

## Measuring Broadband America Program's Mobile Measurement Data Anonymization Policy Review Document

Mobile Collaborative Unofficial Review Draft 12/21/2012

The FCC Measuring Broadband America (MBA) program is based on principles of openness and transparency. The Commission is committed to releasing data used to produce each report coincident with the report's release, and releasing all data within one year of collection. As the Commission expands the MBA program to include measurements of wireless performance, the trust that volunteer subscribers place in us to protect their privacy remains critical. All data collected in this program will be subject to thorough analysis and processing prior to release to ensure that subscribers' privacy interests are protected. We believe that by subjecting all data to a rigorous review to ensure that it has been rendered anonymous, we can maintain our commitment to open data without ever compromising consumer privacy. We propose four high-level principles that will guide the technical approaches of our program: anonymization of all data prior to public release; review of all data categories to identify patterns that might reduce anonymity; setting of minimum numbers of samples for categories of data prior to release to minimize risk of groupings that might reduce anonymity; and coarsening of time or location data prior to release to preserve anonymity.

Our main concern is that we avoid releasing data that might permit identification of a specific consumer by data we release in combination with other data sources.

As part of the MBA mobile program, we will gather data that includes an identifier of the software used by the mobile device, the radio characteristics of the handset, information about the handset type and operating system (OS) version, the GPS coordinates available from the handset at the time each test is run, and the date and time of the test observation. Individually these records may pose no risk of being correlated to other data to identify an individual handset. However, when collected in a database these data sets, third parties might combine other datasets in order to determine a volunteer's identity. By scrutinizing the data prior to release, we will prevent third parties from using our data to identify patterns they might exploit in order to reduce the anonymity of the database.

Rows might include such example features as the following with software identifiers deleted and other elements processed to preserve the anonymity in the data.

Row Level Data Example:

~~{Software\_Instance\_ID}~~ {RF\_Cell\_Tower} {RF\_Signal\_Strength} {RF\_Carrier} {RF\_Bearer}  
{Handset\_Type} {Handset\_OS\_Version} *{Handset\_GPS}* {Active\_Test\_Result} *{Time\_Stamp}*

## **Data Anonymization and Processing Principles**

1. We will anonymize all data prior to releasing it to the public in order to prevent the information you provide from being linked directly with your handset.

Through the software that you install on your handset, we run various tests and collect information about your broadband performance. We also save that information on our secure servers. The information is tagged with an identifier for your handset to maintain the test infrastructure, and to enable us to deliver information to you about your individual performance. We delete references to your handset identifier unless strictly necessary and will never release that identifier or publicly release data that includes your identifier.

2. We will review all categories of data prior to release to identify patterns that might reduce the anonymity of our data.

The release of the raw data set will not compromise your privacy because any direct relationship between the data gathered from your individual handset will be deleted. We further ensure that identifiable patterns in the data will be presented in a way to prevent a third party from grouping data gathered from your handset in ways that would reduce the data's anonymity.

3. We will review the combinations of aspects of the data prior to release to set minimum numbers for groupings of handsets associated with categories within the dataset.

We will never release data that reveals patterns that would allow a third-party to filter data gathered from an individual handset from our datasets. We also process the data to make sure no patterns could be established from the data that would allow your data to be filtered as part of a small group. In the analysis, we will define minimum numbers for groupings of handsets associated with categories within the dataset. The analysis will ensure data from your handset is only released as part of a larger set of results, and not a smaller group that might pose more risk to your anonymity. We believe this step strengthens the data set against a third-party who might try to use a narrow group of results from our data to link results with other data sets.

4. We will review data prior to release to coarsen time or location data.

The time at which measurements were made and the location of the device at the time the tests were run are crucial to understanding mobile broadband performance. To ensure anonymity of MBA mobile volunteers, we may coarsen the level of detail of any time or location information in measurement results prior to release. We believe this step strengthens the data set against a third-party who might try to correlate information in other datasets about time or location of your broadband usage data to link results with other data sets.