Guidelines for Open Data Policies

Compiled by the Sunlight Foundation
Version 2 | August 2013

Introduction

The Sunlight Foundation created this living document to present a broad vision of the kinds of challenges that open data policies can actively address.

A few general notes: Although some provisions may carry more importance or heft than others, these Guidelines are not ranked in order of priority, but organized to help define What Data Should be Public, How to Make Data Public, and How to Implement Policy -- three key elements of any legislation, executive order, or other policy seeking to include language about open data. Further, it's worth repeating that these provisions are only a guide. As such, they do not address every question one should consider in preparing a policy. Instead, these provisions attempt to answer the specific question: What can or should an open data policy do?

Index

15. Create a portal or website devoted to data publication or policy
16. Publish bulk data
17. Create public APIs for accessing information
18. Mandate electronic filing
19. Mandate ongoing data publication and updates
20. Create permanent, lasting access to data
21. Build on the values, goals, and mission of the community and government

How To Implement Policy (Page 16)
22. Create or appoint oversight authority
23. Create binding regulations or guidance for implementation
24. Create new legal rights or other mechanisms
25. Incorporate public perspectives into policy implementation
26. Set appropriately ambitious timelines for implementation
27. Create processes to ensure data quality
28. Create a public, comprehensive list of all information holdings
29. Ensure sufficient funding for implementation
30. Tie contract awards to transparency requirements for new systems
31. Create or explore potential public/private partnerships
32. Mandate future review for potential changes to this policy

What Data Should Be Public

1. **Set the default to open**

Most public records systems, including the Freedom of Information Act itself, are systems of reactive disclosure -- meaning that a question has to be asked before an answer given; public information requested, before it is disclosed.

Proactive disclosure is the opposite. Proactive disclosure is the release of public information -- online and in open formats (see Provisions 8 and 9) -- before it is asked for. This is no simple task, but, in a way, it's what all "open data" is aiming to accomplish. Setting the default to open means that the government and parties acting on its behalf will make public information available proactively and that they'll put that information within reach of the public (online), with low to no barriers for its reuse and consumption. Open formats may help us maximize on the value we can extract from certain kinds of public data today, but to ensure that data publishing is sustained and, in fact, made easier over time, we need to reset the default for how data is released and disclosed.

Setting the default to open is about living up to the potential of our information, about looking at comprehensive information management, and making determinations that fall in the public interest. It's about purely practical government improvements, too, and taking steps that not only keep government systems up to date, but ensure that we have the foresight to survive changes in technology that we can't predict.

Usually, for information to be defined as public, important restrictions have already been applied. Therefore, policy language can be used to outline that "all public data and information must be considered open and accessible." Whether listed as part of a statement of intent (as Austin, Texas[1] does; a concept explored more in Provision 21), as direction to a new oversight authority (Provision 22), or as the underlying aim of new data guidance (Provision 20), openness by default is a critical tool in crafting open data policies that are both ambitious and sustainable.

SAMPLE LANGUAGE

Austin, Texas cites the concept of "open by default" in a WHEREAS clause noting that

"Open Data, proactively disclosing City data, is the foundation of Open Government, is consistent with citizens' right to public information" and has benefits to government service delivery.

http://www.ci.austin.tx.us/edims/document.cfm?id=161941

2. **Reference and build on existing public accountability and access policies.**

Open data policies should be informed by provisions that are already on the books as, in most cases, they are a natural extension of existing laws, executive orders, and other policies that defend and establish public access and/or define standards for information quality, disclosure, and publishing. Pre-existing provisions in accountability policies are commonly found in open meetings acts, open records acts, ethics protections, campaign finance regulation, and lobbying disclosure laws, to name a few. Building on precedent from these policies and others, as applicable, can both help strengthen new open data requirements and inform where policy updates or revisions are necessary that an open data policy can address. Madison, Wisconsin for example, bases its definition of "public data" on both Wisconsin's state and Madison's local public records laws and ordinances in its open data policy[2].

---

[1] Austin, Texas Open Data Resolution, http://www.ci.austin.tx.us/edims/document.cfm?id=161941
[2] Madison, Wisconsin Open Data Policy,
http://madison.legistar.com/ViewReport.ashx?M=R&N=Text&GID=205&ID=1201083&GUID=2EC21911-798D-4499-BFAF-96BDACBCD8C7&Title=Legislation+Text

SAMPLE LANGUAGE

Although the provision goes on to define overreaching exemptions to data release which could have a negative impact, Madison, Wisconsin demonstrates how to use existing public records laws to help reinforce public data.

"Public data set" means a comprehensive collection of interrelated data that is available for inspection by the public in accordance with any provision of the Wisconsin Public Records Laws (Wis. Stats. §§ 19.31-19.37) and the Madison Public Records Ordinance (Sec. 3.70, MGO) and is maintained on a computer system by, or on behalf of, an agency.

http://madison.legistar.com/ViewReport.ashx?M=R&N=Text&GID=205&ID=1201083&GUID=2EC21911-798D-4499-BFAF-96BDACBCD8C7&Title=Legislation+Text

3. Mandate the release of specific new information

Open data laws provide an opportunity not just to update and improve access to information that is already open and/or public, but to specify that new data sets and records to be published.

Specific mandates can be made about a variety of kinds of data -- information ranging from transportation data to lobbying registration databases to the video and audio of public meetings -- though careful consideration should be given to the language used to describe what information is affected. Descriptive phrases such as "high-value" or "high priority", when used without direction or indication of how to assign value or priority, can open up loopholes that slow or prevent the release of information desired by the public.

