

PLUTo: Phyloinformatic Literature Unlocking Tools

Ross Mounce ¹, Peter Murray-Rust ², Matthew Wills ¹

¹ Department of Biology & Biochemistry, University of Bath, England, United Kingdom

² [Shuttleworth Foundation Fellow](#)

Contact Email: ross.mounce@gmail.com
Homepage: <https://bitbucket.org/petermr/pluto/wiki/Home/>
Repositories: <https://bitbucket.org/petermr/crawlerrepo/>
<https://bitbucket.org/petermr/svg2xml-dev>
<https://bitbucket.org/petermr/imageanalysis>
<https://bitbucket.org/petermr/svgbuilder>
<https://bitbucket.org/petermr/xhtml2stm-dev/>

License: [Apache License 2.0](#)

Full Abstract

Approximately 4% of published phylogenetic analyses make their underlying data & results publicly available in an immediately re-usable, machine-readable form [1]. Furthermore, if one makes the effort to email the authors for the underlying data; only 16% of such requests are successful [2]. Phylogenetic data can be and *is* re-used in a multitude of different ways by other projects subsequent to the publication of the original analysis – it is well understood, valuable, and eminently re-usable data. This [BBSRC-funded PLUTo project](#) aims to develop software tools which can extract phylogenetic data directly from the PDFs in which these data are often siloed. Why PDF? Many biodiversity journals do not provide XML and are not deposited in PMC, indeed some like those published by Magnolia Press (e.g. *Zootaxa*, *Phytotaxa*) are *only* made available as PDF. Therefore tools that handle PDFs are needed if we are to discover, reclaim & re-use all the phylogenetic data that is otherwise buried in the literature. This talk will also discuss helpful changes to UK copyright law (to be introduced sometime this year) which will legally enable and empower this type of data discovery, liberation & re-use for 'non-commercial'[3] research purposes.

References

- [1] Stoltzfus, A., O'Meara, B., Whitacre, J., Mounce, R., Gillespie, E., Kumar, S., Rosauer, D., and Vos, R. 2012. Sharing and re-use of phylogenetic trees (and associated data) to facilitate synthesis. *BMC Research Notes* <http://dx.doi.org/10.1186/1756-0500-5-574>
- [2] Drew, B. T., Gazis, R., Cabezas, P., Swithers, K. S., Deng, J., Rodriguez, R., Katz, L. A., Crandall, K. A., Hibbett, D. S., and Soltis, D. E. 2013. Lost branches on the tree of life. *PLoS Biology* <http://dx.doi.org/10.1371/journal.pbio.1001636>
- [3] Klimpel, P. 2013. Consequences, Risks, and side-effects of the license module Non-Commercial – NC [English translation] http://openglam.org/files/2013/01/iRights_CC-NC_Guide_English.pdf