Written Testimony of Daniel Schuman, policy director, Demand Progress,
Before the Select Committee on the Modernization of Congress,
On "Opening up the Process: Recommendations for Making Legislative Information More Transparent"
May 10, 2019

Chairman Kilmer, Vice Chairman Graves, and esteemed members of the committee:

Thank you for the opportunity to testify before you today. My name is Daniel Schuman, and I am the policy director with Demand Progress and the Demand Progress Education Fund. I am also the co-founder of the Congressional Data Coalition, the co-founder of the Advisory Committee on Transparency, and I wear more hats than is comfortable to recount. I have had the longstanding pleasure of working with fellow witnesses Bob Reeves, the Deputy Clerk of the House of Representatives, and Josh Tauberer, creator of the legislative information website GovTrack.us.

WHY I CARE AND HOW WE GOT HERE

I started working on congressional transparency more than a decade ago because I saw that Congress was struggling with its responsibilities and I wanted to help. At that time, I saw the possibility of improving Congress's ability to meet its legislative and oversight obligations by marrying transparency with technology.

Even then, it was apparent that Congress struggled with technology. My time as a congressional intern, junior staffer, and legislative attorney with the Congressional Research Service (as well as working at several non-profits) made that struggle pellucidly clear. So, like a good empiricist, I started gathering data and figuring out where I could make the biggest positive impact.

As I rolled up my sleeves, three things became apparent. First, there was a small but smart cadre of people working on improving public access to legislative data, but, generally speaking, they weren't well connected to congressional insiders. Second, there was very little practical data on what was broken in Congress at a granular level. And third, there wasn't really anyone lobbying on improving congressional operations. So, I jumped in with both feet.

I started with the assumption that it would be next to impossible to convince Congress to hire more staff, but it might be possible to turn staffers into super staffers by giving them better tools. In addition, I also assumed that a lot of what people view as undue influence of special interests on Congress is really about information asymmetries — where well-heeled lobbyists have access to information and tools unavailable to public-interest lobbyists and congressional staff.

I began to tinker. I successfully pushed to have all congressional hearings webcast — building upon a change to House rules, then shaming committees that did not comply. If the hearing isn't webcast, after all, it's harder for staff to monitor the proceedings. Moreover, why should you have to be in the U.S. Capitol to see what's happening at an official proceeding?

I spent a year scouring libraries across the country to get all the old Congressional Management Foundation reports on staff pay and retention, typed the data into a spreadsheet, analyzed it, and published the first longitudinal report covering 25 years of House data, entitled "Keeping Congress Competent." No one had ever looked at whether staff were earning more than their counterparts over such a long period of time. Two years later, I analyzed similar data for the Senate.

We made use of the House's Statement of Expenditures, newly published online as a PDF file, thanks to our request. Our developers built a tool to try to make sense of the information — a PDF is a terrible format for data storage — and we were able to perform a real-time study of staff pay ranges for each title. We also used this approach to look at turnover rates by office, and compare turnover rates across offices. To my knowledge, this was the first time that had ever been done. While it was crude, we now had a tool that could do this on an ongoing basis. We could start to figure out which offices were functioning well and what a career trajectory looked like for staff.

The hardest issue to crack was tracking legislation. We knew there were a ton of things that one could do with legislative data. For example, we could figure out if a bill had been reintroduced from a prior congress, or compare different versions of a bill to see what had changed in committee, or build a map of all the official documents connecting to a bill, or keep track of co-sponsors.

Imagine that you could pick a bill and see all the related IG, CRS, GAO, and CBO reports, all the committee reports and amendments, the votes at the various stages, the Dear Colleague letters, statements from the administration, and the floor speeches. Or even better, imagine that you're a staffer who is following a particular issue and you could enter a few key terms in your search engine and get email alerts whenever any matching documents are found. Think of all the time that would save — and all the knowledge it would put at the hands of overworked congressional staff or non-profit advocates. As it turned out, we eventually got a prototype of this working before a funding crunch put an end to the effort. This is the downside of relying on outside groups to build and maintain these tools once they've proven their worth.

