

Ellen S. Miller  
Co-Founder and Executive Director  
The Sunlight Foundation  
1818 N Street, NW  
Suite 300  
Washington, DC 20036  
July 15, 2010

Comment on WT Docket No. 10-131 - Review of Wireless Telecommunications Bureau Data Practices

To the Commission:

The Sunlight Foundation welcomes the opportunity to comment on the data practices of the FCC's Wireless Telecommunications Bureau.

Founded in 2006, the Sunlight Foundation is a non-partisan non-profit designed to use the power of the Internet to catalyze greater government openness and transparency. We are committed to improving access to government information by making it available online—indeed redefining “public information” as meaning “online”—and by creating new tools and websites to access that information and engage communities in their use.

Our comments will address section (D) of the Wireless Telecommunications Bureau's request, relating to “recommendations on how it may improve dissemination of the reports and analyses it produces.”

These comments will primarily relate to subsection (1) of this request, in how best to implement the National Broadband Plan recommendation that the Commission have a “general policy of making the data it collects available to the public, including via the Internet in a broadband data depository, except in certain circumstances such as when the data are competitively sensitive or protected by copyright.”<sup>1</sup>

## **Introduction**

Through the FCC's National Broadband Plan, the FCC has already recommended that it publish its data publicly and online in a central depository. In fact, the FCC has already begun to do so, hosting a number of documents and datasets at the FCC's Reboot initiative.<sup>2</sup>

However, many of the FCC's datasets are still unavailable electronically, and not accessible through Reboot. It's not clear from the working inventories of datasets provided alongside this public notice<sup>3</sup> exactly which of these datasets are available through Reboot, but the number of bulk data entries listed on the “Database Exports” and “Mapping Data” pages are quite small by comparison.

---

1 <http://www.broadband.gov/plan/4-broadband-competition-and-innovation-policy/#s4-1>

2 <http://fcc.gov/data/>

3 <http://reboot.fcc.gov/data/review>

We support the efforts the FCC has taken thus far to publish its data through Reboot, and urge it to continue expanding its offerings and improving its systems for disseminating data. We also recommend that the FCC adhere to certain principles of open data we believe are essential to fulfilling the FCC's stated goal to "transform the FCC into a model of excellence in government."<sup>1</sup>

## **Machine Processable**

The FCC's data depository should publish as much data as possible in machine-processable formats. Machine-processable information is data that can be easily parsed and manipulated by computers. CSV<sup>2</sup> spreadsheets, XML documents, and GIS<sup>3</sup> data all fit this description; graphs, PDF documents, and written articles discussing data, while important in their own right, are not machine processable. Machine processable data should also be accompanied by human-readable documentation that explains the nature of the data and how to process it.

Citizens and organizations who wish to do their own analysis, visualization, or discussion of data using spreadsheets, databases, or their own custom software are greatly served by providing this data in machine-processable formats. If a dataset is not in a machine-processable format, third party analysis of this data can be weakened, incomplete, or even impossible.

## **Non-proprietary**

Any machine-processable formats the FCC uses to release data should be non-proprietary. A non-proprietary format means that no one particular piece of commercial software or vendor is required to access the data in this format. This ensures that cost will not be a barrier to accessing the FCC's datasets, even indirectly.

For example, a Microsoft Access database file would be considered proprietary, by virtue of its dependence upon a commercial software product (Microsoft Office), and thus unsuitable for the release of government data. A non-proprietary alternative format, such as CSV or SQL, would be appropriate.

## **Primary and Complete**

Datasets released by the FCC should include complete primary source data, containing the entirety of what the FCC collected in its original, raw form. This should include details about how the data was collected, and the original source documents recording the collection of the data.

This does not exclude releasing data in secondary formats; if the agency performs any aggregation, reformatting, or "cleaning" of the data, the results of these processes should be

---

1 <http://reboot.fcc.gov/about-reboot>

2 CSV stands for "Comma-separated values", a common and non-proprietary way to publish tables of data.

3 GIS stands for "Geographic information systems", and can include any type of geographic data. One example would be the Census' shapefiles representing congressional districts.  
<http://www.census.gov/geo/www/cob/cd110.html#shp>

published as well. These should include details about the process, including any formulas, explanations, or metadata, so that they may be reproducible and verifiable by third parties using the primary source data.

### **Timely**

The FCC should make datasets available to the public in a timely fashion. Whenever feasible, information collected by the government should be released as quickly as it is collected. Priority should be given to data whose utility is time sensitive.

Ideally, the FCC will eventually be able to release data quickly enough to offer real-time updates; this maximizes the utility the public can obtain from this information, and empowers analysis and discussion of the data in real-time as well.

### **Permanent and Accessible**

Datasets released by the FCC should be as accessible as possible, by being placed in public, permanent locations. A given dataset should be downloadable from a single URL, and this URL should not change over time.

The FCC should ensure data is accessible via automated electronic access. Barriers to automated access include making data accessible only by submitting web-based forms, or by requiring browser-oriented technologies (e.g., Flash, Javascript, cookies, or Java applets) in order to access this data.

It is expected that datasets may be updated over time. When a dataset is updated, older versions of this dataset should remain available, and the reasons for the update made clear.

### **Identifiable Entities**

A frequent problem when working with government datasets, especially across agencies but even within individual agencies, is an inability to uniquely identify the entities referred to in the datasets. This can create situations where two separate entities share the same name, or where one entity is referred to by multiple names, which in turn creates confusion and uncertainty about how to draw conclusions from data.

Assigning entities unique identifiers, or including existing unique identifiers from other public systems, wherever possible, vastly increases the reliability and quality of datasets.

This is especially important when these entities are referenced in other datasets, including datasets published by other agencies. Where possible, the FCC should collaborate with any other agencies whose datasets reference the same entities and attempt to standardize on unique identifiers for these entities. Even in the absence of collaboration, the FCC should still create unique identifiers for entities in datasets which lack them.

## **Free of Cost**

The FCC should not impose a cost barrier in front of its datasets. Imposing fees for access skews the pool of who is willing (or able) to access information, even when the cost is *de minimus*.

## **Licensing**

Any depository of datasets should be clearly labeled as containing works of the government, and available without restrictions on use as part of the public domain.

## **Data.gov**

As of this writing, the FCC has eight datasets listed on Data.gov<sup>1</sup>, a small subset of the data available at FCC.gov/data. We recommend that all data the FCC makes available online also be listed on the federal government's central data portal. Doing so will improve discoverability and dissemination of the FCC's public data.

The FCC's current data offerings already embrace some of the principles listed above; it is our hope that the agency will reaffirm and expand its commitment to these principles as its data offerings continue to grow.

Sincerely yours,

Ellen S. Miller

---

1 <http://www.data.gov/catalog/raw/category/0/agency/95/filter/type/sort/page/1/count/25>