

MetaData can be easy!

Rainer M. Krug¹

&

Owen L. Petchey¹

¹ Department of Evolutionary Biology and Environmental Studies, University of Zürich, Zürich, Switzerland

Abstract

The data generated in and for research increases dramatically not only in size but also in the number of data sets. For (later) re-use of data (by the creator of the data as well as by others), the metadata associated with data sets is as important as the data itself. But when 'normal' researchers are asked to provide metadata, the enthusiasm is severely dampened due to the complexities of most metadata schemes.

It is generally easy to provide general ("bibliographic") metadata, but this type of metadata is often not very useful when trying to find the data one is looking for.

Going further, the definitions of metadata schemes are often so complex, that only somebody very familiar with these (and preferably fluent in reading and writing of xml) is able to use them to their full potential.

Consequently, metadata remains the poor stepchild of data sets.

I will present an approach which tries to overcome this weakness by

1. defining a domain specific metadata scheme which is being driven by scientists in a specific domain and their needs to find other data sets which can be used in their research
2. use an R package to define the metadata scheme
3. use a spreadsheet to enter the metadata per experiment and not per dataset
4. validate the metadata and creating a user friendly report
5. export the metadata to xml format for further usage
6. bundling functionality as well as scheme definition into one convenient R package.

It is planned to bundle this into a shiny application, so that users do need no knowledge of R

to use these metadata schemes.