May 3, 2017

House Committee on Appropriations Subcommittee on Legislative Branch
HT-2, The Capitol
Washington, DC

Dear members of the subcommittee,

On behalf of the Center for Data Innovation (datainnovation.org), we are pleased to submit this testimony to the House Appropriations Legislative Branch Subcommittee.

The Center for Data Innovation is the leading think tank studying the intersection of data, technology, and public policy. With staff in Washington, DC and Brussels, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as predictive analytics, open data, cloud computing, and the Internet of Things. The Center is a non-profit, non-partisan research institute affiliated with the Information Technology and Innovation Foundation.

Congress has made significant strides in publishing open data.[1] Congressional data sets include decades of information on legislation, votes, hearings, member biographies, and federal nominations. This data is a valuable resource for innovation, and companies such as Quorum and FiscalNote have built successful businesses using these datasets.[2] However, obtaining this data and making it usable can be very resource intensive. While Congress publishes many data sets in machine-readable formats and without license restrictions, most of its data sets are only available as a bulk download (i.e. a single large file). Congress does not provide access to most of its data via an application programming interface (API), a series of computer functions that allows developers to build software that interacts directly and seamlessly with discrete portions of a dataset.[3] As a result, Congressional data is unnecessarily difficult for developers to access and integrate into other apps and online services. For example, developers must download

complete datasets, such as the entirety of the U.S. Code, and then extract the information they need, rather than downloading only the relevant portions. For datasets that are updated frequently, this type of limitation is particularly cumbersome.

Should Congress publish information via API as well as in bulk, more businesses could integrate Congressional data into their apps and services. Furthermore, tools such as these can enable staffers in the House and Senate to be more productive by reducing the time they have to spend tracking down legislative information.[4] In the absence of an API provided by Congress, other organizations, such as the nonprofit investigative reporting organization ProPublica, have created APIs to provide access to their own collections of legislative data. However, these APIs are not authoritative sources of data. Moreover, the API owners can impose restrictions on how developers use the APIs, such as limiting to noncommercial uses only or charging commercial users additional fees.[5]

We recommend that Congress should begin publishing legislative data via API in addition to publishing this data in bulk format. As the Library of Congress already transmits legislative data internally via API, Congress should publish this data with a public-facing API as well. Given the success of the Congressional Bulk Data Task Force which led Congress's earlier efforts to make machine-readable data from the legislative branch available for download, Congress should direct the task force to investigate and implement these changes to existing legislative information systems.[6]

Sincerely,

Daniel Castro
Director
Center for Data Innovation
dcastro@datainnovation.org

Joshua New
Policy Analyst
Center for Data Innovation
jnew@datainnovation.org

[1] Legislative Branch Appropriations Bill, 2013, H. Rept 112-511, 112th Cong. (2012)

[2] Joshua New, "5 Q's for Jonathan Marks, Cofounder of Quorum," Center for Data Innovation, November 2015, https://www.datainnovation.org/2015/11/5-qs-for-jonathan-marks-cofounder-of-quorum/2015.

[3] The U.S. Government Publishing Office (GPO) does provide a limited API at https://api.fdsys.gov/.

[4] Joshua New, "Event Recap: Open Legislative Data Could Help Solve Some of the Public's Biggest Frustrations with Congress," Center for Data Innovation, July 2016, https://www.datainnovation.org/2016/06/event-recap-open-legislative-data-could-help-solve-some-of-the-publics-biggest-frustrations-with-congress/.

[5] ProPublica Congress API, accessed April 12, 2017, https://www.propublica.org/datastore/api/propublica-congress-api

[6] Legislative Branch Appropriations Bill, 2013, H. Rept 112-511, 112th Cong. (2012)