It is important, therefore, that the scope of this provision be clearly defined. As with other provisions listed here, the scope can be broad or narrow, but to provide not only clarity but executable strength, the provision's scope should be explicitly defined, the limitations noted, and key agencies, committees or other relevant agents identified. Similarly, policies should be specific about what "new" data can mean: In some instances, this provision can be used to require that that new data be created, collected and released for the first time. In others (or, in addition), it could define the identification of existing data sets and (newly) mandate their release. That was Utah's approach. In the state's 2013 open data policy (SB283[3]), Utah required the identification of information that is not online and a process, including a timeline, for making that information available online. A 2006 memorandum in Washington, DC[4] took a narrower

---

[3] Utah Open Data Policy, http://le.utah.gov/~2013/bills/sbillenr/SB0283.pdf
[4] Washington, DC Open Data Policy,
http://www.scribd.com/fullscreen/26442622?access_key=key-20rfsh26eu0ob66xlbmu

approach, listing specific data sets (i.e. registered vacant properties and crime incidents) and timelines for release.

Other provisions noted on this page address how to bring the public into the process of determining how data sets can be prioritized for release.

4. **Stipulate that provisions apply to contractors or quasi-governmental agencies**

The government often uses third party entities or contractors to handle, research, or generate government information, and the use of outside services should not necessitate sacrificing important public protections. Chicago, for example, specifically directs its chief data officer to work with the chief procurement officer to develop contract provisions to promote open data standards in technology-related procurements in its [2012 open data policy](#)[5].

Similarly, these public protections should generally apply to quasi-governmental agencies and other similar actors, such as multi-state agencies, government-sponsored entities, publicly-funded universities, and self-regulatory organizations (like [FINRA](#)[6]).

> **SAMPLE LANGUAGE**
>
> **Chicago, Illinois has a specific provision about "Technology-Related Procurements."**
>
> *The chief data officer shall work with the chief procurement officer to develop contract provisions to promote open data policies in technology-related procurements. These provisions shall promote the City's open data policies, including, where appropriate, requirements to post data on data.cityofchicago.org or to make data available through other means.*
>
> http://www.cityofchicago.org/city/en/narr/foia/open_data_executiveorder.html

5. **Appropriately safeguard sensitive information**

---

[5] Chicago Open Data Executive Order, http://www.cityofchicago.org/city/en/narr/foia/open_data_executiveorder.html
[6] FINRA: Open data not on the table, Nancy Watzman, Sunlight Foundation, http://reporting.sunlightfoundation.com/2012/Finra_open_data/

Open data policy should be complementary to pre-existing legislation and directives about access to public information (see Provision 2 for more details), which means taking into consideration pre-existing protections for sensitive information for privacy, security, or other reasons. While these protections should be upheld, careful thought should be given to the language used to describe what (if any) additional information will be exempt from the policy, as overbroad terms can create loopholes that undermine the soundness of provisions requiring openness. For example, in a 2006 memorandum[7], Washington, DC, required the identification of information that should be designated private on account of law or other privacy reasons, but also requires agencies specify how the information can be aggregated, generalized, or otherwise de-identified so it can be made public. Utah's 2013 open data policy[8] simply notes that disclosure of public information should appropriately safeguard "sensitive information."

6. **Require exemptions to data release be balance-tested in the public interest**

Exemptions to disclosure are a necessary component of many transparency requirements. Unfortunately, these exemptions are often crafted as blanket categories for entire types of information, without consideration for competing interests. Valid privacy and security concerns should be addressed through provisions that recognize the public interest in determining whether information will be disclosed or not. For example, rather than saying "information relating to X topic is exempt from disclosure", provisions should require that "information relating to X topic is exempt from disclosure *if the potential for harm outweighs the public interest their disclosure.*" Public interest here does not mean public attention, but instead refers to interests like democratic accountability, justice, and effective oversight. There are some examples of broad approaches to this. Utah's 2013 open data policy[9] notes that factors "in favor of excluding public information from an information website" will be balanced against the public interest in having the information available online. San Francisco, California[10] balances privacy concerns against the "benefits of open data."

7. **Require code sharing or publishing open source**

Not only the data, but the code used to create government websites, portals, tools, and other online resources can provide further benefits, as valuable open data itself. Governments should employ open source solutions whenever possible to enable sharing and make the most out of

---

[7] Washington, DC Open Data Policy, http://www.scribd.com/fullscreen/26442622?access_key=key-20rfsh26eu0ob66xlbmu
[8] Utah Open Data Policy, http://le.utah.gov/~2013/bills/sbillenr/SB0283.pdf
[9] Utah Open Data Policy, http://le.utah.gov/~2013/bills/sbillenr/SB0283.pdf
[10] San Francisco Open Data Policy, http://sfbos.org/ftp/uploadedfiles/bdsupvrs/committees/materials/gao_032813_121017.pdf

these benefits. The Consumer Finance Protection Bureau (CFPB) began publishing open code on the social code site GitHub in 2012, citing that doing so helped them fulfill the mission of their agency and facilitated their technical work. (More information is available in the announcement blogpost on the CFPB's website.[11])

How to Make Data Public

8. **Mandate open formats for government data**

The utility, quality, and permanence of information depends on the format in which it's published. "Open" formats are considered best practice by technology and transparency communities because of their versatility: To quote Josh Tauberer[12], open formats "tend to promote a wide range of uses, backward and forward compatibility, and an independence from short-term commercial interests". In other words, these formats are machine-readable (structured), serve searchable, sortable data, and tend to be non-proprietary and/or implemented in open source software. When combined with appropriate methods of distribution, these traits maximize the degree of access, use, and quality of published information. This degree of access and interaction allows citizens and government alike to get the most out of the data.

