The Center for Internet as Infrastructure, LLC

369 Montezuma Avenue., #631
Santa Fe, NM 87501-2626

September 20, 2019

Comments r.e. WC Docket Nos. 11–10 and 19-195, FCC 19–79

Robert A. Ballance, Ph.D, Principal
The Center for Internet as Infrastructure, LLC

Thank you for the opportunity to comment on the new data submission rules for FCC 477 polygon-based data.

The Center for Internet as Infrastructure, LLC, is a consumer of the current FCC 477 Wireline and Mobile data in both tabular and Shapefile formats. The Center supports the I3 Connectivity Explorer, an application that combines data from U.S. Government agencies — including the FCC, US Census, Environmental Protection Agency (EPA), US Department of Agriculture (USDA), The National Center for Educational Statistics (NCES), Institute for Museum and Library Services (IMLS) — and other open data sources such as the Measurement Lab (M-Lab) and the ProPublica Congress Application Programming Interface (API). The platform then localizes the national data to locations of interest. Our comments are based on real-world experience and background.

These comments assume that the data reported to the FCC will (i) be used internally for FCC purposes, (ii) will continue to be openly available to all data users in a non-proprietary format that can correctly encode the polygonal data without loss of precision, and (iii) will continue to be used by third-party organizations, including community groups, to track the status of broadband connectivity in their local areas. These comments apply to all uses of shape data.

Recasting data reporting into "polygon shapes" alone will not resolve any issues with the FCC's broadband mapping data. However, the shift to shape data will increase the demands on the FCC Staff to manage and interpret more complex data submissions. Reliance on shape data has the potential to introduce new errors and inaccuracies into the reports derived from the data. Most importantly, the shift from tabular to shape data will increase the demands on consumers that load, process, or analyze the data.

A data quality assurance program will be essential to the success of the FCC's transition from census-block data to graphical shapes. With every data provider required to report in the new format, the Commission should anticipate and proactively address likely new data quality problems akin to the well-studied ones encountered when maintaining quality in crowd-sourced applications. As one paper states, "Quality is a verb [7]". An active program to assure the quality and validity of the new data will benefit the FCC, the providers, and the end-users.

The FCC should take the following steps to assure a successful transition from Census-area based reporting to shape-based reporting:

1. Develop open technical standards for all data submissions. The results should build upon the FCC instructions entitled "How Should I Format My Mobile Broadband Deployment Data? [1]"
2. Ensure that all FCC data continue to interoperate with Census-supported geographic data.
3. Fully fund and deploy active processes to verify that all submissions are accurate according to the reporting requirements and valid according to the standards.
4. Clearly document any loss of accuracy or other changes if redistributed data are modified in any way from that submitted to the FCC.
5. Ensure that all data are redistributed as open data [3] and are valid according to the published standards.
6. Release all documentation and computer code that are used in the validation process as Open Source components to ensure transparency and repeatability.

Each step is discussed further below.

1. Develop open technical standards for all data submissions

Without appropriate standards, submitted data will contain errors and other artifacts that will hinder the mapping program. Throughout the development of the standards, the FCC standards themselves should be developed and managed using open-source tools and platforms. In particular, the Center strongly recommends that all polygons must be "valid" according to the Open Geospatial Consortium standards [4, 5, 6], and comply with more stringent technical standards than those currently required by the FCC.

The standards must resolve the following questions; other questions will undoubtedly need to be addressed as they are identified.

1. What are the minimum standards for polygons? Again, the Center strongly recommends that all polygons must be "valid" according to the Open Geospatial Consortium standards [4, 5, 6], and be easily processed with Open Source Geographic Information System (GIS) platforms such as PostGIS as well as commercial mapping tools.
2. What is the accuracy required of the polygon shapes?
3. What spatial reference systems will be permitted? It's unlikely that the FCC will want to allow every submission (or every polygon) to use an arbitrarily chosen map datum and projection for submissions. Limiting the set of allowable reference systems to a small number will yield large operational benefits.
4. Will the polygons be allowed to specify internal "holes"?

5. Will "multi-polygons" be supported?
6. What level and types of simplification of polygons (*e.g.* smoothing) will be allowed?
7. Will providers submit multiple polygons in a single file?
8. What is the maximum allowable size (number of vertices) for a polygon? Large, complicated polygons can incur high computational overhead.
9. What are the size and submission limitations on polygons? Multitudes of small polygons can incur high computational overhead.
10. What metadata fields will be required for each polygon?
11. Can submitted data span state boundaries?
12. Can submitted data span county boundaries?

    The treatment of state and county boundaries is important for smaller teams that are interested only in limited areas; they may not be in the position to manipulate national data sets.