But I'm getting ahead of myself. We worked assiduously to advocate for public access to legislative data. We built bridges between the outside technologists and congressional

institutionalists. It took years and years, with stiff opposition from some institutional players. For example, the Library of Congress and GPO were most reluctant to publish legislative information as data. And at the moment we were about to win in a vote on the floor, a compromise was struck and a task force established to "evaluate" our request. The expected result was that the task force — the Bulk Data Task Force — would bury the idea. But that didn't happen. Instead, something wonderful happened.

The Task Force brought together many of the institutional players inside Congress and across the legislative branch. Many didn't know each other that well, others had been quietly working behind the scenes, and the Task Force gave them a mandate to collaborate. So they started meeting, and they met with us, and we presented our request. And they agreed to it — and implemented public access to legislative data — i.e., bill summaries, status information, and text. But they kept meeting with each other, and with the public. They continue to meet and collaborate to this day.

Congress has continued to make progress on transparency on a number of fronts. In recent years, the data publication initiatives have included the online publishing of bills in formats that support reuse of the data by others; committee schedules, documents, and videos; an online House phone directory; the bills and amendments scheduled for a floor vote in the House; the Statement of Disbursements; publication of the US Code as structured data; and the new joint meetings calendar. The Clerk has a non-partisan repository of committee documents at docs.house.gov. We have public access to CRS reports (although not the underlying data). There's a requirement that bills (but not resolutions) be available online 72 hours prior to a floor vote. There's a new Whistleblower Ombudsman. There are regular meetings of the Bulk Data Task Force and the annual Legislative Data and Transparency Conference. And the House has continued its longstanding efforts to draft legislation in the structured data format XML.

<u>WHAT IF</u>

There is still more to do with respect to online publication of information and the publication of that information as data. But we are finally getting to the point where we can have better integration of these data sets to provide contextual information to legislators and the public alike. Some of what we imagined a decade ago — document alerts for staff and contextualized legislative information — is possible in a way that was hard to sustainably implement previously.

Imagine being able to push a button to see how an amendment would change a bill or a bill would change the law. Imagine a dashboard that provides a view into Congress's vital statistics on staff pay, retention, the movement of legislation, and lobbying data. Imagine being able to

identify at a glance the unlikely allies who might join you in co-sponsoring legislation. Data transparency combined with technology makes this possible.

THE TRANSPARENCY AND TECHNOLOGY MIND MELD

To meld transparency and technology, it was important to shift mindsets on how information is handled by Congress. I identify a few of the significant issues problem we had to overcome below. Please look to the appendix for the *10 Principles for Opening Up Government Information* that discusses best practices for publishing government data in greater detail.

For what it's worth, the Clerk's office has described the qualities of government data in slightly different terms than is described in the principles, and I'll reproduce that list here: government data should be accessible, accurate, complete, described, free, machine readable, permanent, searchable, timely, and usable. This is a good list.

Here are some of the areas where we ran into bumps along the road.

Authenticity. The way to determine that information is authentic — that it hasn't been altered in any way — is by comparing the information you have against an official data source. So, if you want to know that your bill or committee report or dataset is accurate, you must have an official source to compare it against. While this seems straightforward, it represented a significant shift in mindset from the view that authenticity could be inherent in a digital file or a physical document. The older perspective was premised on the false belief that documents couldn't be altered or faked, which simply isn't true. Any type of document — print or digital — can be manipulated to appear authentic: but an official online version can act as a check to validate authenticity.

Data Accessibility. The government routinely publishes information as a PDF, but PDFs alone aren't good enough for most purposes. Governments routinely use PDFs to provide a visual representation of a print document. If what you want is a visual representation only, this is a fine format, but if you want to be able to make use of the data inside the file, a PDF is not sufficient. You should also publish that information as data, ideally as structured data. Sometimes that means publishing the data separately, and sometimes it means publishing the structured data in a way that piggy-backs on the PDF. This is the difference between being given a picture of a spreadsheet and given the actual spreadsheet.

Structured Data. To get the most use out of your information, you've got to make it understandable. That means information must be published as data, and ideally that data should have structure to it. For example, when you fill out a government form, you're often asked for your Social Security Number. This is how the government knows it is you and not someone else.

