

The Austronesian Basic Vocabulary Database

Simon Greenhill, Russell D. Gray, and Robert Blust

University of Auckland and University of Hawai'i

Much of the valuable linguistic data that has been collected over the years is languishing in filing cabinets and is not immediately available to linguists and interested members of the public. Over the last twenty years, Blust has collected a large number of Swadesh word lists from Austronesian languages, which we (Gray and Greenhill) are attempting to use in large computational analyses. As part of this project we have taken this opportunity to "computerize" these word lists and construct a large scale comparative database (<http://language.psy.auckland.ac.nz>) of this vocabulary. This database currently has word lists from 335 languages, for a total of 60,617 words. First, we shall describe some of the technologies required to build a repository such as this. Second, we will talk about the benefits of releasing data onto the internet for collaborative purposes. Finally, we will discuss our plans for expansion and consolidation of this database and take requests.