2. Ensure that all FCC data continue to interoperate with Census-supported geographic data.

The FCC should ensure that its provided data can easily interoperate with the Census-provided geographies. The US Census provides official shapes for a wide range of geographic areas. The move away from Census blocks toward arbitrary shape data will require users to deploy spatial database techniques in order to align shape data to geographic areas.

For example, the Department of Housing and Urban Development (HUD) denotes opportunity zones using Census tracts. To correlate broadband data expressed as polygons with opportunity zones, it will be necessary to intersect the broadband coverage shapes with the shapes of the tracts that defined the opportunity zones. Under the census-block approach, one needs only to examine the Census block code and the Census tract code, a computationally low-cost operation. Using shapes, much of the down-stream processing of the FCC provider data will require spatial techniques. The need to consumers to perform spatial operations is a compelling reason for requiring adherence to the OpenGIS standards, limiting the number of spatial reference systems, and ensuring that the data can be processed using freely available tools.

3. Fully fund and deploy processes to verify that all submissions are accurate according to the reporting requirements and valid according to the standards.

The FCC must develop and deploy processes to verify that all submissions are both accurate and valid according to the standards. Basic validity checking should be performed on submitted data. It will be necessary to validate both the syntactic and semantic forms of the data (are the polygons valid per the standards?) and the accuracy of the content (do the polygons properly represent the coverage areas?). The

FCC should plan and budget to handle an increased number of erroneous or non-conforming submissions during the initial rounds of data submissions as the submitting organizations get used to the new requirements.

Statistical methods will be needed to check the accuracy of the content. Even simple statistical checks might have caught the issues that recently necessitated the re-release of the December 2017 Wireline data sets. Sampling techniques to check accuracy might also play a role, analogous to risk-limiting audits in elections. Data quality checks such as those described in [2] should be applied to review the overall consistency of the submissions. A challenge process may also be needed.

4. Clearly document any loss of accuracy or other changes if redistributed data are modified in any way from that submitted to the FCC.

The FCC may need to revise or correct polygon data before it is released publicly. All changes, including any loss of accuracy, should be clearly documented on the distribution web site.

5. Ensure that all data are redistributed as open data [3] and are valid according to the published standards.

The FCC has a long, and admirable, history of releasing the 477 data sets as open data. This activity should continue even as the form of the data changes. Since the resulting polygon data will contain additional complexity, the FCC should take steps to ensure that any data delivered openly maintains the same quality standards as required of the original data providers. The verification processes used on data submitted to the FCC will provide a strong first-step in ensuring that any data originating from the FCC adheres to standards.

- The FCC data itself should not accept or redistribute data represented in proprietary data formats. The data should be accessible by all users. This, goal, in itself, is achieved with the adoption of the Shapefile format. However, discussed next, it is not sufficient for the Shapefile submissions to be syntactically correct. The contents of the Shapefiles must also adhere to the required technical standards.
- All polygon-based data sets released by the FCC should be conform to standards supported by common, Open Source, spatial database platforms [4, 5, 6]. Validating the data for conformance will ensure higher quality data. For example, the FCC 477 Mobile Shapefiles dated June 2016 were syntactically valid Shapefiles, but contained self-intersections and other pathologies, making the data much less useful than originally intended. End-users, and even the FCC staff, are not in a position to correct bad data.
- Compatibility with Open Geospatial Consortium standards is essential since many users of FCC Open Data are also users of Open Source GIS platforms. One approach is to ensure that all processing in the data validation and publication pipelines checks be run on at least one Open Geospatial Consortium compliant platform such as PostGIS.

6. Release all documentation and computer code that are used in the validation process as Open Source components to ensure transparency and repeatability.

The verification processes will require code development. All code should be released as Open Source using non-proprietary programming languages. This will allow for third-party verification of the programs and to assess the repeatability of the verification process. For example, the quality checks implemented for both submitted and published data should be released as documentation and working code.

References

[1] How should I format my mobile broadband deployment data? https://transition.fcc.gov/form477/MBD/formatting_mbd.pdf.

[2] Peter Mooney, Padraig Corcoran, and Adam C. Winstanley. Towards quality metrics for OpenStreetMap. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '10)*, pages 514–517. ACM, 2010.

[3] Open definition 2.1. https://opendefinition.org/od/2.1/en/.

[4] OpenGIS implementation specification for geographic information — simple feature access — part 1: Common architecture. http://www.opengeospatial.org/standards/sfa.

[5] OGC® best practice volume 4: OGC CDB best practice use of shapefiles for vector data storage. http://www.opengis.net/doc/BP/cdb-shapefile-guidance/1.0.

[6] Ensuring OpenGIS compliancy of geometries. https://postgis.net/docs/using_postgis_dbmanagement.html#OGC_Validity.

[7] Andrew Sheppard and Loren Terveen. Quality is a verb: the operationalization of data quality in a citizen science community. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*, pages 29–38. ACM, 2011.