

ALA Data Cleaning

Table of contents

Welcome

Data cleaning is the exploration, detection and correction of erroneous data. The standard of what ‘clean’ data looks like varies considerably across projects and data sources. This means that there is no ‘one-size-fits-all’ approach.

Our goal in creating this resource was to assist researchers and decision makers that may have limited experience with cleaning geo-referenced biodiversity data in R.

In this book, we provide an overview of a typical data cleaning workflow - from acquisition, to identifying potential errors to correction. Throughout each chapter, we include practical guidelines, R code blocks and additional resources that may aid with each data cleaning step.

The content of this book is guided by the current state of biodiversity literature on preparing data for species distribution modelling. For more details about how this was done, please refer to the [Appendix](#). All resources that have been consulted for this book can be found in [References](#).

How to contribute

We would like to preface that we are not experts in data cleaning, but felt there was need for a consolidated resource to guide data cleaning decisions.

We welcome contributions to this document and suggest folks to submit pull requests at the [GitHub repository](#) of this document. Contributing guidelines can be found at XX (Maybe another section or just a MD in GitHub Repo)

Alternatively, if you have questions please submit a [GitHub issue](#).