Specific open data formats include JSON, CSV, and XML (for data sets), and HTML and plain text (which are only semi-structured, but can provide more flexibility for documents). The Open States Project[13] has explored how these formats relate to legislative data in more detail here[14]. More details about file formats can be found in the Open Knowledge Foundation's Open Data Handbook[15], Josh Tauberer's Open Data is Civic Capital[16], the 8 Open Government Data Principles[17], the 10 Open Government Data Principles[18], and The Power of Information[19] report.

Open format provisions can be broad or specific in scope. More broadly defined provisions (such as those that call for the release of "open data" with no definition) are generally hard to enforce,

---

[11] The CFPB's source code policy: open and shared, Matthew Burton,
http://www.consumerfinance.gov/blog/the-cfpbs-source-code-policy-open-and-shared/
[12] Open Data is Civic Capital: Best Practices for "Open Government Data," Josh Tauberer,
http://razor.occams.info/pubdocs/opendataciviccapital.html#bestprac
[13] Open States Project, http://openstates.org/
[14] Open States Project Best Practices Wiki, https://code.google.com/p/openstates/wiki/StateBestPractices
[15] Open Data Handbook File Formats Overview, http://opendatahandbook.org/en/appendices/file-formats.html
[16] Open Data is Civic Capital: Why Data Formats Matter, Josh Tauberer,
http://razor.occams.info/pubdocs/opendataciviccapital.html#format
[17] 8 Open Government Data Principles, Public.Resource.Org, https://public.resource.org/8_principles.html
[18] Ten Principles for Opening Up Government Information, Sunlight Foundation,
http://sunlightfoundation.com/policy/documents/ten-open-data-principles/
[19] The Power of Information, Ed Mayo and Tom Steinberg,
http://www.opsi.gov.uk/advice/poi/power-of-information-review.pdf

but can still be helpful as statements of general policy. Provisions that use more specific wording (e.g. those that define both specific data sets and the formats that they'll be released in) are more likely to cause meaningful change but take more effort to craft.

It should be noted that in this context, data refers broadly to information published in electronic formats. By this definition, data can include a variety of databases, analytics, documents, transcripts, and audio and video recordings. Although each of these examples represent different kinds of data, each can be published in an open format. Portland's 2009 open data policy[20], for example, directs the development of a strategy to adopt prevailing open standards for data, documents, maps, and other formats of media.

> **SAMPLE LANGUAGE**
>
> **A simple, but strong, top-level definition of open formats could include the following two provisions.**
>
> *[Data shall] be published in a non-proprietary, searchable, sortable, platform-independent, machine-readable format;*
>
> *Any data reporting standards designated under this subsection shall be capable of being continually upgraded as necessary.*

9. **Require public information to be posted online**

The government makes tremendous amounts of information available to the public, but only a small subset is available on the Internet, even as more and more people look online first to find these records. To close this gap, public information should be published online in a timely fashion subject only to common-sense exceptions (such as redacting personally identifiable information in certain contexts). Online publication can be enhanced by the creation of a specific webpage or data portal (see Provision 15) but to ensure sustainability of public access, it is important that the data not be tied to the existence of any one webpage or portal. Webpages and portals are good vehicles for public distribution, but the goal of this provision is to shift the foundation of public access to information more broadly so that it can be sustained even as technology and our use of online services change over time. The "Public Online Information Act" has been introduced on the federal level[21] to require public information to be available online.

> **SAMPLE LANGUAGE**

---

[20] Portland, Oregon Open Data Policy, http://www.portlandonline.com/shared/cfm/image.cfm?id=275696
[21] The Public Online Information Act, Sunlight Foundation, http://sunlightfoundation.com/policy/poia/

Utah both has provision specific to an "information website" or portal, but also makes online publishing itself a goal.

[The Transparency Advisory Board shall determine 'guidance that will make recommendations about how to make public information more readily available to the public', including]

the identification of public information not currently made available online and the implementation of a process, including a timeline and benchmarks, for making that public information available online

http://le.utah.gov/~2013/bills/sbillenr/SB0283.pdf

10. **Remove restrictions for accessing information**

Open data that is out of reach of the public is hardly open. To provide truly open access, you must provide both the right to reuse government information (explored in Provision 11) and remove arbitrary technical restrictions, such as registration requirements, access fees, and usage limitations, among others. Whether these technical restrictions have been specifically put in place (i.e. access fees) or are the accidental result of the choice of data format or software (i.e. usage limits or copyright restrictions), it is appropriate for an open data policy to address and remove these barriers to access. The aim should be to be to provide broad, non-discriminatory, free access to data so that any person can access information at any time without having to identify him/herself or provide any justification for doing so. More detailed exploration of these limitations can be found in Josh Tauberer's Open Data is Civic Capital: Best Practices for "Open Government Data"[22].

11. **Remove restrictions on reuse of information**

Most restrictions on the reuse of government information serve no purpose but to restrict the public value of important information. If information is to be truly public, there should be no license-related barrier to the public's interaction with or use of that information. Outside of data legally exempted from public use or access because of privacy or security restrictions (see Provisions 5 and 6), to be completely "open," public government information should be released completely into the public domain and clearly labeled as such. At a minimum, licenses that grant the right to use, download, and reproduce government data can be applied. The fewer

---

[22] Open Data is Civic Capital: Why Data Formats Matter, Josh Tauberer, http://razor.occams.info/pubdocs/opendataciviccapital.html#format

restrictions the better. Opening data into the public domain (or at least explicitly into free public use) removes arbitrary barriers to information access (more explored in Provision 7), helps disseminate knowledge, aids in data preservation, promotes civic engagement and entrepreneurial activity, and extends the longevity of the technological investments used to open information in the first place.

