors in the field of forest sciences will submit their data-papers for review and publication in the new "data-papers" section of Annals of Forest Science. We believe that making data available to the community is a process that will benefit all members of the community, and we join the global movement with great anticipation. We list on the homepage of the AFS website the recommendations to authors for this new type of publication.

In addition to this new data-paper section in the journal, we are now urging our authors to submit their manuscripts with data archived in accessible databases committed to long-term storage and with metadata providing the key information to understand the structure of the data. Such manuscripts will receive increased attention from the editorial board. We will gradually incite all authors to provide a clear description of the data sets that back the presented demonstrations as well as the metadata required to understand the structure of the data sets and a reliable access to them. We do believe that in a few years, this requirement will become a standard practice for publishing in forest research as well as in other fields.

To incite and help authors in the process of providing data sets following international standards and ensure customized access to data to potential users, we provide a template for describing the metadata associated to a database. Completed metadata files will be hosted in a repository managed by Annals of Forest Science (https://metadata-afs.nancy.inra.fr/geonetwork), while the actual data set will be made available by the authors at their convenience under a number of conditions: guaranteed access during 5 years at least,

maintenance of the access under clarified authorship and citation rules, and identification of the data set with a Digital Object Identifier or any other perennial and secure identification system. This change fits into the new editorial policy described in an earlier editorial (Dreyer et al. 2014). We hope this new move of Annals of Forest Science will contribute to the data-sharing process in forest science and research, and contribute to new ideas, new theories, and the development of forest science as a major component of ecosystems sciences and research.

References

Benito-Garzón M, Ruiz-Benito P, Zavala MA (2013) Interspecific differences in tree growth and mortality responses to environmental drivers determine potential species distributional limits in Iberian forests. Glob Ecol Biogeogr 22:1141–1151. doi:10.1111/geb.12075

Cappa E, Yanchuk A, Cartwright C (2012) Bayesian inference for multi-environment spatial individual-tree models with additive and full-sib family genetic effects for large forest genetic trials. Ann For Sci 69:627–640. doi:10.1007/s13595-011-0179-7

Chevin LM, Lande R (2011) Adaptation to marginal habitats by evolution of increased phenotypic plasticity. J Evol Biol 24:1462–1476. doi:10.1111/j.1420-9101.2011.02279.x

Diaz-Balteiro L, Romero C (2008) Making forestry decisions with multiple criteria: a review and an assessment. For Ecol Manag 255:3222–3241. doi:10.1016/j.foreco.2008.01.038

Dreyer E, Peiffer M, Bontemps J-D, Leban J-M (2014) Editorial: Annals of Forest Science changes its scope and complies with green open access rules. Ann For Sci doi:10.1007/s13595-014-0370-8

Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. Genetics 164:1205–1219

Luyssaert S, Inglima I, Jung M, Richardson AD, Reichstein M, Papale D, Piao SL, Schulze ED, Wingate L, Matteucci G, Aragao L, Aubinet M, Beer C, Bernhofer C, Black KG, Bonal D, Bonnefond JM, Chambers J, Ciais P, Cook B, Davis KJ, Dolman AJ, Gielen B, Goulden M, Grace J, Granier A, Grelle A, Griffis T, GrÜNwald T, Guidolotti G, Hanson PJ, Harding R, Hollinger DY, Hutyra LR, Kolari P, Kruijt B, Kutsch W, Lagergren F, Laurila T, Law BE, Le Maire G, Lindroth A, Loustau D, Malhi Y, Mateus J, Migliavacca M, Misson L, Montagnani L, Moncrieff J, Moors E, Munger JW, Nikinmaa E, Ollinger SV, Pita G, Rebmann C, Roupsard O, Saigusa N, Sanz MJ, Seufert G, Sierra C, Smith ML, Tang J, Valentini R, Vesala T, Janssens IA (2007) CO_2 balance of boreal, temperate, and tropical forests derived from a global database. Glob Chang Biol 13:2509–2537. doi:10.1111/j.1365-2486.2007.01439.x

Michener WK, Jones MB (2012) Ecoinformatics: supporting ecology as a data-intensive science. Trends Ecol Evol 27:85–93. doi:10.1016/j.tree.2011.11.016

Porth I, Klapšte J, Skyba O, Hannemann J, McKown AD, Guy RD, DiFazio SP, Muchero W, Ranjan P, Tuskan GA, Friedmann MC, Ehlting J, Cronk QCB, El-Kassaby YA, Douglas CJ, Mansfield SD (2013) Genome-wide association mapping for wood characteristics in Populus identifies an array of candidate single nucleotide polymorphisms. New Phytol 200:710–726. doi:10.1111/nph.12422

Rausher MD, McPeek MA, Moore AJ, Rieseberg L, Whitlock MC (2010) Data archiving. Evolution 64:603–604. doi:10.1111/j.1558-5646.2009.00940.x

Rehfeldt G, Wykoff W, Ying C (2001) Physiologic plasticity, evolution, and impacts of a changing climate on *Pinus contorta*. Clim Chang 50:355

Rehfeldt GE, Tchebakova NM, Parfenova YI, Wykoff WR, Kuzmina NA, Milyutin LI (2002) Intraspecific responses to climate in *Pinus sylvestris*. Glob Chang Biol 8:912–929. doi:10.1046/j.1365-2486.2002.00516.x

Rosindell J, Cornell SJ, Hubbell SP, Etienne RS (2010) Protracted speciation revitalizes the neutral theory of biodiversity. Ecol Lett 13:716–727. doi:10.1111/j.1461-0248.2010.01463.x

Stephenson NL, Das AJ, Condit R, Russo SE, Baker PJ, Beckman NG, Coomes DA, Lines ER, Morris WK, Ruger N, Alvarez E, Blundo C, Bunyavejchewin S, Chuyong G, Davies SJ, Duque A, Ewango CN, Flores O, Franklin JF, Grau HR, Hao Z, Harmon ME, Hubbell SP, Kenfack D, Lin Y, Makana JR, Malizia A, Malizia LR, Pabst RJ, Pongpattananurak N, Su SH, Sun IF, Tan S, Thomas D, van Mantgem PJ, Wang X, Wiser SK, Zavala MA (2014) Rate of tree carbon accumulation increases continuously with tree size. Nature. doi:10.1038/nature12914, Advance online publication