The same is true for legislative data. What that means is if you refer to an item (like a bill) in a legislative document, you should have something in the code that tells the computer: "hey, this is a bill!" There are many of these kinds of IDs in use. Legislative documents must be infused with structure to make it understandable to computers.

This is where things get really nerdy. There are a lot of structured data formats with weird names like XML and JSON. You don't really need to know what they are called. These languages are different ways of encoding legislative information, to say "Hello computer, this thing right here is HR 123." In fact, Congress has its own version of a structured data format, USLM, or United States Legislative Markup, which is a standard way to describe legislative information. It's pretty amazing what's been accomplished here, and the expanding deployment and adoption of USLM makes it possible to do many useful things.

Findable and downloadable. You've got to be able to find government information. That means that it should be routinely published in a central location and updated on a regular basis. And you should be able to get it in the way that you need. Sometimes you want all the information at once — this is called bulk data. For example, you can download every single bill introduced in Congress over the last 10 years. And sometimes you just have a particular question, like what happened to HR 123 in the 114th Congress. You can use a tool called an API — think of it as your local Library reference desk — and ask it that specific question and it will give you the answer. For this to be useful, the data itself must be considered official when used for official purposes, which can require amending the House rules in certain instances.

Eat your own dogfood. To make sure that the information you are publishing is accurate and useful, the people publishing the information must also be using it. This is colloquially called eating your own dog food. What it means is that you're using the information that you're publishing in your own computer systems so you have an incentive to get it right. It also means you have to be publishing information in a timely way, so that you've got it when you need it.

Free. Government data should be publicly available at no cost. There's a long history of government agencies charging for access to information, and all it does is empower the creation of gatekeepers between people and the information they need. Because of the siloed nature of the legislative branch, what it has meant in practice is that components of the legislative branch often could not obtain the information they needed from other components of the legislative branch. In one particularly notable example, House Democratic Leadership was turning to civil society to obtain data on legislation because the Library of Congress was not publishing the information online as data, and for a while the only way to obtain the official data was to buy it from the government or to reassemble it yourself in an awful and laborious process.

Not every tool can or should be built by Congress, but some should. There's real value in having Congress put out information and having civil society or entrepreneurs make use of it and combine it with information from other sources. Nevertheless, Congress should be in the business of building tools for members and staff and building tools for public use, although it should be the only game in town. While Congress shouldn't be building everything, there are some things it absolutely must build itself. This tension is a good one.

Congress should have available to it — and available to civil society — enough data and technological tools that it can easily access and use its own information and not have to purchase information about Congressional activities from the private sector. Staff shouldn't be wasting their time with tasks that computers can do better. They have more important things to do. And lobbyists, who spend the money to turn information into data or to purchase information from providers that do, shouldn't know more about congressional activities than Congress itself.

TRANSPARENCY BEYOND TECHNOLOGY

There's a lot more to transparency than improved technology, of course. There's also making sure there's enough time for members to consider legislation, or that meetings are truly open, or that support offices and agencies regularly report out on their activities, or that the watchdogs (like the House IG) are given bite by making some of their reports publicly available. Some congressional support offices and agencies lean into their mission and routinely publish information online. Others do everything they can to avoid online publication until ordered to do so. There should be a presumption that information should be published online by default unless there's a good reason to hold it back.

For a comprehensive list of how the House Rules should be changed to strengthen Congress, please see our recommendations.[1] Many of them concern improving transparency that go far beyond technology. In addition, see our recommendations to House and Senate appropriators on strengthening Congress, where a few changes can yield outsized benefits.[2]

So, what has changed with respect to making the House strong and more resilient as the result of our efforts? Inside Congress, there is better access to information and more of it is usable.

---

[1] The House Rules Reform Recommendations are available online here: https://s3.amazonaws.com/demandprogress/reports/House_Rules_Reform_Recommendation_2018-09-12.pdf. Draft legislative language is available here: https://s3.amazonaws.com/demandprogress/reports/116th_Congress_House_Rules_Reform_Legislative_Language_2018-12-05.pdf.