The state of Utah required that recommendations for data disclosure and format selection will remove restrictions on the reuse of public information in their 2013 open data law[23]. In 2012, New Hampshire[24] passed a bill that required its data to be made license-free, meaning "not subject to any copyright, patent, trademark, or trade secret regulation," and goes on to elaborate even further.

12. **Require publishing metadata or other documentation**

Metadata and other documentation about the data provided by the government can be useful to the public and government alike. Notations such as these both add potentially helpful context about the data's creation that will aid in the public's use of that information and support archival and data quality efforts. The Open Data White Paper[25] released by the UK's Ministers of State for the Cabinet Office and Paymaster General in June 2012 notes that the UK data portal www.data.gov.uk already includes "basic metadata about all its data sets, including timing and geographical scope" as well as "a link to a departmentally supplied description of the data and details of a contact point within the department who data users can ask for further details" (2.46, Principle 14[26]). Madison, Wisconsin's open data policy directs public data sets to include metadata[27] and to make it available to the public through the web portal.

13. **Mandate the use of unique identifiers**

Unique identifiers within data sets empower analysis and reuse by allowing disparate data sets to be combined and to help data to more be more carefully mapped to real-world entities. Without unique identifiers, some analyses can become difficult or impossible, since similar names may or may not refer to the same entities. Importantly, identifiers should be

---

[23] Utah Open Data Policy, http://le.utah.gov/~2013/bills/sbillenr/SB0283.pdf
[24] New Hampshire Open Data Policy, http://www.nhliberty.org/bills/view/2012/HB418
[25] UK Open Data White Paper, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf
[26] UK Open Data White Paper, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf
[27] Madison, Wisconsin Open Data Policy, http://madison.legistar.com/ViewReport.ashx?M=R&N=Text&GID=205&ID=1201083&GUID=2EC21911-798D-4499-BFAF-96BDACBCD8C7&Title=Legislation+Text

non-proprietary and public. Unique Identifiers are often required for [lobbying disclosures](28), for example. See also this list of [extensive resources](29) about the need for unique identifiers for corporate entities.

14. **Require digitization and distribution of archival materials**

Open data policies can address not only information currently or soon to be available in an electronic format, but also undigitized archival material. See for example [Vancouver's Open Data motion](30), which critically notes not only the importance of thoughtful digitization of archival information but the imperative to release this data to the public, ideally eventually in the same formats and in the same locations as modern data.

15. **Create a portal or website devoted to data publication or policy**

Data portals and similar websites can facilitate the distribution of open data by providing an easy-to-access, searchable hub for multiple data sets. At their best, these portals or hubs promote interaction with and reuse of open data (see the note about bulk data in Provision 16) and provide documentation for the use of information (see Provision 12). Portals can be generalized (e.g. an "open data portal", like Data.gov) or specific (e.g. a spending or ethics portal, like Colorado's [TRACER](31)), and can vary in terms of their sophistication.

[Chicago](32), Illinois, and [New York](33) state's open data policies require data to be shared on specific sites, like data.cityofchicago.org and data.ny.gov, or a successor website still maintained by or on behalf of the government.

Portals and other related websites also provide governments with the opportunity to go into detail about issues and policies related to its commitment to openness and transparency. NASA details their open government and data activities on http://www.nasa.gov/open/ as part of their compliance with the White House's Open Government Initiative. Austin, Texas also keeps a

---

[28] A State by State Look at Lobbyist Disclosure, Eric Dunn, Sunlight Foundation, http://sunlightfoundation.com/blog/2011/07/19/a-state-by-state-look-at-lobbyist-disclosure/
[29] Six Degrees of Corporations, Sunlight Foundation, http://sunlightfoundation.com/sixdegrees/resources/
[30] Vancouver Open Data Motion, http://former.vancouver.ca/ctyclerk/cclerk//20090519/documents/motionb2.pdf
[31] Colorado TRACER Campaign Finance Disclosure Website, http://tracer.sos.colorado.gov/PublicSite/homepage.aspx
[32] Chicago Open Data Executive Order, http://www.cityofchicago.org/city/en/narr/foia/open_data_executiveorder.html
[33] New York Open Data Policy, http://www.governor.ny.gov/executiveorder/95

top-level website, AustinGO2.0[34], as a hub for public communication and access to the city's ongoing open government-related activities and uses this platform to guide residents to their data portal[35].

To facilitate their "findability" these websites (and others) should be allowed to be indexed and searched by third parties (such as search engines).

