By design, altmetrics providers aggregate data from many different sources. Our overarching goal is to observe all citations and mentions on the web, and, for better or for worse, the web is always changing. Link rot and content drift have been discussed at length in the literature, but there are no easy fixes, and, most importantly, they are not the only obstacles to making altmetrics data FAIR-er. Continuous updates, streaming APIs, and rolling releases are great to deliver fresh data. However, with any web service, if new or updated data is made available to the users (near-)instantaneously, this creates a cascade of challenges to version datasets and guarantee the repeatability and the reproducibility of all studies that build upon data collected from that service. For example, if a user runs queries to compare the attention received by two publications, and the underlying database is updated to account for (potentially unrelated) new content between the two queries, how do we guarantee that the responses represent comparable snapshots of the web?

Altmetric.com, for instance, offers no version information via their API. The best users can do in data citations is to mention when the data was retrieved. Preliminary results from an ongoing analysis of quantitative studies that analyze altmetric data (N=141) show that 50% of studies that build upon Altmetric.com data mention no version information at all, and that only 32% mention the day the data was collected.\(^3\)

With Cobaltmetrics, we balance two main goals. On the one hand, when API calls are made, we want to offer fresh data. On the other hand, when API calls are made over short periods of time, we want the responses to be sensibly comparable. In order to control when and how data is updated, we ingest entire snapshots of source datasets, and we now only publish updates and additions once a week.\(^4\) In a similar vein, our API adds a machine-readable fingerprint to all API responses.\(^5\) A fingerprint uniquely identifies all the resources (code and data) that were used to build a given version of Cobaltmetrics. The fingerprint changes whenever any of the resources (code or data) changes. Fingerprints make citation data citable. If users make multiple API calls, and the responses include different fingerprints, they might be comparing apples to oranges. To address this issue, fingerprints can be used in content negotiation mechanisms. We support various headers, inspired by HTTP entity tags and the Memento project. Fingerprints should also be used in data citations, along with the persistent identifiers obtained from the PID Services Registry.\(^6\)

These features are available to all users of Cobaltmetrics by default and at no cost. We believe that these recent developments are empirical evidence that altmetrics providers can offer more guarantees to their users, without compromising on other values. The goal of this short talk is to present these features to the community, and to gather feedback and start conversations to further improve the quality of altmetrics data.