[2] Demand Progress Appropriations Requests, FY 2020, available at https://s3.amazonaws.com/demandprogress/documents/Demand_Progress_FY_2020_Appropriations_Requests.pdf.

Members can see legislation set to go to the floor and all proposed amendments at the new rules.house.gov website, and all committee documents are managed in a central repository. We have a better understanding of how staff work in Congress and there is a growing effort to improve their working conditions. There is now a central committee calendar, so it is no longer necessary to pay a private service to find out about congressional activities. And there's more happening behind the scenes.

Outside of Congress, it is increasingly possible for people to meaningfully participate in the legislative and oversight process. They can better understand what Congress is doing, educate themselves about the issues before Congress, and provide meaningful input. In addition, the disparity in power between expensive lobbyists and their public sector counterparts is growing at a slower rate. There's a lot more that can be done here.

CLOSING THOUGHTS

I am not interested in transparency for transparency's sake, but transparency for democracy's sake. Transparency is an engine that powers congressional deliberations — in committees, in caucuses, in small groups, on the floor, in the media. When used properly, transparency has the ability to propel the body politic into action, and when used improperly it merely has the ability to shock.

I want to use transparency to help build trust in Congress — trust that every member of Congress has a real opportunity to do his or her job and can be held accountable for it. On the House floor, this is called regular order. In a committee it's called fair play. In government, it's called a fair shake. I just think that's fair.

Transparency alone won't make our politics work. It won't restore faith in our political system. It won't guarantee that everyone will be treated fairly. But it can be a bridge to all those things, when we use it appropriately.

———

I have two documents to this testimony in the hopes that they will provide useful context. They are: (1) our written testimony to House Legislative Branch Appropriators on our technology requests; (2) 10 open government data principles.

Ultimately, what is needed is better coordination of technology-empowered transparency efforts, additional House rules changes to create a presumption of openness that empowers all members of Congress, and more funding. I recommend the creation of a Legislative Branch Chief Data Officer, who can help to encourage the publication of legislative information and a

harmonization of data transparency efforts across the different siloes, without being viewed as favoring any particular component. I've described what that position could do in the appendix. For more as my thoughts continue to evolve, please feel encouraged to read my weekly newsletter on congressional capacity, available at firstbranchforecast.com.

Thank you for the opportunity to testify.

Testimony of Daniel Schuman, policy director, Demand Progress
Before the House Legislative Branch Appropriations Committee
Concerning Legislative Branch Technology Appropriations Requests
For the FY 2020 Appropriations Bill

Dear Chairman Ryan, Ranking Member Herrera Beutler, and members of the subcommittee:

Thank you for the invitation to testify again before the House Legislative Branch Appropriations subcommittee. Our testimony focuses on technological innovation in the legislative branch, with a particular focus on the Bulk Data Task Force and the Library of Congress.

But before we get into the weeds, thank you for your bipartisan leadership in the 115th Congress in support of a number of initiatives to modernize the House of Representatives. You included more than a half-dozen significant reforms — more than we have seen in my decade-long observation of this committee — and we can already see the positive effects. Thank you.

As you know, Congressional technological innovation is important because it implicates the very ability of the House to carry out its legislative, oversight, and constituent service duties in an effective, efficient, and responsive manner. The offices and agencies that support the work of Members of the House of Representatives rely upon a complex series of interdependent technologies that together affect how easy or difficult it is for Congress to do its job. When the Legislative Bulk Data Task Force was created by this Committee in 2013, we saw marked improvements in how these offices and agencies worked with one another and communicated with the general public. The Task Force had a limited purpose, but the collaboration it fostered changed the culture of Capitol Hill for the better.

We propose to build upon the accomplishments of the Bulk Data Task Force and to address a concern that has arisen concerning communications with the Library of Congress. **We make the following four requests**:

1. Create a legislative branch Chief Data Officer
2. Expand the Bulk Data Task Force into the Congressional Data Task Force
3. Increase technology funding for the Clerk of the House
4. Establish a Public Information Advisory Committee for the Library of Congress

The Bulk Data Task Force and a Chief Data Officer

In recent years, the legislative branch has made significant advances in releasing legislative information to the public as data. This has served Congress well, as it has facilitated Congress's

access to its own data — both as raw structured data and as data refined by third parties. These data publication initiatives have included the online publishing of bills; committee schedules, documents, and videos; an online House phone directory; CRS reports; the bills and amendments scheduled for a floor vote in the House; the Statement of Disbursements; the new joint meetings calendar; as well as holding regular meetings of the Bulk Data Task Force and the annual Legislative Data and Transparency Conference. These efforts are welcome and encouraged.

