

The case for (support for) public tools and longitudinal measurements

Matt Calder (Columbia) and Ethan Katz-Bassett (Columbia)

The value in much Internet measurement research is in the results, but the Internet can change quickly, meaning that published results go stale quickly—one cannot know whether they still reflect the state of the Internet—and measurements, like cars, lose much of their value as soon as they are driven off the lot. We provide an illustrative example from our own experience, drawing from our IMC 2013 paper "Mapping the Expansion of Google's Serving Infrastructure." This paper developed a methodology for uncovering a complete client-server mapping for `www.google.com` (or similar sites), including the geographic and logical (AS) locations of servers. This mapping provides an insight into Google's global deployment and strategy and can be joined with other datasets to explain performance, perform what-if analysis, evaluate approaches for censorship evasion, or explore a range of other questions. However, the results vary significantly depending on when the measurements are performed. During IMC 2012, Google used servers in 100 ASes (and, we suspect, the answer would have been ~1 a year earlier). When we submitted to IMC 2013, the number had increased to 200 ASes, and the paper was accepted based on this ~2x growth. By IMC 2013, Google used servers in more than 800 ASes. By IMC 2015, it was up to 1300 ASes. Today, `www.google.com` uses servers in only Google's AS, but, invisible to the technique from the IMC 2013 paper, Google also caches static content in over 3000 ASes. Anyone relying on the published numbers a year—or two, or five—later could significantly misunderstand Google's deployment.

This example shows how aspects of the Internet can change rapidly and unpredictably, and so the community needs up-to-date, public, longitudinal data, but instead the incentives for publication, funding, and graduation/promotion often favor one-off snapshots that may quickly be stale. Program committees favor novelty rather than revisiting previous results, especially if those previous results have not changed. Most funding sources fund shorter periods than the eight years to capture the example Google results, and most calls favor innovation and funding at the granularity of a "grad student year," rather than the smaller amounts of money over longer periods of time needed to maintain projects like this after the initial development. The need for graduate students to publish and graduate gives an incentive to abandon projects after publication. However, if existing results are not publicly refreshed, one cannot know which remain "safe" to use. Efforts including Community Awards, reproducibility badges, and poster Reproducibility Track intend to encourage public data and revisiting of results, but most are assessed at submission time and provide limited incentive compared to the opportunity cost.

We believe that strategic and achievable changes to the incentives can encourage long-running public measurements/tools, which will support better science and more impactful research. Ideas include:

- **Funding:** Maintenance funding with an easier submission process and significant consideration for the impact of the previous project. Summer REUs/fellowships to productionize existing measurements. Funding for longterm adoption by CAIDA/MLab/RIPE NCC/etc.
- **Publication:** Review forms that encourage consideration of coverage across space/time and of community contribution. New tracks for expedited acceptance of short papers revisiting earlier studies, perhaps without a live talk at the conference.
- **Curation:** Curated guides to the data available from a conference (updated yearly for what has been refreshed). Common application platforms and tooling for maintaining and sharing best-practice pipelines for issuing, processing, and publishing measurements.
- **Recognition:** Test-of-time community awards to recognize efforts that helped the community (vs the current award that predicts what will be helpful). Recertification of reproducibility badges (and similar for public, up-to-date data) over time, with recognition every N years. Exit polls (and prizes to help fund?) at conferences: what data do you most want?