Title: **Open as a strategy for durability, reproducibility and scalability**
Authors: Jonathan A. Rees, Karen Cranston
Author affiliations: National Evolutionary Synthesis Center (NESCent), rees@nescent.org
URL for the overall project web site: http://opentreeoflife.org
URL for accessing the code: https://github.com/OpenTreeOfLife
Open source license: GPL v. 3 and BSD 2-clause

Open Tree of Life aims to create a complete and dynamic evolutionary history of all species by combining published phylogenetic trees with taxonomic hierarchies. Being a grant-funded academic project, our strategic decisions have been driven by the goals of scientific reproducibility of computational processes, scalability through automation that replaces what in other projects have been manual steps, and project durability beyond the end of the grant period. We have tried to make the project as open as possible because long-term success depends on scientific data sharing and volunteer curation, and because forks of the project may, we hope, eventually end up being as scientifically productive as the 'master' branch. The project practices the following:

- Free software, open access publications, and open data (CC0 when possible), with inputs, intermediate artifacts, and outputs available on the web
- Software development in the open on github, with biologist user/curators encouraged to use the issue tracker
- A largely open approach to scientific decision-making and progress reporting, using Google groups, Google drive, and similar tools
- Technical "opening" of artifacts through reliance on text files on github, as contrasted with a database, for live maintenance of data, using NeXML (an open standard for phylogenetic study data), json, tsv. In addition, provenance is tracked so that file origins are clear, improving transparency
- Transparency of study methods through scripting (mostly make and python); anyone can rebuild the outputs from the inputs. For example, taxonomy construction involves alignment of input taxonomies and correction of errors. Detecting errors is manual, but all correction steps are recorded, with provenance and evidence, as operations that are applied by the build script, similar to a log replay. This makes corrections reuseable.

As of April 2014 we are pre-launch, but already enjoying the benefits of this approach. The decentralized group of biologists and programmers can coordinate using open services and use resources on the web without having to worry about credentials or secrecy, and individuals not directly connected with the project can see what we're doing and contribute suggestions and data. In the future we are planning and hoping for community contribution of phylogenetic trees to our repository, taxonomy improvements, and of course research that builds on anything and everything we've done.