In fact, Deputy Clerk Bob Reeves performed phenomenally in coordinating the Bulk Data Task Force, and we are grateful to him. Indeed, the vast majority of participating offices and agencies have gone out of their way to be helpful and collaborative. House Administration Committee Technology Policy Director Reynold Schweickhardt played a particularly notable role.

With the complexity and distributed governance of information in Congress, it is helpful to have a touchstone that can help facilitate a coordinated approach to manage that data and support ongoing work to transform it into useful information.

We respectfully request that you establish a Legislative Branch Chief Data Officer to add further support to the Bulk Data Task Force. The CDO should have the responsibility for tracking datasets released by the legislative branch; providing advice, guidance, and encouragement to offices regarding the publication of legislative branch information as data; supporting the work of the Bulk Data Task Force, including assisting Deputy Clerk Reeves; coordinating the annual Legislative Data and Transparency Conference; and providing assistance to the public with finding and obtaining legislative data.

We additionally recommend an expansion of the role of the very successful Bulk Data Task Force into the Congressional Data Task Force. Congress established the Legislative Bulk Data Task Force with a focus on the question of determining whether Congress should make the legislative data behind Congress's information system, THOMAS and LIS, available to the public as structured data. Ultimately the Task Force recommended and GPO implemented the publication of bill summary, status, and text information online as structured data.

Perhaps more importantly, the Task Force — which brought together many of the technology stakeholders inside the legislative branch as well as members of civil society — continues to hold public meetings on a quarterly basis as well as innumerable Congress-only meetings. This has led to ongoing collaboration among all the stakeholders that has changed the culture of Congress and quietly led to many technological advances concerning legislative operations and transparency. The Task Force serves as a platform for people inside and outside Congress to develop innovative products and tools that help Congress using information released by Congress. Leadership of both parties have quietly blessed this group's activities.

We encourage you to expand the Bulk Data Task Force into the Congressional Data Task Force. The legislative language establishing the Task Force focuses on bulk access to legislative data, with bulk access being one mechanism by which data can be published, and legislative data being narrowly construed to information only about legislation. On its original mission, the Task Force has surpassed expectations. An expanded mission would formally allow the Task Force to look at how data is handled throughout the legislative branch. It would officially allow it to expand its scope beyond bills and the data attendant to them. This would allow consideration of other legislative documents, the handling of information used for oversight, information used and published in responding to constituents, and providing key insights about the operations of Congress itself.

In addition, **we support the request of the Clerk of the House for additional funds to create three new positions inside its Legislative Computer Systems Office to assist with its transparency and technology efforts.** Given the work of this Committee, the House Administration Committee, and the Select Committee on the Modernization of Congress, it is likely that new technology projects will be requested of the Clerk's office, which already is stretched thin and has been forced to reallocate resources because of unplanned projects. Accordingly, the Committee may want to consider providing additional funds beyond their request to the Clerk's office for technology given the likelihood of new project requests.

Public Information Advisory Committee for the Library of Congress

The Library of Congress is proud of its reputation and role as the largest library in the world. The Library plays an important role in providing information about Congress to Congress and the general public (such as through the website Congress.gov), but the Library — at least in our experience — has not prioritized its role as a source of information and is not in regular contact with civil society, especially those with expertise in facilitating public access to congressional information. This is a missed opportunity and reflects an unfortunate pattern of behavior.

