SamKnows Ltd
135 Park Street
London SE1 9EA

www.samknows.com

FROM:	Charlotte Wilson (charlotte@samknows.com)
TO:	FCC
DATE:	Friday 19th May, 2017
SUBJECT:	WEIGHTING DESCRIPTION

Following the Measuring Broadband America Fixed Collaborative meeting on Wednesday 3rd May 2017 there was an action for SamKnows to prepare a document that would include the following information:

- New dashboard tooltips content
- Updated home dashboard screenshots
- Simplified weighting explanation with example

This weighting information is included in the document below.

Weighting Explanation

In 2016, the MBA report introduced weighting into the calculations used to produce some of the statistics in the report. Most notably, the ISP-level averages were now calculated as a weighted mean of the service tier medians. Note that each service tier's median was calculated by taking the median of the average speed of each Whitebox within that service tier.

The 2016, the weighting approach used two variables in calculating the weights for each service tier:

1) The market share of that service tier

2) The number of Whiteboxes deployed on that service tier

Whilst the first variable (the market share) is clearly critical, the second one is not. In fact, including the number of Whiteboxes in the weighting has a downside: Adding or removing Whiteboxes will affect the weights, and thus the overall results (very slightly). Moreover, if we deem that 25-30 Whiteboxes is sufficient to represent the performance of a service tier, then we should regard any figure produced from that sample as a ground truth, and should not try to correct it with weighting. For this reason, in 2017 we have decided to simplify the weighting approach such that only the market share of the service tier is used to calculate the weights.

This change has been tested on the 2017 data set and has been found to produce <1% difference in the results. This slight difference is caused by the old weighting approach correcting for the number of Whiteboxes, whose distribution does not precisely match the market share. However, as noted above, providing the number of Whiteboxes on a tier meets the minimum sample size, we should not consider the number of Whiteboxes in our calculations.

(continued)

A question arose on the call regarding whether we still use the number of subscribers across all ISPs for the universe in the weighting calculation, or just consider the subscriber count of the ISP in question. SamKnows pointed out that the outcome is the same regardless. The below example demonstrates why.

For calculating the ISP level averages, it does not matter whether we take the total N in the weighting calculation to be "the sum of the total N for all subscribers on ISP A (for plans that are covered)" or "the sum of the total N for all subscribers on all ISPs (for plans that are covered)". Please see the following example below:

A hypothetical ISP, ISP A has two tiers, 6Mbps with 2000 subscribers and 12Mbps with 3000 subscribers. The total is 5000 subscribers. Both tiers are included in our study. Let's say there are another group of tiers from other ISPs all included in our study, with a further 5000 subscribers in total. So we have 10,000 subscribers in the US.

If we take the total N to be 5000 (i.e. just ISP A's users), we have:

ISP A 6Mbps: 2000/5000 = 0.4
ISP A 12Mbps: 3000/5000 = 0.6

If we take the total N to be 10000 (i.e. the whole of the US), we have:
ISP A 6Mbps: 2000/10000 = 0.2
ISP A 12Mbps: 3000/10000 = 0.3

But, crucially, the ratio between both of those two sets of fractions is the same (2:3)!

Let's put some performance figures behind these tiers too:
ISP A 6Mbps: 90%
ISP A 12Mbps: 110%

Now let's try calculating the weighted average:
Using the ISP A only weights:
- ((90*0.4) + (110*0.6)) / (0.4+0.6) = 102%

Using the entire US weights:
- ((90*0.2) + (110*0.3)) / (0.2+0.3) = 102%

As shown above, it does not matter which when you work out the final weighted average for an ISP you get the same answer, because the ratios between the tiers are the same.

(continued)

ENDS