>
> **SAMPLE LANGUAGE**
>
> **New York state's 2013 policy created an "Open Data Website" and outlines steps to make it sustainable and comprehensive.**
>
> *An online Open Data Website for the collection and public dissemination of Publishable State data, and, to the extent feasible, reports is hereby established. The Open Data Website shall be maintained at data.ny.gov or such other successor website maintained by, or on behalf of, the State, as deemed appropriate by the New York State Office of Information Technology Services in consultation with the Governor's Office and Data Working Group established below. The Open Data Website will provide "single-stop" access to Publishable State data that is owned, controlled, collected or otherwise maintained by covered State entities as defined herein and, to the extent feasible, reports of such covered State entities.*
>
> http://www.governor.ny.gov/executiveorder/95

16. Publish bulk data

Bulk access is a simple, but effective means of publishing data sets in full, giving the public the ability to download all of the information stored in a database at once. This is a step beyond simply making select data sets or search results available for download or export, and is critical for supporting the maximal reuse and analysis of data. Whether offered as a feature of a data portal or even as a simple "click to download" button on a government agency webpage describing or displaying information, bulk access to information is often one of the simplest and most direct steps a government entity can take to share information with the public. Many states, like Arkansas and Kentucky, allow for "authorized subscribers" to download records in bulk from different departments, but bulk data should ideally be made available to the public at large without restrictions to access (see Provision 10), as Utah[36] calls for in its 2013 open data policy. Daniel Schuman and Eric Mill explore bulk data further as it relates to legislative data and the federal

---

[34] Austin Government Online, http://austintexas.gov/austingo2.0
[35] Austin, Texas data portal, https://data.austintexas.gov/
[36] Utah Open Data Policy, http://le.utah.gov/~2013/bills/sbillenr/SB0283.pdf

THOMAS system in this blog post[37].

17. **Create public APIs for accessing information**

Although bulk data (Provision 16) provides the most basic access to searching and retrieving government data, government bodies can also develop APIs, or Application Programming Interfaces, that allow third parties to automatically search, retrieve, or submit information directly from databases online (see Josh Tauberer's Open Data is Civic Capital[38]). Navigating requirements for bulk data and APIs should be done in consultation with people with technical expertise as well as likely users of the information. For a lengthier discussion of the benefits of APIs, see the recently developed Federal Web Policy[39]. For a slightly more critical take on APIs (and their relationship to bulk data), see this post by Eric Mill of the Sunlight Foundation[40].

18. **Mandate electronic filing**

Many existing disclosure requirements were created as inefficient, paper-based requirements and should be updated to require online, electronic filing, as long as the filers can be reasonably expected to have access to the necessary technology. Electronic filing requirements save money, make real-time disclosure possible, and allow structured data to be created at the same moment information is being filed, whereas paper filings only make reuse and analysis more difficult.

This practice is currently in place in the United States Federal government, from the Federal Election Commission[41], where "electronic filing [is] the preferred method for committees to file reports and statements" and in state governments. In 2012, Delaware passed a bill[42] (SB 185) mandating that all lobbyist registration and disclosure be filed electronically by default.

Electronic filing provisions can be broad or narrow, but more specific clauses can be useful to

---

[37] Appropriators May Undercut Legislative Transparency, Daniel Schuman and Eric Mill, Sunlight Foundation,
http://sunlightfoundation.com/blog/2012/05/30/appropriators-may-undercut-legislative-transparency/
[38] Open Data is Civic Capital: Best Practices for "Open Government Data," Josh Tauberer,
http://razor.occams.info/pubdocs/opendataciviccapital.html#bestprac
[39] U.S. Federal Web Policy,
http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html
[40] Government: Do You Really Need an API?, Eric Mill, Sunlight Foundation,
http://sunlightfoundation.com/blog/2012/03/21/government-do-you-really-need-an-api/
[41] Electronic Filing, Federal Election Commission, http://www.fec.gov/elecfil/electron.shtml
[42] Delaware Senate Bill 185,
http://legis.delaware.gov/LIS/lis146.nsf/vwLegislation/SB+185/$file/legis.html?open

ensure the completeness of the data captured and what to do if the online e-filing service is down.

SAMPLE LANGUAGE

What follows is language for mandating electronic filing for campaign finance reports, allow the language can be used more generally.

All campaign finance reports required to be filed with the Secretary of the State shall be filed electronically using the electronic filing system developed by the secretary that is consistent with the purpose of this article and in a manner that allows the public to review such information.

A campaign finance report submitted electronically shall: (1) include the electronic signatures of the treasurer or assistant treasurer of the political committee serving at the time of the filing of the campaign finance report; (2) be published online in the Campaign Finance Online Reporting portal or another designated publicly-accessible database immediately; (3) be published in a widely accepted, non-proprietary, searchable, platform-independent, machine-readable format;

19. **Mandate ongoing data publication and updates**

It is not enough to mandate the one-time release of information: Data is often created on an ongoing basis and should be released the same way. A one-time release of data is in some sense incomplete the minute additional information is generated, but not included in the published set. Therefore, in order to ensure that the information published is as accurate and useful as possible, specific requirements should be put in place to make sure that government data is released as quickly as it is gathered and collected (in "real time"). Utah's Transparency Advisory Board's[43] processes call for continuous publication of and updates to public information, for example. This kind of rapid publishing becomes less of a burden when combined with others measures for online publishing, such as electronic filing (Provision 18), data portals (Provision 15), and APIs (Provision 17). This was evident in the rational behind provisions in Colorado legislation improving the state's campaign finance database[44]. The bill, which passed into law in 2007, set requirements that reports submitted to the state's online system to be "electronically filed...[and] made available immediately on the website."

---

[43] Utah Open Data Policy, http://le.utah.gov/~2013/bills/sbillenr/SB0283.pdf
[44] Colorado SB 07-259, http://www.leg.state.co.us/clics/clics2007a/csl.nsf/fsbillcont3/6611A31F8BAF456B872572C300557B0C?open&file=259_enr.pdf

20. **Create permanent, lasting access to data**

Information released by the government should be sticky: Once released, it must remain "findable" at a stable location or through archives in perpetuity. Although portals and websites can be vehicles for accessing this data over the long term (see Provision 13), it is critical that the data's permanent release & accessibility is defined so as to apply to the data itself, not just the means of access. Utah, for example, requires guidance to create "permanent, lasting, open access to public information[45]," in addition to requirements about publishing on a to-be-determined website.