The Library of Congress did not consult with civil society prior to releasing its Digital Strategy, which notably did not address the Library's role in collecting, organizing, preserving, digitizing, publishing, and contextualizing the legislative activities of Congress for the American people. There are significant deficiencies in the Library's implementation of the congressional calendar that you requested in last year's appropriations bill, most notably in how the information is displayed. We continue to have deep concerns with its implementation of the CRS Reports

website, especially in that information is published only as a PDF. For a decade we have asked that the Constitution Annotated be publicly available in a more usable format, but the Library has not engaged with us even as it apparently moves forward with plans for a major upgrade. We have deep concerns with the Library's plan to create a Congress.gov app for $750,000. And we note its decades-long opposition to public access to the legislative data under prior leadership.

This is not intended as a broadside of criticism against the Library, especially as it has been under new leadership for the last few years. We believe the Library is a pivotal institution in providing Congressional and public access to information about Congress's work. We support its funding request in full. But we in civil society are bewildered when we hear that Library staff feel discouraged from participating in the House's Legislative Data and Transparency Conference or in talking with its participants. We are dismayed when the Library does not fulfill a request from the now-Chair of the House Administration Committee to have someone from the Library talk with civil society about the CRS Reports website. And we are saddened when the Library's implementation of requests from Congress do not to satisfy the purposes for which the request was made. The Library's difficulties in managing its information technology are well documented by the Government Accountability Office, and we welcome the creation of the position of Chief Information Officer and the hiring of Bud Barton as the first CIO. There is no doubt there are good people at the Library who strive to support Congress and the Library's public mission, and we want to empower them.

It is not unusual for agencies to show reticence to talk with civil society, but there is a model that can support changing an agency's culture to one of inclusion and conversation. Other legislative and executive branch agencies and entities routinely meet with civil society stakeholders to share information and provide a foundation for collaboration. Inside the Legislative Branch, the aforementioned Bulk Data Task Force meets quarterly concerning bulk access to congressional data, the Advisory Committee on the Records of Congress semi-annually convenes congressional historians, and the Federal Depository Library Council is an ongoing point of contact for depository libraries. In the executive branch, the FOIA Advisory Committee meets monthly as a point of focus for FOIA practitioners and agency officials, the Archivist meets regularly with civil society, and so on. While we note that the Library participates in the Bulk Data Task Force, there are significant limits to its engagement that reflect its functional units and institutional reluctance.

To our knowledge, the Library of Congress does not have any regular mechanism by which it convenes external and internal stakeholders to share information on the Library's legislative information activities. Because of the Library's outsized role as an information provider, we believe it is important for it to scale its public-facing engagement to match. We recommend that such an advisory body be established with broad internal and external stakeholder representation

that would hold regular public meetings where a productive interchange can take place. These stakeholders should reflect the functional units inside the Library and the civil society organizations that are well known to Congress regarding public access to congressional information. Many of our concerns are rooted in a lack of conversation about what the Library is doing, and what it plans to do concerning public access to legislative information.

Accordingly, we urge the creation of a Library of Congress Public Information Advisory Committee. We recommend the following report language:

> *The Library of Congress is encouraged to create an Advisory Committee on Public Access to Congressional Information, composed of internal and external stakeholders that may be a source, consumer, or republisher of information or data concerning Congress, with a particular focus on legislative information. The Advisory Committee shall meet no fewer than 6 times a year in open session. The Library is encouraged to consult the Advisory Committee on a regular basis, not just at its meetings, concerning the information it gathers, holds, or publishes regarding Congress, and how that information is presented and released to the public.*

We understand that the Library may not initially welcome the creation of such an advisory committee. Nevertheless, we believe that deepening engagement with civil society on technology will help the Library of Congress fulfill its mission to "engage, inspire, and inform Congress and the American people with a universal and enduring source of knowledge and creativity." Conversation across government silos and with those on the outside often results in the sharing of new approaches to addressing technology challenges, the resolution of problems before they crop-up, greater understanding of the opportunities and constraints posed by new technology, and increased adaptability of technology for more uses and for more users. In short, this would be a win for Congress, a win for the Library, and a win for the public.

Thank you for the opportunity to testify.

