Appendix 1 – ACNC Data Cleaning Rules

VCOSS adopted the data cleaning rules that were developed by the Centre for Social Impact and the Social Policy Research Centre in their report Australian Charities 2014..

Syntax for data cleaning and data analysis

SPSS was used to clean and analyse the data. Syntax for the data cleaning and analysis is available . Much of the cleaning is to turn string (text) variables into numeric variables, which is easier for analysis. We also apply rules as outlined in the following table.

The analysis is easy to follow/recreate using the tables and charts we have publically reported as long as we use the valid data filters as created through the cleaning process.

Year on year we have needed to update the syntax because the ACNC will often change variable names which means we need to update the syntax.

Data rules:

Item/variable	Rule
Creation of Victorian Community Services Industry Dataset	We usually keep all the data from across Australia in the data set and use filters in SPSS to use the data we want for Victoria. **If Main Activity** is one of: Aged Care Activities Civic and Advocacy Activities Economic, social and community development Emergency Relief Employment and Training Housing Activities Income support and maintenance International activities Law and Legal Services Mental health and crisis intervention Other Education Other health service delivery Social Services **If Other Activity** offered by the charity is from one of the above list. We create Community Sector Charity variable. Then to create the Victorian Community Services Charity variable it is if orgs operate activities as above and it is they are based in Victoria (State=Victoria) and operating in Victoria (Operates in VIC). This means that this dataset will pick up organisations based in Victoria, operating in Victoria, but also operating in other states.

Item/variable	Rule
Size of organisation	Organisations self-select their organisational size Small = Revenue Less than $250,000 Medium= Revenue of $250,000 to $999,999 Large = Revenue of $1 million or more This is checked against Total Gross Income of the organisation We have also added an additional size category for organisations into the following: Extra Small= <$50,000 Small = $50,000-<$250,000 Medium = $250,000 to <$1m Large-$1m to < $10m Extra Large $10m to < $100m Extra-extra- large >$100m This will be useful to track over time in terms of whether the number of large to extra/extra-large organisations has grown
Income	Data is checked, if there is no income for the organisation then these are excluded from the analysis
Inaccurate Income Data	The sum of individual income fields and total income differ by more than $25,000 for small charities The sum of individual income fields and total income differ by more than $100,000 for medium charities The sum of individual income fields and total income differ by more than $1,000,000 for large charities
Inaccurate expenditure data	The sum of individual expense fields and total expenses differ by more than $25,000 for small charities The sum of individual expense fields and total expenses differ by more than $100,000 for medium charities The sum of individual expense fields and total expenses differ by more than $1,000,000 for large charities Employee expenses per reported employee exceeds $300,000.
Ratios	Ratios (e.g. Government grant per total gross income) exceed 100% Negative values (e.g. negative employee expenses or liability).
Valid Financial data	Variable created to ensure that after all checks data is valid