Provisions relating to permanence can also be expanded to relate to updates, changes, or other alterations to the data. For best use by the public, these changes should be documented to include appropriate version-tracking and archiving over time (discussed in a little more detail in Provision 12). These provisions should build on the strengths of existing records management laws and procedures (see Provision 2).

21. **Build on the values, goals, and mission of the community and government**

Publishing data proactively and in open formats has many practical and normative implications which can be noted and explored in the text of an open data policy. An explicit statement of goals, values, or intention can help highlight the importance of open data and the release of information for the particular political context in which the policy is being formed and executed, and can be an important tool in bringing together support for the policy both internal and external to government. Many policies touch on a broad range of values and goals that will be furthered by allowing public access to government data, including greater government transparency, honesty, accountability, efficiency, civic engagement, and economic growth. Other policies outline how providing open data will support and expand specific employment and commerce opportunities, internal and community innovation, and general public services provision.

For a detailed look at statements of intent already in use, see the Sunlight Foundation's Open Data Policy Comparison chart[46].

---

[45] Utah Open Data Policy, http://le.utah.gov/~2013/bills/sbillenr/SB0283.pdf
[46] Open Data Policy Comparison: Best Practices, Sunlight Foundation, https://docs.google.com/a/sunlightfoundation.com/spreadsheet/ccc?key=0Ap0CEAgs-R_odFc0Qk1WNHdIUDE3bUNVd1U3WUlnblE#gid=0

How to Implement Policy

22. Create or appoint oversight authority

Some questions may defy easy treatment in the process of creating an open data policy, so it's appropriate to define a single authority that can be empowered to resolve conflicts and ensure compliance with new open data measures. Commonly, policies direct a pre-existing officer (i.e. a chief technology or administrative officer) or a specific department to oversee execution and compliance, although new positions and authorities can also be created. Utah, for example, refurbished and completely redefined the composition of a [Transparency Advisory Board](#)[47] to serve as the oversight authority for its 2013 open data policy, and the Recovery Act created a [Recovery Accountability and Transparency Board](#)[48] for similar oversight. Specifying an authority, review board, or similar body is an important step to making sure that an open data policy can actually be executed and provides a resource to address unforeseen hurdles in implementation. New oversight bodies should conduct their work independently and publicly, and can be bolstered by creating new regulations or guidance for implementation (Provision 23).

23. Create binding regulations or guidance for implementation

Open data policies should be practically aspirational, meaning that they should both define a vision for why the policy is being implemented, but also be able to provide actionable steps for the government and oversight authorities to follow to see the policy through to implementation. Creating regulations or guidance can ensure a strong, reliable policy, and usually mean the difference between policy passed for show versus policy passed for substance. Regulations help make the work of oversight and implementation authorities possible. Several US cities, like [Chicago](#)[49] and [San Francisco](#)[50], use their open data policies to give their chief information officer or chief data officers not only direction to oversee the implementation of technical standards for new open data policies, but to determine compliance. Although the state of [New Hampshire](#)[51] doesn't direct a single authority to take action in the same way, it does regulate that each state agency to adopt and review the statewide information policy. The Dodd-Frank Financial Reform bill empowers regulations to require public reporting of royalty payments made by the extractive industry (see Section [1504](#)[52]). A similar approach is taken in the proposed [DATA Act](#)[53].

---

[47] Utah Open Data Policy: http://le.utah.gov/~2013/bills/sbillenr/SB0283.pdf
[48] The Recovery Accountability and Transparency Board, http://www.recovery.gov/About/board/Pages/TheBoard.aspx
[49] Chicago Open Data Executive Order, http://www.cityofchicago.org/city/en/narr/foia/open_data_executiveorder.html
[50] San Francisco Open Data Policy, http://sfbos.org/ftp/uploadedfiles/bdsupvrs/committees/materials/gao_032813_121017.pdf
[51] New Hampshire Open Data Policy, http://www.nhliberty.org/bills/view/2012/HB418
[52] H.R. 4173 The Dodd-Frank Wall Street Reform and Consumer Protection Act, 111th Congress,

Open data policies can also direct that guidance is created from a basic framework created in the policy. So, rather than spelling out the entirety of data standards in the original policy document, some governments, like Utah[54] and Montgomery County, Maryland[55], have used their policies to direct that guidance is created to help agencies comply with online public access to non-proprietary, machine-readable data published in open formats. New York City's[56] open data policy, Public Law 11, resulted in the creation of extensive technical guidance, which you can find here. A similar approach is outlined in the proposed Public Online Information Act[57], which would create a central regulator empowered to create and set data standards for the US federal government.

24. **Create new legal rights or other mechanisms**

An open data policy can create mechanisms that allow individual members of the public to play a dynamic role in policy oversight and compliance. For example, the right to sue serves as the ultimate enforcement mechanism of the Freedom of Information Act (Section 4.K, Page 33[58]), some countries (like Canada[59]) have FOI ombudsmen with special legal enforcement powers, and some countries also have special anti-corruption agencies.