TEN PRINCIPLES FOR OPENING UP GOVERNMENT INFORMATION[3]

1. Completeness

Datasets released by the government should be as complete as possible, reflecting the entirety of what is recorded about a particular subject. All raw information from a dataset should be released to the public, except to the extent necessary to comply with federal law regarding the release of personally identifiable information. Metadata that defines and explains the raw data should be included as well, along with formulas and explanations for how derived data was calculated. Doing so will permit users to understand the scope of information available and examine each data item at the greatest possible level of detail.

2. Primacy

Datasets released by the government should be primary source data. This includes the original information collected by the government, details on how the data was collected and the original source documents recording the collection of the data. Public dissemination will allow users to verify that information was collected properly and recorded accurately.

3. Timeliness

Datasets released by the government should be available to the public in a timely fashion. Whenever feasible, information collected by the government should be released as quickly as it is gathered and collected. Priority should be given to data whose utility is time sensitive. Real-time information updates would maximize the utility the public can obtain from this information.

4. Ease of Physical and Electronic Access

Datasets released by the government should as accessible as possible, with accessibility defined as the ease with which information can be obtained, whether through physical or electronic means. Barriers to physical access include requirements to visit a particular office in person or requirements to comply with particular procedures (such as completing forms or submitting FOIA requests). Barriers to automated electronic access include making data accessible only via submitted forms or systems that require browser-oriented technologies (e.g., Flash, Javascript, cookies or Java applets). By contrast, providing an interface for users to download all of the information stored in a database at once (known as "bulk" access) and the means to make specific calls for data through an Application Programming Interface (API) make data much more readily accessible. (An aspect of this is "findability," which is the ability to easily locate and download content.)

5. Machine readability

[3] Available at https://sunlightfoundation.com/policy/documents/ten-open-data-principles/.

Machines can handle certain kinds of inputs much better than others. For example, handwritten notes on paper are very difficult for machines to process. Scanning text via Optical Character Recognition (OCR) results in many matching and formatting errors. Information shared in the widely-used PDF format, for example, is very difficult for machines to parse. Thus, information should be stored in widely-used file formats that easily lend themselves to machine processing. (When other factors necessitate the use of difficult-to-parse formats, data should also be available in machine-friendly formats.) These files should be accompanied by documentation related to the format and how to use it in relation to the data.

6. Non-discrimination

"Non-discrimination" refers to who can access data and how they must do so. Barriers to use of data can include registration or membership requirements. Another barrier is the uses of "walled garden," which is when only some applications are allowed access to data. At its broadest, non-discriminatory access to data means that any person can access the data at any time without having to identify him/herself or provide any justification for doing so.

7. Commonly owned or open Standards

Commonly owned or open standards refer to who owns the format in which data is stored. For example, if only one company manufactures the program that can read a file where data is stored, access to that information is dependent upon use of the company's processing program. Sometimes that program is unavailable to the public at any cost, or is available, but for a fee. For example, Microsoft Excel is a fairly commonly-used spreadsheet program which costs money to use. Freely available alternative formats often exist by which stored data can be accessed without the need for a software license. Removing this cost makes the data available to a wider pool of potential users.

8. Licensing

The imposition of "Terms of Service," attribution requirements, restrictions on dissemination and so on acts as barriers to public use of data. Maximal openness includes clearly labeling public information as a work of the government and available without restrictions on use as part of the public domain.

9. Permanence

The capability of finding information over time is referred to as permanence. Information released by the government online should be sticky: It should be available online in archives in perpetuity. Often times, information is updated, changed or removed without any indication that an alteration has been made. Or, it is made available as a stream of data, but not archived anywhere. For best use by the public, information made available online should remain online, with appropriate version-tracking and archiving over time.

10. Usage Costs

One of the greatest barriers to access to ostensibly publicly-available information is the cost imposed on the public for access–even when the cost is de minimus. Governments use a number of bases for charging the public for access to their own documents: the costs of creating the information; a cost-recovery basis (cost to produce the information divided by the expected number of purchasers); the cost to retrieve information; a per page or per inquiry cost; processing cost; the cost of duplication etc.

Most government information is collected for governmental purposes, and the existence of user fees has little to no effect on whether the government gathers the data in the first place. Imposing fees for access skews the pool of who is willing (or able) to access information. It also may preclude transformative uses of the data that in turn generates business growth and tax revenues.