Issued to:

Regulator	Australian Competition and Consumer Commission (ACCC)
Date:	March 2021
Subject:	Guide to data release

Introduction

This short guide is intended to accompany the data release for Measuring Broadband Australia (MBA) periodic reports and provide clarifying information, including:

- How it is produced
- How to use the data release
- A comprehensive data dictionary

The data

The data provided includes Whitebox level aggregates of a selection of test results for all Whiteboxes used in MBA periodic reports. The test results covered are those for download, upload, latency and outage tests. Additionally, the data provides descriptive information (metadata) about each measuring device including, access technology, speed tier and its location.

In preparing the periodic reports, each measuring device goes through a process of validation to ensure sufficient test results have been recorded and that all accompanying metadata is accurate. For those Whiteboxes which pass this process of validation, test results for each metric are then aggregated to give Whitebox level summary data. A portion of this summary data makes up the data release.

The Whitebox level summary data includes a minimum of five test results in the reporting period, otherwise the Whitebox is excluded.

The data provided includes three different aggregations of underlying raw data for each Whitebox. These are:

- 1% trimmed mean – the mean of test results for each unit excluding the largest 1% and smallest 1% of values. The trimmed mean is used so as to dampen the effect of any extreme observations for any measuring device.
- Minimum – the single lowest value recorded for a Whitebox during the measurement period.
- Maximum – the single largest value recorded for a Whitebox during the measurement period.

Download, upload and latency data in this release is provided for all hours and busy hours. These are defined as follows:

- All hours include tests run at any time on any day.
- Busy hours include tests run from 19:00:00 to 22:59:59 Monday to Friday.

Outage data is the average number of outages of 30 seconds or longer recorded by the Whitebox each day for all hours of the day except for between 00.00 and 05.59 (the time period when network maintenance and upgrades are commonly performed).

The RSP is provided for each fixed-line NBN Whitebox. This includes RSPs with insufficient Whiteboxes to be directly included in the report. In this case, the data is still available but the RSP is labelled as part of 'Other RSPs'. For NBN fixed wireless (NBN_FW, in the 'connection_type' column of the data), the RSP has been masked due to the small sample of fixed wireless Whiteboxes currently within the sample.

All metrics in the data release are rounded to two decimal places. Due to this rounding, metrics in Measuring Broadband Australia (MBA) periodic reports may not precisely match.

Guide to produce the report results

What is included in the dataset

The results presented in the periodic reports that are published under the Measuring Broadband Australia program are aggregates of Whitebox summary data. Therefore, results from the report can be recreated broadly by calculating simple averages of individual measuring device results (most commonly grouped by metadata fields such as RSP, tier or connection type for NBN fixed-line services). In the report download and upload results are broadly presented as a percentage of maximum download or upload speed for the participants' speed plan. This can be calculated by dividing each Whitebox measure by the respective maximum speed.

The data release includes additional data and will allow observations to be drawn in addition to those that have been presented in a MBA program report. These data or observations could have been omitted from the MBA report for editorial reasons, or alternatively for reasons such as small sample size that may make them not reliable indicators of general performance.

For outages a 30 second threshold has been chosen because it is at a point where users are much more likely to start noticing the impact, either due to video streams starting to stall, web pages failing to load or video calls pausing.

What is excluded from the dataset, e.g. empty cells

Some data values may be missing for some units and this is due to insufficient tests produced by these units. For NBN fixed wireless, the RSP has been masked due to the small sample of fixed wireless Whiteboxes currently within the sample.

The data release does not include all metrics that are included in the MBA periodic reports. This includes the results of additional metrics (e.g. web loading times, Netflix streaming results, video conferencing) and results calculated directly from raw test data (e.g. busiest hour results, distributions based on raw tests).

Data dictionary

The following table includes a description of each variable in the raw data release.

Variable	Description	Unit of measurement
unit_id	A unique identifier for each Whitebox	n/a
connection_type	NBN (fixed-line) or NBN_FW (fixed wireless)	n/a
technology	The access technology used for the service	n/a
tier	The download and upload speed associated with the wholesale NBN speed tier of the service.	If an NBN connection: download speed / upload speed Megabits per second (Mbps)
rsp	Retail Service Provider	n/a
state_or_territory	The State and/or Territory in Australia the Whitebox is located in. NB: South Australia (SA) and the Northern Territory (NT) have been combined due to low sample sizes.	n/a
is_this_service_impaired	A service is classified as 'impaired' for a Fibre to the Node and Fibre to the Building service where NBN Co acknowledges that there is a physical limitation on the service's attainable speeds such that it is less than the relevant NBN tier.	n/a
is_this_service_ underperforming	A service is classified as 'underperforming' if no more than 5 percent of speed tests conducted over the service achieved a download speed that was above 75 percent of the download speed of the relevant NBN tier.	n/a
gateway_mac	The fuzzed MAC address of the participants' router	n/a
All hour trimmed mean download speed	The 1% trimmed mean of download test results run in all hours	Mbps
All hour minimum download speed	The minimum download speed test result recorded during all hours	Mbps
All hour maximum download speed	The maximum download speed test result recorded during all hours	Mbps
Busy hour trimmed mean download speed	The 1% trimmed mean of download test results run in busy hours	Mbps
Busy hour minimum download speed	The minimum download speed test result recorded during busy hours	Mbps
Busy hour maximum download speed	The maximum download speed test result recorded during busy hours	Mbps
All hour trimmed mean upload speed	The 1% trimmed mean of upload test results run in all hours	Mbps
All hour minimum upload speed	The minimum upload speed test result recorded during all hours	Mbps

Variable	Description	Unit of measurement
All hour maximum upload speed	The maximum upload speed test result recorded during all hours	Mbps
Busy hour trimmed mean upload speed	The 1% trimmed mean of upload test results run in busy hours	Mbps
Busy hour minimum upload speed	The minimum upload speed test result recorded during busy hours	Mbps
Busy hour maximum upload speed	The maximum upload speed test result recorded during busy hours	Mbps
All hour trimmed mean latency	The 1% trimmed mean of latency test results run in all hours	Milliseconds (ms)
All hour minimum latency	The minimum latency speed test result recorded during all hours	ms
All hour maximum latency	The maximum latency speed test result recorded during all hours	ms
Busy hour trimmed mean latency	The 1% trimmed mean of latency test results run in busy hours	ms
Busy hour minimum latency	The minimum latency speed test result recorded during busy hours	ms
Busy hour maximum latency	The maximum latency speed test result recorded during busy hours	ms
Average daily outages	The average number of outages recorded each day lasting 30 seconds or longer excluding between 00.00 and 05.59	Count

#end of document#