25. **Incorporate public perspectives into policy implementation**

Implementing the details of an open data policy will benefit from public participation. Open data policies not only have effects government-wide, which will require consideration, but also have consequences for a variety of stakeholder groups outside of the government. Allowing these groups to participate in the decision-making process (and make real contributions) can have great benefits for policy creation and execution. Stakeholders and experts can bring to the table valuable new perspectives that highlight challenges or opportunities that might not otherwise be obvious. Formal mechanisms for collaboration can include hearings, draft proposals open for public comment and contribution, and online resources like wikis and email lists. In 2012, New

---

http://www.govtrack.us/congress/bills/111/hr4173/text
[53] H.R. 2146 The DATA Act, http://www.opencongress.org/bill/112-h2146/show
[54] Utah Open Data Policy, http://le.utah.gov/~2013/bills/sbillenr/SB0283.pdf
[55] Montgomery County, Maryland, Open Data Policy,
http://www6.montgomerycountymd.gov/content/council/pdf/agenda/cm/2012/121126/20121126_GO2.pdf
[56] New York City Open Data Wiki,
http://nycopendata.pediacities.com/wiki/index.php/Local_Law_11_of_2012
[57] The Public Online Information Act, Sunlight Foundation, http://sunlightfoundation.com/policy/poia/
[58] A Citizen's Guide On Using the Freedom of Information Act and the Privacy Act of 1974 to Request Government Records, http://www.gpo.gov/fdsys/pkg/CRPT-106hrpt50/pdf/CRPT-106hrpt50.pdf
[59] Office of the Information Commissioner of Canada, http://www.oic-ci.gc.ca/eng/

York City created a wiki[60] to encourage collaborative input on the open data policies, standards, and guidelines that would be enacted as part of its then-newly passed open data law. In Ottawa, Canada[61], the city's open data policy directs staff to explore ways to consult the public and receive input on high-value data sets. Some cities go the distance to create working groups that contain members of the public, media, local businesses, as well as government staff, like the Utah Transparency Advisory Board[62] described in Provision 22. Other governments, like the cities of Chicago and San Francisco, call for less defined online forums to solicit feedback from the public on data sets and policies.

26. **Set appropriately ambitious timelines for implementation**

Setting clear deadlines can demonstrate the strength of a commitment and will help translate commitments into results. Deadlines can also help to identify failures clearly, opening the door to public oversight. Relevant actors should be given enough time to prepare for the changes brought on by the new open data policy, but not so much time that the policy becomes inoperable. The timeline should be firm, provide motivation for action, and have actionable goals and benchmarks that can be used as a metric for compliance. These goals or checkpoints can include qualitative and quantitative measurements.

27. **Create processes to ensure data quality**

Data quality will not be ensured through data release alone: efforts need to be made to keep the data up to date, clean, accurate, and accessible. In the executive memorandum that established that Washington, DC would share internal data on DC.gov[63], the city specified not only the need to maintain data quality but also touched, broadly, on the responsibilities of the agencies involved in resolving discrepancies or inconsistent results. In a 2012 law, New Hampshire[64] required that state government data be collected at the sources, "not in aggregate or modified forms." Other approaches to ensuring data quality include assigning specific staff responsible for maintenance (as was done in the Open Government Directive on financial data[65] (Section 2.a.)) and creating audit processes.

---

[60] New York City Open Data Wiki, http://nycopendata.pediacities.com/wiki/index.php/NYC_Open_Data
[61] Ottawa Open Data Policy, http://www.ottawa.ca/calendar/ottawa/citycouncil/occ/2010/05-12/csedc/08-ACS2010-COS-ITS-0005-Open%20data%20%282%29.htm
[62] Utah Open Data Policy, http://le.utah.gov/~2013/bills/sbillenr/SB0283.pdf
[63] Washington, DC, Open Data Policy, http://www.scribd.com/fullscreen/26442622?access_key=key-20rfsh26eu0ob66xlbmu
[64] New Hampshire Open Data Policy, http://www.nhliberty.org/bills/view/2012/HB418
[65] Open Government Directive, Executive Office of the President, http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf

In any case, data quality concerns should not be accepted as an excuse for exempting or restricting the release of information, but should rather be seen as a challenge that becomes clearer and easier to address as data is released. Data with serious accuracy and quality concerns should be adequately documented to avoid creating confusion or misinformation.

Similarly, public data reporting streams that are separate from what is used within government should be avoided whenever possible, as redundant or parallel data streams can create opportunities for data quality to suffer.

28. **Create a public, comprehensive list of all information holdings**

Government bodies often do not know what information they have. Open data policies should require a full public listing of government information. This comprehensive listing empowers policymakers and administrators to determine whether information is being appropriately managed, and empowers the public oversight of those determinations. Publicly accounting for agency information helps ensure that information is managed to benefit the public interest, and can create efficiencies among government departments, all while empowering journalists and policymakers. To provide up-to-date information, agencies can also be required to regularly audit their information holdings.

In addition to noting the data sets themselves, a data listing should note the department or agency (or agencies) responsible for the collection and maintenance of the data, its public or private classification, and, when possible, information about where to access the public data. To the extent practicable, additional details can also be given about private classifications, allowing the public to understand why certain information is marked as not-public. This is a step taken by the United Kingdom's Department for Communities and Local Government[66], which not only publishes a list of information assets held by the department, but which notes, whenever an entry on their listing cannot be made available to the public in its entire raw form, the rationale for not publishing the information.

In 2010, the US Department of Transportation released (and to this day continues to maintains) an inventory[67] of its high-value data for the public as part of its Open Government Plan[68]. Many more governments[69] are in the early stages of directing their agencies to create indexes. In 2013,

---

[66] UK Department for Communities and Local Government,
https://www.gov.uk/government/publications/dclg-data-inventory
[67] U.S. Department of Transportation Data Inventory Release, http://www.dot.gov/open/data-inventory-release
[68] U.S. Department of Transportation Data Inventory, http://www.dot.gov/open/data-archive
[69] How Unique is the New U.S. Open Data Policy?, Laurenellen McCann and Júlia Keserű,
http://sunlightfoundation.com/blog/2013/05/13/how-unique-is-the-new-u-s-open-data-policy/

a [memorandum](#)[70] accompanying a [new federal executive order for open data](#)[71] directed all agencies to compile data inventories to [a similar rigor](#)[72] defined by the UK's Department of Communities and the US Department of Transportation. Several state and local governments have also used open data policies to direct departments to create listings, including San Francisco and Utah. See also this [blog post](#)[73] from John Wonderlich explaining the need for indexes.

29. **Ensure sufficient funding for implementation**

Like any other initiative, implementing an open data policy should be done with an eye on long-term sustainability. One way to do this is to consider funding sources for the implementation of the policy as well as its future maintenance. Sufficient funding can mean the difference between successful and unsuccessful policies.

For example, in 2011, the Electronic Government Fund, which supports Data.gov, the IT Spending Dashboard, and USASpending.gov, among other programs, was sliced from over $34 million to $12.4 million. Without the work of the advocacy community, funds would have dropped as low as $8 million. This dramatic change in funding has continued implications for federal data, some of which can be explored in [the posts collected here from the Sunlight Foundation blog](#)[74].

By contrast, in 2012, California approved a bill (SB 1001) to pay for maintenance, repair, and improvements to their Cal-Access public disclosure database by [increasing the registration fees for those engaged in lobbying and with political action committees](#)[75]. Hawaii's 2013 open data policy ([HB 632](#)[76]) included provisions to appropriate two-years worth of funds to the department charged with executing the policy, the Office of Information Practices, to ensure that the implementation process was appropriately staffed.

---

[70] White House Open Data Policy Memorandum,
http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf
[71] Open Data Executive Order Shows the Path Forward, John Wonderlich, Sunlight Foundation,
http://sunlightfoundation.com/blog/2013/05/09/open-data-executive-order-shows-path-forward/
[72] Open Data Executive Order Shows the Path Forward, John Wonderlich, Sunlight Foundation,
http://sunlightfoundation.com/blog/2013/05/09/open-data-executive-order-shows-path-forward/
[73] The Missing Open Data Policy, John Wonderlich, Sunlight Foundation,
http://sunlightfoundation.com/blog/2012/03/22/the-missing-open-data-policy/
[74] Save the Data blogposts, Sunlight Foundation,
http://sunlightfoundation.com/blog/taxonomy/term/savethedata/
[75] Yee's Cal-Access Bill Reconsidered, Approved Along With Other Open Government Bills,
http://sd08.senate.ca.gov/news/2012-05-31-yee-s-cal-access-bill-reconsidered-approved-along-other-open-government-bills
[76] Hawaii Open Data Policy, http://www.capitol.hawaii.gov/session2013/bills/HB632_CD1_.htm

30. **Tie contract awards to transparency requirements for new systems**

Existing procurement, contracting, or planning processes can be used to create new defaults and requirements for IT systems and databases to ensure that open data requirements are baked into new systems as they're being planned. See for example the White House's Digital Government strategy[77], which proposes the creation of similar new requirements and encourages agencies to share best practices in procuring these contracts and solutions with each other.

31. **Create or explore potential public/private partnerships**

Partnerships can be useful in a variety of important efforts related to data release, from increasing awareness of the availability of open data to identifying constituent priorities for data release to connecting government information to that held by non-profits, think tanks, academic institutions and others. Ed Mayo and Tom Steinberg have noted[78] that such partnerships can aid civic participation, help identify the gaps in services delivery, among other benefits. Philadelphia, Pennsylvania hosts its open data portal, Open Data Philly[79], through a public/private partnership involving local journalism, business, and non-profit organizations, rather than hosting the portal on its own, solely-government platform.

Poorly planned public/private partnerships run the risk of subsidizing private sector actors at the expense of the public. For example, see the Government Accountability Office's digitization project described here[80].

Public/private partnerships are increasingly being explored in cities as a way to collaborate with regional governments and ensure that government data handled by third parties is also made open and available to the public (see Provision 4). To that end, San Francisco, California's 2013 open data policy[81] approaches partnerships by suggesting that data standards be established within and outside the city through collaboration with external organizations. Lexington, Kentucky[82] included aims to develop agreements with regional partners to publish and maintain public

---

[77] U.S. Federal Web Policy,
http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html
[78] The Power of Information, Ed Mayo and Tom Steinberg,
http://www.opsi.gov.uk/advice/poi/power-of-information-review.pdf
[79] Open Data Philly, http://opendataphilly.org/
[80] GAO did sell exclusive access to legislative history to Thomson West, Free Government Information,
http://freegovinfo.info/node/1798
[81] San Francisco Open Data Policy,
http://sfbos.org/ftp/uploadedfiles/bdsupvrs/committees/materials/gao_032813_121017.pdf
[82] Lexington Open Data Policy,
http://www.lexingtonky.gov/Modules/ShowDocument.aspx?documentid=16343

data sets.

32. **Mandate future review for potential changes to this policy**

Just as publishing open data is an ongoing process that requires attention to its quality and upkeep (Provision 26), so too does the policy that establishes it. In order to keep up with the times, current best practices, and feedback from existing policy oversight, open data policies should be written in a way that makes them open to future revision. Open data policies should acknowledge that the context in which they operate is rapidly changing over time, and will likely need sustained attention to remain relevant.